UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE INFORMÁTICA
DEPARTAMENTO DE INGENIERÍA DEL SOFTWARE E INTELIGENCIA ARTIFICIAL

TESIS DOCTORAL

Mejorando la evaluación de juegos serios mediante el uso de analíticas de aprendizaje

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Ángel Serrano Laguna

DIRECTOR

Baltasar Fernández Manjón

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA

DEPARTAMENTO DE INGENIERÍA DEL SOFTWARE E INTELIGENCIA ARTIFICIAL

TESIS DOCTORAL

Mejorando la evaluación de juegos serios mediante el uso de analíticas de aprendizaje

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Ángel Serrano Laguna

Director

Baltasar Fernández Manjón

Madrid, 2017

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA

Departamento de Ingeniería del Software e Inteligencia Artificial

MEJORANDO LA EVALUACIÓN DE JUEGOS SERIOS MEDIANTE EL USO DE ANALÍTICAS DE APRENDIZAJE

MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA POR

Ángel Serrano Laguna

Bajo la dirección del doctor

Baltasar Fernández Manjón

Madrid, abril de 2017

De aquella época recuerdo que arrojaba con alegría el tiempo por la borda,
en la esperanza de que el globo alzara el vuelo y me llevara a un futuro mejor.

Eduardo Mendoza

El enredo de la bolsa y la vida

Agradecimientos

Me senté frente a Pablo sin saber muy bien qué esperar. Durante el último año había sido mi profesor de LPS y a partir de aquel día iba a empezar a trabajar en su grupo de investigación. Yo no sabía qué era un grupo de investigación ni cuáles iban a ser mis tareas, pero el trabajo era en la facultad y tenía horario flexible, algo especialmente conveniente para seguir estudiando Ingeniería Informática (la de 5) una vez acabada la de Sistemas (la de 3).

Pablo me habló sobre e-UCM (el grupo de investigación) y sobre lo que hacían. Después se quedó en silencio esperando. Yo había entendido poco de su explicación así que asentí convencido con la esperanza de que no me preguntara nada. Tras unos segundos, me asignó mi primera tarea: traducir un videojuego educativo al inglés. Y así me familiarizaba con la herramienta que ellos utilizaban para desarrollar videojuegos educativos (algo llamado “e-Adventure”, un nombre en la línea del grupo). Le dije que yo el inglés de aquella manera, y me dijo que no me preocupara, que no tenía que quedar perfecto. Y asentí, porque yo el inglés de aquella manera y perfecto no iba a quedar.

Pasaron un par de semanas y completé mi primera tarea. Pablo entonces me llevó al que sería mi hábitat durante los siguientes cinco años: el Aula 16. Un aula (la 16) reconvertida en oficina para investigadores. Allí estaban Ángel, Javi y Eugenio, más miembros del grupo e-UCM. Ángel me saludó en tono jocoso, con un volumen algo alto y soltando una carcajada. Javi se presentó con un volumen más de interior, y me saludó como si ya nos conociéramos. Eugenio no dijo nada. Yo tampoco.

Javi me asignó una tarea: añadir una nueva funcionalidad a e-Adventure. Cuando la completé me miró sorprendido y satisfecho. Me siguió asignando tareas durante un tiempo.

Balta llegó un día al Aula 16 y nos presentaron. Balta era el director del grupo y me dijo que cómo iba a todo, que a lo mejor había que apretarme un poco las tuercas. Como yo no dije nada, me aclaró que era broma. Como seguí sin decir nada añadió que había oído cosas muy buenas de mí y yo asentí aliviado.

Después empecé a trabajar con Eugenio en una nueva versión de e-Adventure. Eugenio hablaba poco y yo hablaba poco, lo cual estaba bien porque así los dos hablábamos poco.

Con el tiempo fui conociendo al resto del grupo: Iván, Manu, Borja, Blanca… y entre todos me liaron para que hiciera el máster, y después el doctorado.

También di un curso de Android en Ecuador, donde me llamaban doctor hasta que les dije que aún no tenía el doctorado y entonces empezaron a llamarme ingeniero aunque yo les dije que Ángel estaba bien. Asistí a un curso de videojuegos educativos en Estonia y hubo una noche en la que no se puso el sol. Hice una estancia de investigación en Boston y cayó la nevada más grande de los últimos 20 años. Y estuve también cuando los de Microsoft vinieron a presentar la Surface a la facultad y nosotros bajamos sólo a que nos dieran bolígrafos.

Y ahora, estoy aquí, al final del camino.

Quiero dar las gracias a toda la gente de la Facultad de Informática. He estado 10 años en ella y es difícil no cogerles cariño. En especial al departamento de Ingeniería del Software e Inteligencia Artificial, por su apoyo en general y su ayuda en papeleos variados en particular, especialmente a Luis en su momento, a Juanan en la actualidad y a Lourdes siempre. También al resto de profesores, con los que he compartido mesa durante las comidas.

Y quiero dar las gracias a toda la gente que me he encontrado en e-UCM.

A Blanca, por hacerme ver que la multitarea es posible si tienes la motivación suficiente.

A Cristian, por recordarme la importancia de divertirse y hacer cosas que a veces dan miedo.

A Toni, por darme perspectiva sobre cómo evolucionan algunas cosas.

A Pablo, por darme la primera oportunidad y por sus críticas cinematográficas durante las comidas.

A Eugenio, por guiarme en los inicios, y por haber seguido ayudándome aun cuando estaba lejos.

A Ángel, por la ayuda, por el ingenio, por la energía desbordante y por las risas.

A Manu, por su minuciosidad y por prestar una mano siempre que la he necesitado.

A Borja, por llevarme a los coles a experimentar ciencia en acción, por su visión de la vida y por sus ganas de hacer cosas nuevas.

A Iván, por siempre saber más cosas que yo, y siempre estar dispuesto a contármelas, las veces que haga falta, con la profundidad que haga falta.

A Javi, por ayudarme a articular y mejorar mis ideas. Por las largas tardes en el Aula 16. Por las planes que no funcionaron y las cosas que aprendimos de ellos. Y por ser uno de los mejores tíos que me he encontrado en años. Gracias, Javi.

A Balta, por dejarme hacer mi camino libremente, aunque a veces haya llevado a callejones sin salida. Por ser paciente en momentos donde otros no lo habrían sido. Por confiar en mí desde el primer día. En definitiva, por todo lo que ha trabajado para traerme hasta el final de esta tesis. Gracias, Balta.

Han sido 10 años en esta facultad y he aprendido mucho (en su mayoría, cosas no relacionadas con la informática). Y aunque llego al final del camino algo cansado, sé que esta tesis doctoral es el culmen de una época que me ha definido como persona.

Ah. Y se me olvidaba. Gracias, María, por inspirarme para ser mejor cada día.

Índice de contenidos

Índice de figuras

Resumen

Este trabajo de tesis propone el uso de analíticas de aprendizaje para mejorar la evaluación en videojuegos educativos o juegos serios. El trabajo aborda dos aspectos necesarios para una evaluación de aprendizaje efectiva: 1) los requisitos necesarios en el diseño de juegos serios; 2) la infraestructura y el modelo de análisis necesario para las analíticas de aprendizaje.

Estos dos trabajos, combinados, marcan un camino para aplicar analíticas de aprendizaje a juegos serios de manera efectiva. Esta aplicación se fundamenta en proporcionar a profesores e instructores resultados de aprendizaje (o niveles de conocimiento) de sus alumnos tras interactuar con un juego serio. El proceso empieza por el diseño de un juego serio efectivo, capaz de capturar, a partir de las interacciones de los estudiantes, el proceso de aprendizaje. A partir de ahí, se define un modelo de captura de datos que recoja estas interacciones. Después, estas interacciones son analizadas para obtener resultados de aprendizaje, y, finalmente, son mostradas a través de visualizaciones a los profesores e instructores, que pueden tomar acción a partir de ahí.

El diseño de juego serios se aborda desarrollando una metodología que modela los aspectos fundamentales en el desarrollo e implementación de un juego serio, desde su incepción inicial a partir de objetivos educativos, pasando por su validación formal con usuarios expertos y terminando con el despliegue final, en el que los usuarios objetivo interactúan con el juego para lograr sus objetivos educativos. Esta metodología se ve acompañada, para cada uno de sus pasos, de las consideraciones necesarias para habilitar la evaluación automática del juego serio, y de un patrón de diseño destinado a estructurar la manera en la que los juegos serios presentan los contenidos educativos a sus jugadores.

Por otro lado, este trabajo propone e implementa una arquitectura centrada en el análisis de interacciones procedentes de juegos serios, cuyo objetivo final es la evaluación de aprendizaje o la medida de conocimiento. Esta arquitectura hace énfasis en los pasos principales de cualquier proceso de análisis: captura y selección de datos, dónde se propone un modelo de interacciones capaz de representar las interacciones más comunes dentro de los videojuegos educativos; análisis, dónde se propone una aproximación para obtener resultados de aprendizaje, tomando como eje vertebrador el patrón de diseño propuesto en la metodología de diseño; y visualización, explorando las representaciones más beneficiosas a la hora de presentar los resultados de aprendizaje de los jugadores, y centrándose en los profesores como principales beneficiarios de las mismas.

Finalmente, para que otros investigadores y desarrolladores puedan reutilizar el trabajo desarrollado, se propone un proceso de estandarización para el modelo de interacciones de juego, y se implementa una especificación haciendo uso de la especificación educativa xAPI.

Este proceso, además de haber sido desarrollado teóricamente, ha sido aplicado parcial o totalmente a diferentes juegos serios que han sido utilizados en entornos educativos reales. Esta aplicación ha permitido iterar sobre el proceso, logrando una metodología final validada.

En este trabajo se presentan los resultados de experimentos reales, llevados a cabo durante las diferentes fases del desarrollo del trabajo. En un primer experimento se probó una implementación inicial de la arquitectura de análisis, que sirvió para pulir el tipo de datos que quería capturarse. En un experimento posterior, se probó un juego serio desarrollado con las consideraciones necesarias para la evaluación de aprendizaje como principal objetivo. De esta prueba se obtuvieron los primeros requisitos de la metodología de diseño, así como una validación extra de la arquitectura de análisis. En el tercer experimento, se puso en práctica todo lo aprendido en los experimentos anteriores, completando la formalización de la metodología de diseño, así como la arquitectura de análisis. El cuarto y último experimento se utilizó como prueba de concepto para el proceso de estandarización de la captura de datos.

Abstract

This thesis proposes the use of learning analytics to improve assessment in educational videogames or serious games. The work focuses on two aspects needed for an effective learning assessment: 1) requirements related to serious games design; 2) the framework and the data analysis model needed for learning analytics.

This work presents an analysis framework and a design methodology focused on serious games. These two pieces combined set the basis for a full application of learning analytics into serious games. The main goal of this process is to offer assessment results (or knowledge levels) to teachers and instructors after their students interact with a serious game. The process starts with the design and implementation of an effective serious games. The serious game must be able to capture the interactions representing the learning process. These interactions are the basis for a capture data model that is later analyze to obtain assessment results. Finally, these results are presented to teachers and instructors through visualizations that allow them to take action.

The game design process is approached through a methodology in charge of modeling the fundamental aspects of the serious game development and implementation. The methodology starts on the serious game inception, using the learning goals as guide for the whole process. It proposes steps to validate the serious game formally, using domain experts as players of the game. And it end ups modeling the final deployment of the game, where real users play with the game and are assessed. Each of the steps in the methodology considers the needs and requirements for a successful and automatic learning assessment. Additionally, it defines a game design pattern that structure how the different educational contents of the game are delivered to the player.

On the other hand, this work proposes and implements a framework focused on the analysis of serious games interactions. The final goal of this framework is to assess students' knowledge using these interactions as primary source of data. The framework includes all the steps present in any kind of analysis process. It proposes a data model for the data selection and data capture. This model represents the most common interactions a player can do in a serious game. It proposes an analysis process, which uses the game design pattern proposed in the design methodology as basis for the assessment. And it proposes a set of visualizations, exploring the best way of representing assessment results, and focusing on instructors and teacher as main stakeholders.

Finally, the work proposes a standardization process so others can reuse some of the results. This process uses xAPI, an educational specification, to model the possible interactions coming from the serious game.

This process has been applied (partially or totally) to several serious games deployed in real educational environments. These case studies have allowed to iterate and validate the whole methodology.

This work presents the result of four real experiments carried out during the different phases of the work. The first experiment tested the initial implementation of analytics framework. The results helped to polish the type of data we needed to capture from the games. The second experiment tested a serious game developed with the learning analytics requirements in mind. The results help to start defining the design methodology and further validation of the analysis framework. The third experiment validated both the analysis framework and the design methodology. The final experiment was used as a proof of concept of the standardization process.

Acerca de este documento

Este documento incluye una introducción, un estudio del dominio, una descripción de los objetivos de la tesis y una discusión que integra los seis artículos incluidos, relacionándolos con los objetivos mencionados. A continuación, se presentan las conclusiones y el trabajo futuro derivado del trabajo. Finalmente, se incluye una bibliografía que complementa las referencias incluidas en los artículos.

Este trabajo es presentado como una recopilación de publicaciones editadas. Los artículos presentados son los siguientes:

- Ángel Serrano-Laguna, Javier Torrente, Borja Manero, Baltasar Fernández-Manjón (2015), ***Building a Scalable Game Engine to Teach Computer Science Languages***, IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, Vol. 10, Issue 4, pp. 253-261. DOI: 10.1109/RITA.2015.2486386
- Ángel Serrano-Laguna, Eugenio Jorge Marchiori, Ángel del Blanco, Javier Torrente, Baltasar Fernández-Manjón (2012), ***A framework to improve evaluation in educational games,*** Global Engineering Education Conference (EDUCON), 2012 IEEE, 17-20 abril 2012. DOI: 10.1109/EDUCON.2012.6201154
- Ángel Serrano-Laguna, Javier Torrente, Pablo Moreno-Ger, Baltasar Fernández-Manjón (2014), ***Application of Learning Analytics in educational videogames,*** Revista Entertainment Computing, Vol. 5, Issue 4, pp. 313-322. DOI: http://dx.doi.org/10.1016/j.entcom.2014.02.003
- Ángel Serrano-Laguna, Baltasar Fernández-Manjón (2014), ***Applying learning analytics to simplify serious games deployment in the classroom,*** Global Engineering Education Conference (EDUCON), 2014 IEEE, 3-5 abril 2014. DOI: 10.1109/EDUCON.2014.6826199
- Ángel Serrano-Laguna, Borja Manero, Manuel Freire, Baltasar Fernández-Manjón (2017), ***A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes,*** Multimedia Tools and Applications, 2017, pp 1-23, DOI: 10.1007/s11042-017-4467-6 [Factor de impacto JCR: 1.331; Segundo cuartil (Q2) 31/106 en "Computer Science, Software Engineering"]
- Ángel Serrano-Laguna, Javier Torrente, Borja Manero, Ángel del Blanco, Blanca Borro-Escribano, Iván Martínez, Manuel Freire, Baltasar Fernández-Manjón (2013), ***Learning Analytics and Educational Games: Lessons Learned from Practical Experience,*** International Conference on Games and Learning Alliance: GALA 2013, pp 16-28. DOI: 10.1007/978-3-319-12157-4_2
- Ángel Serrano-Laguna, Iván Martínez-Ortiz, Jason Haag, Damon Regan, Andy Johnson, Baltasar Fernández-Manjón (2017), ***Applying standards to systematize learning analytics in serious games***, Computer Standards and Interfaces, Vol.

50, pp. 116-123. DOI: 10.1016/j.csi.2016.09.014 [Factor de impacto JCR 2015: 1.268; Segundo cuartil (Q2) 35/106 en “Computer Science, Software Engineering”]

Estructura del trabajo

El núcleo de este trabajo es una recopilación de publicaciones editadas que se reproducen en su totalidad en el último capítulo. Los capítulos que los preceden comentan e integran las contribuciones particulares realizadas en cada una de ellas.

La memoria de esta tesis sigue la siguiente estructura:

- Capítulo 1. Introducción y motivación
- Capítulo 2. Estudio del dominio
- Capítulo 3. Objetivos y planteamiento del trabajo
- Capítulo 4. Discusión y contribuciones
- Capítulo 5. Conclusiones y trabajo futuro
- Capítulo 6. Artículos presentados

Las referencias bibliográficas se presentan en un apartado al final de esta memoria.

Capítulo 1. Introducción y motivación

En este capítulo se presenta la motivación inicial para la realización de esta tesis, junto a una primera aproximación al contexto en el que se desarrolla (que se verá extendido ampliamente en el Capítulo 2). A continuación, se enumeran los objetivos principales de la misma.

1.1. Motivación de la investigación

El análisis masivo de datos está cada vez más presente en la mayoría de aspectos que nos rodean. Estos análisis toman como fuente de datos cualquier indicador externo que emita información relevante, que después son sometidos a procesos de limpiado y procesado, para finalmente proveer de información accionable que se aplica para mejorar todo tipo de procesos. Muchos de estos procesos están alcanzado grandes niveles de sofisticación en diferentes áreas de estudio. Por ejemplo, en el campo de la Inteligencia Empresarial (Business Intelligence), llevan décadas analizando datos de comercio para encontrar puntos débiles en sus modelos de negocios, así como para optimizar ventas (Rud, 2009).

El campo de la educación tardó un poco más en empezar a beneficiarse de estas técnicas. Una de las primeras incursiones de las instituciones educativas en el mundo del análisis de datos fue con el uso de Analíticas Académicas (Academic Analytics). Esta disciplina está centrada en la aplicación de las técnicas de Inteligencia Empresarial a datos administrativos procedentes de instituciones educativas (Campbell, DeBlois, & Oblinger, 2007). Estos datos son, en su mayoría, información demográfica, historiales académicos, datos económicos o información institucional, que pueden ser utilizados para predecir diferentes aspectos de la institución académica. Por ejemplo, utilizando los resultados académicos de los alumnos de un curso, puede predecirse el número de alumnos que habrá en el curso siguiente.

Cada vez más, las instituciones educativas tienden a evolucionar a modelos de enseñanza on-line. En este modelo los alumnos realizan su aprendizaje a través de un dispositivo conectado a Internet, a veces de manera parcial, a veces de manera total. Una de las primeras tecnologías que hizo posible la conexión de instituciones educativas y alumnos a través de Internet fueron los Sistemas de Gestión de Aprendizaje (Learning Management System), donde profesores y alumnos pueden comunicarse a través de diferentes canales (Dougiamas & Taylor, 2003). Existen foros donde los alumnos interactúan, la posibilidad de realizar exámenes y repositorios donde los profesores pueden compartir recursos educativos son sus alumnos.

Estos nuevos tipos de interacciones on-line crearon un cuerpo de datos que representaba algo que hasta el momento no podía verse en los datos manejados por las Analíticas Académicas: el proceso de aprendizaje de los alumnos (en oposición únicamente a su resultado final). Esto abría la ventana a nuevos tipos de estudios, más centrados en las interacciones directas de los estudiantes, que pondrían algo de luz sobre los procesos internos que ocurren durante el aprendizaje de los alumnos. En ese momento nació como disciplina la Minería de Datos Educativa (Educational Data Mining), centrada en aplicar algoritmos de minería de datos en los datos provistos por los Sistemas de Gestión de Aprendizaje y otras plataformas on-line (C Romero & Ventura, 2010).

Una de las plataformas más relevantes en la masificación de datos de aprendizaje son los Cursos Masivos Abiertos On-line (Massive On-line Open Courses). Estos cursos pueden llegar a ser completados por cientos de miles de personas y en la mayoría de ellos el proceso de aprendizaje ocurre exclusivamente on-line. Esto supuso un cambio en el formato de la enseñanza on-line y en los tipos de datos de aprendizaje que podían capturarse: por primera vez, se tenían datos de aprendizaje completos de manera masiva sobre un mismo material educativo.

Aunque la Minería de Datos Educativa también se encargó de realizar estudios sobre este nuevo cuerpo de datos, una nueva disciplina empezó a emerger, con la vocación de aportar una visión propia al mundo educativo: Analíticas de Aprendizaje (Learning Analytics). Esta disciplina recoge el testigo de sus antecesoras y, además de seguir aplicando las técnicas de análisis usadas en otros campos, pretende crear sus propios modelos y procesos aplicados exclusivamente a los datos procedentes de los procesos de aprendizaje (Long & Siemens, 2011).

El informe anual Horizon, encargado de enumerar las tendencias educativas más influyentes del momento, marcó, en su edición de 2012, las Analíticas de Aprendizaje como uno de los campos a tener en cuenta a medio plazo. Las define como "el análisis de una amplia variedad de datos producidos por estudiantes, con el objetivo de evaluar su progreso académico, predecir sus futuros resultados e identificar riesgos potenciales" (L. Johnson, Adams, & Cummins, 2012). En su edición de 2016 (Johnson, L., Adams Becker, S., Cummins, M., Estrada & Freeman, A., and Hall, 2015) marca la medición de aprendizaje como uno de los campos donde más esfuerzos se van a focalizar a corto plazo

Estas analíticas se basan en el estudio de los datos de interacción, lo que hace que los contenidos educativos altamente interactivos sean especialmente interesantes para este tipo de análisis. Los juegos serios son un recurso destacado dentro de estos contenidos interactivos.

Un juego serio tiene como característica definitoria que su propósito vaya más allá del puro entretenimiento, en oposición al objetivo principal de los videojuegos comerciales

(Djaouti, Alvarez, & Jessel, 2011). En su mayoría, el aspecto que va más allá de estos videojuegos es el educativo, puesto que la mayoría de juegos serios tienen una vocación pedagógica.

Los juegos serios han sido probados y validados como herramientas educativas en diferentes dominios y niveles de educación. Se han utilizado con resultados satisfactorios en el campo de la ingeniería (Hauge & Riedel, 2012), la enseñanza de idiomas (W. L. Johnson & Wang, 2007; Ye, 2014), la medicina (Baranowski, Buday, Thompson, & Baranowski, 2008; Evans et al., 2015), física (Price, 2007; Kurt Squire, Barnett, Grant, & Higginbotham, 2004), programación (Malliarakis, Satratzemi, & Xinogalos, 2014) o el teatro y la literatura (Manero, Torrente, Serrano, Martínez-Ortiz, & Fernández-Manjón, 2015).

Entre sus bondades se encuentra su capacidad de sumergir a los alumnos en mundos virtuales, donde pueden experimentar y aprender en entornos seguros (K. Squire, 2003), y, en ocasiones, entornos que serían caros o peligrosos de reproducir en la vida real (por ejemplo, una simulación que enseña a un médico a operar a corazón abierto).

Tradicionalmente, la validación formal de los juegos serios se ha hecho a través de experimentos formales. Estas validaciones, en su amplia mayoría, utilizan como instrumentos de medida cuestionarios escritos que los alumnos deben rellenar antes y/o después de interactuar con el juego (Calderón & Ruiz, 2015). La naturaleza interactiva de los videojuegos, combinada con la aparición de disciplinas como Analíticas de Aprendizaje abre una puerta para nuevos tipos de validaciones que permitan capturar el aprendizaje dentro del juego serio mientras sucede. Es decir, capturar las interacciones de los estudiantes para poder hacer un análisis más meticuloso de los procesos de aprendizaje.

Esta idea no parte de cero: los videojuegos comerciales llevan años capturando las interacciones de sus jugadores. Las Analíticas de Juego (Game Analytics) estudian datos de interacción procedentes de videojuegos con múltiples propósitos: detección de errores, validación y pulido de mecánicas de juego, obtención de datos de retención de usuarios o análisis de la monetización (El-Nasr, Drachen, & Canossa, 2013). Y aunque alguno de estos objetivos pueden ser extrapolados, las Analíticas de Juego no tienen en cuenta un objetivo fundamental para los juegos serios: medir el aprendizaje.

A la vista de las posibilidades que se plantean en la combinación de juegos serios, con analíticas de juego y aprendizaje, esta tesis plantea la exploración de caminos que combinen el estudio de los procesos de aprendizaje en los juegos serios con las técnicas emergentes de análisis de datos presentes en otras disciplinas. Su objetivo principal es modelar, automatizar y mejorar el análisis de los procesos relacionados con el aprendizaje de los juegos serios.

1.2. Objetivos de la línea de investigación

La línea de investigación principal de esta tesis se fundamenta en mejorar la evaluación de juegos serios a través de analíticas de aprendizaje, fundamentadas en datos de interacción.

Se pretende realizar un trabajo eminentemente práctico e iterativo. En un campo tan nuevo, se considera fundamental contrastar los avances que se vayan realizando. Para ello, el trabajo será incremental y estará acompañaos de casos de estudio que validen el progreso. Al final del trabajo quiere obtenerse un modelo abstracto y una metodología genérica que incluya todo lo aprendido durante la investigación.

El enfoque de este objetivo estará dividido en dos facetas diferenciadas:

1. Modelado e infraestructura para el análisis de juegos serios. En este objetivo pretende crearse un modelo de análisis ligado a los procesos de aprendizaje dentro de juegos serios. El modelo tendrá como base las interacciones de los jugadores, y tratará de adaptar procesos presentes en otras disciplinas a la medición de aprendizaje. Como parte de este objetivo, también se modelará y desarrollará una plataforma software que dé soporte a todo el proceso de análisis.
2. Características y diseño de juegos serios para el análisis del aprendizaje. En este objetivo se pretende estudiar y definir las características necesarias para que un juego serio sea analizable por el proceso definido en el primer objetivo. Como parte de este objetivo, se creará una metodología aplicable al diseño de juegos que prepare las mecánicas de juego para el análisis de interacciones.

Los pasos principales para alcanzar este objetivo (que serán descritos en mayor detalle en la Sección 3), son los siguientes:

1. Estudio de estructura interna de juegos serios, que permita discernir los elementos y mecánicas que están diseñados para facilitar el aprendizaje.
2. Creación de un modelo de datos sobre juegos serios que permita capturar las interacciones que reflejan situaciones relevantes para el proceso de aprendizaje.
3. Implementación de una plataforma software que soporte la recepción del modelo de datos.
4. Desarrollo e implementación de juegos que validen tanto la metodología de diseño, como el modelo de captura de datos.
5. Validación de las aportaciones del modelo para el análisis del proceso de aprendizaje.
6. Generalización del diseño de juegos serios a una metodología de desarrollo que considere el análisis del aprendizaje.

7. Generalización del modelado de datos a un estándar educativo que permite su adopción y extensión dentro del mundo de los juegos serios.

Se escoge esta aproximación tras identificar nuevas líneas de investigación aún por explorar, añadido a la experiencia del autor y director de tesis dentro del mundo de los juegos serios, así como la del resto del grupo de investigación que arropa esta tesis (del Blanco, Marchiori, Torrente, Martínez-Ortiz, & Fernández-Manjón, 2013; Manero, Fernández-Vara, & Fernández-Manjón, 2013b; Marchiori et al., n.d.; Moreno-Ger, Burgos, Martínez-Ortiz, Sierra, & Fernández-Manjón, 2008; Torrente, Del Blanco, Marchiori, Moreno-Ger, & Fernández-Manjón, 2010; Torrente, Del Blanco, Moreno-Ger, & Fernández-Manjón, 2012). Adicionalmente, dentro de este contexto existirán múltiples oportunidades de aplicar los resultados de la tesis a otros trabajos y juegos desarrollados dentro del marco del grupo de investigación, lo que permitirá extender y validar resultados.

Finalmente, aunque el planteamiento de esta tesis nace del trabajo realizado en la Tesis de Máster del autor, durante su desarrollado los primeros resultandos empiezan a verse integrados dentro de proyectos europeos de investigación, tales como FP7 NoE GALA (Hauge et al., 2014a), H2020 RAGE (Hollins, Westera, & Manero, 2015) o H2020 BEACONING. Estas integraciones permiten una validación externa de los resultados obtenidos tanto en los productos realizados en el grupo de investigación e-UCM como en los que se han reutilizado por terceros en dichos proyectos y que ha supuesto como un impulso para su desarrollo (y sobre todo para su maduración en productos finales).

Capítulo 2. Estudio del dominio

Como se ha expuesto en el capítulo anterior, el trabajo de esta tesis se centra en la aplicación de analíticas de aprendizaje sobre juegos serios. En consecuencia, en este capítulo se hace un estudio del dominio que enmarca los resultados que se expondrán en los capítulos 4 y 5.

Este estudio comienza analizando las diferentes disciplinas que han dedicado esfuerzos al análisis de los diferentes tipos de datos que pueden capturarse en instituciones educativas, haciendo especial énfasis en las analíticas de aprendizaje. A continuación, se hace un repaso de los juegos serios, sus características, sus beneficios y sus criticismos, así como de cómo se relaciona su diseño con los procesos de aprendizaje. Después, se repasan los diferentes tipos de análisis que se han hecho sobre videojuegos. Finalmente, se repasan las formalizaciones que, a través de estándares, han intentado capturar los diferentes procesos de aprendizaje en contenidos interactivos.

2.1. Análisis de datos educativos

Como explicamos en el Capítulo 1, son varias las disciplinas que a lo largo del tiempo han utilizado datos procedentes de instituciones educativas para realizar análisis. Aunque sus nombres son diferentes, sus características, actores y beneficios son comunes. Lo cierto es que a veces la línea que separa estas disciplinas es muy fina, y muchos de los trabajos pueden pertenecer a más de una categoría, dependiendo de la definición a la que nos atengamos.

Todas las disciplinas de análisis de datos educativos plantean un proceso de análisis (que toma prestado de las Inteligencia de Negocio (Williams & Williams, 2003)) que consta de cinco pasos:

- **Captura**: en este paso se recolectan todos los datos que después serán analizados. Aquí hay que tener en cuenta el propio proceso de captura (qué datos se van a capturar y cómo), así como consideraciones más técnicas sobre el almacenaje y la granularidad de los datos.
- **Informes y análisis**: una vez los datos han sido capturados, es necesario que expertos puedan analizar y visualizar el contenido de esos datos. En algunos casos el análisis y la visualización se consideran pasos separados.
- **Predecir**: una vez los análisis empiezan a establecer patrones, es el momento de crear modelos que los representen, y que puedan ser utilizados para predecir resultados futuros.
- **Actuar**: el objetivo de las predicciones es proveer a instructores y administradores de información que les permita actuar, por ejemplo, corrigiendo problemas

detectados por los modelos de predicción antes de que se conviertan en irreversibles.

- **Refinar**: los procesos de análisis siempre siguen un ciclo iterativo, y necesitan ser refinados y ajustados con cada nueva pieza de información que se va recibiendo para mantener su aplicabilidad.

Este es un proceso cuyos resultados beneficia a distintos actores dentro del mundo académico:

- **Profesores e instructores**: tienen una visión más detallada de los procesos educativos, y cuentan con herramientas y predicciones que les permiten actuar en beneficio de los estudiantes.
- **Estudiantes**: al contar sus profesores con más herramientas para su instrucción, reciben una atención más personalizada. A veces pueden tener información anticipada que les permite, por ejemplo, poner más esfuerzo en aquellas tareas donde su rendimiento es mejorable.
- **Instituciones y personal administrativo**: tienen métricas e informes basados en datos de confianza, que les permiten tomar decisiones en pos de un mejor funcionamiento global de la institución educativa.

Aunque sus potenciales beneficios son múltiples, es un proceso que no está exento de problemas a resolver. Uno de los más importantes es sobre las implicaciones éticas y legales de monitorizar los resultados y avances pormenorizados de los alumnos, y sobre cómo y cuándo éstos deben participar en el proceso (Slade & Prinsloo, 2013). También existen riesgos con la generación de modelos de predicción que, en ocasiones, pueden fallar en la captura de patrones atípicos acabando con la clasificación errónea de alumnos. Dentro de esta categoría, está contemplada una cierta definición de "éxito", que puede no ser la misma para todos los alumnos.

A continuación, se enumeran las diferentes disciplinas que emergen a partir de estos procesos dentro del mundo educativo, enumerando algunas de sus características y ejemplos.

2.1.1. Analíticas Académicas

Como explicamos en la introducción, una de las primeras disciplinas que comenzó a analizar datos académicos con el objetivo de comprender y mejorar el funcionamiento de los procesos educativos fueron las Analíticas Académicas (Campbell et al., 2007).

La Figura 1 muestra los tipos de datos que esta disciplina usa para sus análisis. Destacan los datos demográficos y los resultados académicos. Como se puede observar por el tipo de datos, la disciplina se centraba en información que tradicionalmente se almacenaban en los diferentes sistemas de información de las instituciones académicas.

En un informe elaborado por EDUCAUSE en 2005 (Goldstein, 2005), se entrevistó a un total de 380 instituciones sobre los sistemas de información con los que contaban, y sobre las analíticas que realizaban sobre ellos. El estudio llegó a la conclusión de que, atendiendo a los sistemas de información a su disposición, las diferentes instituciones se dividían en 3 categorías:

- Instituciones que contaban únicamente con sistemas de transacciones.
- Instituciones con sistemas de transacciones que contaban con pequeñas herramientas para la generación de análisis e informes.
- Instituciones con sistemas empresariales completos de análisis, visualizaciones y alertas.

Type of Data	Variable	Source	Frequency of Update
Demographic	Age	SIS	Once
	Ethnicity	SIS	Once
	First-generation college student	SIS	Once
Academic ability	HS rank	SIS	Once
	HS GPA	SIS	Once
	HS coursework (number of math, science, English courses)	SIS	Once
	Placement test results	SIS	Once
	Standardized test scores	SIS	Once
Academic performance	College GPA	SIS	Once per term
Academic history	Initial major	SIS	Rarely
	Credit hours completed	SIS	Once per term
	Current major	SIS	Rarely
	Previous coursework	SIS	Once per term
Financial	Amount of aid	Financial system	Once per term
	Work study student	Financial system	Once per term
Participation information	Help desks	Varies	Varies
	Orientation activities	Varies	Varies
	Student organizations	Varies	Varies
	Supplemental instruction	Varies	Varies
Academic effort	CMS usage	CMS	Varies
	Computer laboratory usage	Varies	Varies
	Electronic reserve usage	Varies	Varies
Institutional information	Course size	SIS	Once per term
	Historic student information (previous grade distribution, number of withdrawals)	Varies	Varies

Figura 1. Datos utilizados por Analíticas Académicas (Campbell et al., 2007)

El estudio encontró que el 48% aún se encontraba en la primera categoría. Adicionalmente, incluyendo a aquellas instituciones que contaban con sistemas más avanzados, la mayoría del personal que hacía uso de la información obtenida de estos sistemas pertenecía a los departamentos financieros y de admisión.

El análisis de estos datos fue el inicio de muchas instituciones en el viaje de las analíticas de aprendizaje. Poco a poco se fueron implantando nuevos sistemas que dieron lugar a nuevos tipos de análisis.

2.1.2. Minería de Datos Educativos

La aparición de la Minería de Datos Educativos (R. S. J. d. Baker & Yacef, 2009) se debe a dos factores que confluyen en el tiempo: 1) la proliferación de las técnicas de minería de datos; y 2) la aparición de nuevos sistemas con información sobre los procesos educativos (por ejemplo, los sistemas de gestión de aprendizaje).

Al contrario que las Analíticas Académicas, esta disciplina se nutre de datos más directos del proceso del aprendizaje de los alumnos. Su principal objetivo es desarrollar métodos que permitan entender mejor los procesos educativos y su desarrollo. Como objetivos secundarios más concretos, esta disciplina plantea los siguientes (Liñán & Pérez, 2015; Siemens & Baker, 2012):

- Automatizar procesos de descubrimiento de patrones, tratando de reducir la intervención humana.
- Focalizar el área de aplicación y procedencia de la información, centrándose en procesos de aprendizaje concretos.
- Crear modelos de estudiantes que permitan adaptación automatizada.

Diferentes trabajos han recolectado las múltiples aplicaciones de estas disciplinas (R. S. J. d. Baker & Yacef, 2009; C Romero & Ventura, 2010; S. Ventura & Romero, 2007), como proveer de retroalimentación a instructores, generar recomendaciones para los estudiantes o detectar patrones no deseados en los comportamientos de los estudiantes.

2.1.3. Analíticas de aprendizaje

Poco después de la aparición de la Minería de Datos Educativa apareció una nueva disciplina relacionada denominada como Analíticas de Aprendizaje (Long & Siemens, 2011). Aun compartiendo muchos puntos en común con la anterior (su objetivo sigue siendo el análisis y la mejora de los procesos educativos), cuenta con algunas diferencias (Liñán & Pérez, 2015; Siemens & Baker, 2012):

- Busca descubrir patrones de conocimiento, cuya interpretación y uso se deja a interpretación humana.
- Aunque se centra en procesos más concretos que las Analíticas Académicas, intenta entender los sistemas en su completitud.

- En vez de modelar estudiantes, se centra más en predicción de resultados así como en sistemas de alarmas que permiten intervenciones.

A pesar de estas diferencias, la superposición de estas tres disciplinas es constante, y como reflexionan varios investigadores (R. S. Baker & Inventado, 2014), las diferencias entre ambas se deben más al uso y las tendencias dentro del mundo de la investigación que a motivos filosóficos que las diferencian claramente.

2.2. Áreas de aplicación de análisis educativos

En esta sección repasamos las diferentes áreas de aplicación de las analíticas de aprendizaje. Se pretende explorar las diferentes áreas dónde se han utilizado estas técnicas, con el objetivo de enmarcar el uso práctico que se les ha dado hasta el momento. Primero presentamos las primeras aproximaciones iniciales, basadas en procesos más manuales, para después pasar a áreas en las que existe un mayor grado automatización.

2.2.1. Educación tradicional

Esta área de aplicación concierne a los primeros estudios que empezaron a realizarse sobre análisis de aprendizaje, y que contaban, en la mayoría de los casos, con un proceso manual de recolección de datos. La mayoría de trabajos en esta área se basan en los datos definidos por las Analíticas Académicas, haciendo especial hincapié en datos académicos y, más secundariamente, en datos demográficos. La mayoría de trabajos se elaboran dentro del contexto universitario.

Por ejemplo, entre los primeros estudios podemos encontrar el trabajo de Fausett y Elwasif, donde recopilaban las respuestas que alumnos hacían en exámenes escritos, para después analizarlas haciendo uso de redes neuronales con el objetivo de predecir la nota final del curso (Fausett & Elwasif, 1994).

La predicción del rendimiento académico a partir de resultados pasados se encuentra entre los principales objetivos de estas primeras aproximaciones. Por ejemplo, Golding & Donaldson realizaron un estudio para comprobar la validez de los exámenes de admisión de la Universidad Tecnológica de Jamaica (Golding & Donaldson, 2006). En este estudio, a través de un proceso manual de recolección de datos, intentaban establecer correlaciones entre los resultados en el examen admisión a un curso de ingeniería de un alumno con su rendimiento final en dicho curso.

En una línea similar, Hien & Haddawy intentaron crear modelos que predijeran, a partir del expediente académico de alumno, las probabilidades de completar con éxito sus estudios académicos dentro de la universidad (Hien & Haddawy, 2007). En este caso, su objetivo era deshacerse por completo de los exámenes de admisión, y basar la aceptación de sus alumnos únicamente en los logros académicos conseguidos hasta el momento.

En otras universidades, los resultados académicos se han utilizado para predecir la posible demanda de cursos, analizando las tendencias de cursos anteriores (Hsia, Shie, & Chen, 2008).

Otros trabajos utilizaban datos más extensos para crear sus modelos. En (Dekker, Pechenizkiy, & Vleeshouwers, 2009) usan como entrada datos demográficos y resultados académicos pasados, e intentan correlacionarlos con los datos de éxito de un curso de ingeniería. Como resultado, obtienen un modelo capaz de prever la probabilidad de éxito de un alumno a partir de sus datos demográficos.

La variedad de estos trabajos comenzó a crecer conforme empezaron a aparecer nuevas fuentes de datos.

2.2.2. Tutores inteligentes

Los tutores inteligentes son uno de los primeros sistemas educativos que trataron de modelar el conocimiento de los estudiantes y adaptarse a su progreso, intentando imitar la figura de un tutor humano (Anderson, Boyle, & Reiser, 1985). En su mayoría, son sistemas expertos diseñados para la enseñanza de un tema concreto, y aunque sus formas son múltiples, la mayoría cuenta con los siguientes componentes (Nwana, 1990):

- El módulo experto, dónde reside el conocimiento a impartir y las reglas que aplican de cara a la interacción del estudiante
- El módulo del estudiante, dónde se captura una representación dinámica del conocimiento del estudiante
- El módulo tutor, que regula la interacción entre el estudiante y el conocimiento
- Y la interfaz de usuario, que controla la interacción del estudiante con todo el sistema

La interacción producida por los estudiantes con estos sistemas permite la realización de unos primeros análisis del aprendizaje más centrados en el proceso de enseñanza y no sólo en resultados finales. Aun así, la mayoría de trabajos en esta área se centran en el modelado de estudiantes.

Por ejemplo, Chang et al. recogen el cambio del modelo de conocimiento de un estudiante durante el uso un tutor inteligente (Chang, Beck, Mostow, & Corbett, 2006). Como entradas utilizan el estado inicial del conocimiento del estudiante, las intervenciones del tutor (si provee de ayuda al estudiante o no), y el resultado final la adquisición de conocimiento. El objetivo es obtener un modelo de estudiante basado en el uso de Redes Bayesianas.

Yudelson aplica técnicas de minería de datos sobre los datos de un tutor inteligente aplicado a medicina (Yudelson et al., 2005). En su investigación descubre patrones de aprendizaje que le permite clasificar a sus usuarios como diferentes tipos de estudiantes:

los que aprenden a partir de pistas, los que aprenden a partir de fallos y los que tienen una aproximación más mixta.

Hämäläinen & Vinni estudian un uso más avanzado de los datos recolectados por estos sistemas inteligentes, proponiendo usar los resultados de los análisis para mejorar los propios tutores inteligentes (Hämäläinen & Vinni, 2006).

2.2.3. Sistemas de Gestión de Aprendizaje

Los Sistemas de Gestión de Aprendizaje (SGA) suponen una revolución en muchas instituciones educativas. Plataformas como Moodle (Dougiamas & Taylor, 2003) o Sakai (Farmer & Dolphin, 2005) permiten a instructores y estudiantes pasar de una relación estrictamente presencial a estar comunicados a través de una plataforma on-line capaz proveer contenidos educativos, foros de discusión e incluso la posibilidad de realizar exámenes on-line.

Las analíticas de aprendizaje pueden ir un paso más allá gracias a estas plataformas: mientras que las investigaciones anteriores se basaban en resultados individuales de alumnos, en estos sistemas se cuentan con nuevos tipos de datos, como las estadísticas relacionadas con el uso de los contenidos educativos alojados en el sistema o las propias interacciones que los alumnos realizan entre ellos. Toda esta nueva información empieza a ser analizada con un nuevo conjunto de técnicas.

2.2.3.1. Analíticas Web

Las Analíticas Web analizan el uso de recursos web por parte de los usuarios que navegan por Internet (Hassler, 2010). Utilizan como principal fuente de datos los accesos que los usuarios realizan a cada una de las páginas que componen una web. También guardan tiempos empleados en cada página, botones pulsados e interacciones de especial relevancia (por ejemplo, las veces que un usuario añade un producto al carrito de un comercio electrónico). Las técnicas de esta disciplina tienen especial interés para el comercio electrónico, donde son capaces de prever tendencias y estimar beneficios. Dada la naturaleza web de los SGA, muchas de estas técnicas también pueden ser empleadas en el análisis del aprendizaje.

Por ejemplo, Baruque desarrolla una herramienta para analizar el tráfico web generado por estudiantes accediendo a un sistema Moodle (Baruque, Amaral, Barcellos, Freitas, & Longo, 2007). Su objetivo es analizar las tendencias y efectividad de los contenidos educativos basándose en la frecuencia de acceso de los estudiantes.

Chanchary va un poco más allá y retoma el objetivo clásico de las analíticas de aprendizaje: predecir resultados académicos. Para ello, trata de relacionar la actividad web de los alumnos en los SGA con su rendimiento académico (Chanchary, Haque, & Khalid, 2008), llegando a resultados positivos.

2.2.3.2. Interacciones sociales

Una de las herramientas más utilizadas dentro de los SGAs son los foros de discusión. Estos lugares suponen puntos de encuentros virtuales donde profesores y alumnos pueden intercambiar, a través de mensajes escritos, opiniones sobre contenidos educativos. Han sido muchos los investigadores que han realizado estudios sobre las interacciones que se realizan en estos foros, basándose en las teorías de análisis de redes sociales (Scott, 2011), cuyas técnicas se han visto enormemente desarrolladas con la aparición de plataformas de gran impacto social como Facebook.

Existen investigaciones que han estudiado los patrones de uso de los alumnos en este tipo de foros (Burr & Spennemann, 2004), buscando contrastar hipótesis relacionadas con el aprendizaje "en cualquier sitio / en cualquier momento", descubriendo, tras analizar más de 200 foros on-line, que la mayoría de interacciones suceden durante el año escolar y en horario lectivo.

Otros investigadores han desarrollado herramientas específicamente diseñada para analizar las interacciones educativas que se dan en estos foros de debate. Por ejemplo, SNAPP es una herramienta capaz de analizar la evolución de comunidades on-line a partir de las interacciones que mantienen cada uno de sus miembros (Dawson, Bakharia, & Heathcote, 2010). Entre sus principales objetivos se encuentra identificar a aquellos alumnos que están más desconectados del resto de la comunidad educativa y ofrecer a los profesores esta información para que puedan actuar en consecuencia.

En esta área también se ha tratado de predecir resultados académicos. Por ejemplo, Romero (Cristóbal Romero, López, Luna, & Ventura, 2013) utiliza técnicas de minería de datos sobre redes sociales para predecir la nota de los estudiantes en un curso de ingeniería, contando su número de participaciones en foros y combinando con datos académicos. En este trabajo establece diferentes conjuntos datos y trata de establecer cuáles son los más óptimos a la hora de predecir resultados.

2.2.3.3. Análisis de textos

En el apartado anterior se presentaban trabajos que utilizan como dato de entrada las conexiones y participaciones de estudiantes en foros de debate. Sin embargo, existe también otra tendencia en la literatura científica que se dedica a analizar los textos que escriben los alumnos en estos foros de debate para obtener conclusiones sobre su rendimiento académico.

Dringus propone técnicas de análisis de textos para analizar la evolución de discusiones dentro de un mismo hilo en un foro educativo (Dringus & Ellis, 2005). El objetivo no es únicamente contabilizar la participación de los estudiantes en el debate, sino extraer conclusiones sobre la evolución del mismo y proveer al instructor de pistas sobre lo que está sucediendo de manera automática.

Kim presenta un análisis de "actos de discusión" (J. Kim, Chern, Feng, Shaw, & Hovy, 2006), que analiza el contenido de los mensajes de estudiantes en busca de patrones de discurso, cuyo objetivo es clasificar discusiones de estudiantes en las que puedan existir preguntas sin contestar o discusiones.

2.2.3.3. Cuestionarios

Además de contenidos educativos y foros de discusión, los SGAs proveen de un método de evaluación de alumnos prestado directamente de la educación tradicional: cuestionarios y/o exámenes. Dentro de los SGAs, los cuestionarios son las herramientas más directas para la evaluación de aprendizaje de los alumnos. Sin embargo, sus usos potenciales pueden variar. La capacidad de estos cuestionarios de auto-corregirse automatiza la labor del profesor. Los resultados de estos cuestionarios no siempre se ven reflejados en una evaluación directa del alumno; en muchas ocasiones, los alumnos las utilizan como herramientas de auto-evaluación. Su uso como fuente para analíticas de aprendizaje es variado.

Una de las aplicaciones más básica es ofrecer los propios resultados de los cuestionarios de manera individual y agregada a los profesores dentro del propio SGAs. Por ejemplo, Moodle ofrece varios plug-ins en los que se hace uso de estos resultados (Diego et al., 2012). Sin embargo, algunos investigadores han intentado explorar caminos diferentes haciendo uso de estos datos.

Bravo (Bravo & Ortigosa, 2009) hace énfasis en los esfuerzos que a veces se llevan a cabo para el desarrollo de materiales educativos, y que a veces no están acompañados de una evaluación que valide su utilidad. En su trabajo propone el uso de cuestionarios como instrumento de validación.

Chen propone el análisis de respuestas en cuestionarios a través de reglas asociativas (Y.-L. Chen & Weng, 2009). El objetivo es crear un algoritmo capaz de encontrar posibles relaciones entre los patrones de acierto y error que se observan en cada una de las preguntas.

Finalmente, como punto de vista alternativo, Hernández alerta de las posibilidades de falsear resultados o hacer trampas en estos cuestionarios (Hernández, Ochoa, Muñoz, & Burlaka, 2006). En su trabajo propone un método que analiza accesos a los exámenes, patrones de respuestas y reintentos de los estudiantes para detectar y prevenir a aquellos alumnos que pueden estar copiando sus respuestas de otros.

2.2.4. Cursos On-line Masivos Abiertos

Los cursos on-line masivos (o MOOCs) vieron su auge a partir del año 2012 (Daniel, 2012). Estos cursos ofrecen contenidos educativos a grandes números de personas que pueden encontrarse en cualquier parte del mundo. Los MOOCs basan su contenido principalmente en vídeo (normalmente, un instructor explicando contenido educativo directamente a cámara, a veces junto a una presentación de diapositivas como apoyo), acompañado de recursos adicionales como documentación, foros de discusión o ejercicios interactivos. Los contenidos ofrecidos suelen centrarse en una materia especializada con el objetivo de ser auto-contenidos.

Una de las mayores ventajas de este tipo de cursos es la gran cantidad de datos de aprendizaje de la que disponen. Mientras que cursos universitarios normales dentro de SGAs pueden llegar, como máximo, a cientos de alumnos, los MOOCs tienen la capacidad de llegar a miles o incluso a cientos de miles. La masividad de estos datos permite hacer análisis que requerían grandes cantidades de datos para poder ofrecer modelos de confianza.

Sin embargo, el cuerpo de datos que en ocasiones se recoge en estos cursos dista de ser perfecto para su análisis. Uno de los fenómenos más comunes dentro de los MOOCs es la proporción de gente que los abandona, y algunos investigadores han estudiado el efecto que pueden producir en las analíticas de aprendizaje (Clow & Doug, 2013).

Otra característica de los MOOCs es que la demografía de los alumnos que participan en ellos es mucho más variada que la se puede encontrar en una carrera universitaria al uso. Algunos estudios proponen la revisión de modelos de adaptación que sean más sensibles a estas diferencias (Daradoumis, Bassi, Xhafa, & Caballe, 2013).

Finalmente, muchos de los análisis que ya han sido presentados a lo largo de esta sección pueden aplicarse también a los MOOCs. Por ejemplo, existen varios estudios que se centran en el análisis de discusiones en foros de estos cursos (Ezen-Can, Boyer, Kellogg, & Booth, 2015).

2.3. Juegos serios y aprendizaje

El mercado de los videojuegos no ha dejado de crecer en los últimos años[1], llegando a superar en beneficios a industrias tan grandes como la del cine[2]. Uno de los factores que ha determinado este crecimiento exponencial es su penetración en la sociedad y la gran diversificación que los videojuegos han experimentado en la última década. Mientras que en sus inicios se les consideraba como un producto que sólo consumía un público joven a través de dispositivos de entretenimiento ad-hoc (*Atari, Nintendo Enterntaiment System, Master System…*), hoy en día podemos encontrar videojuegos en todo tipo de dispositivos (videoconsolas, ordenadores, tablets, dispositivos móviles…) que son jugados por todo tipo de público. Según el informe ESA de 2015 (Entertainment Software Assotiation, 2015), la

[1] https://newzoo.com/insights/articles/global-games-market-reaches-99-6-billion-2016-mobile-generating-37/

[2] http://www.eleconomista.es/tecnologia-videojuegos/noticias/7608246/06/16/Los-videojuegos-facturan-mas-de-mil-millones-de-euros-en-Espana-en-2015.html

edad del jugador medio es de 35 años y el número de hombres y mujeres que juegan a videojuegos está prácticamente igualado (Figura 2).

Adicionalmente, esta diversificación ha supuesto que los videojuegos se hayan adentrado con éxito en otros campos más allá del puro entretenimiento. Los videojuegos han sido utilizados en el terreno de la publicidad o la educación, dando lugar al término juego serio (Djaouti et al., 2011).

Definimos juego serio como todo aquel videojuego cuyo objetivo principal no es el entretenimiento. En el caso de juegos serios para publicidad, se utilizan para dar a conocer algún producto. En el caso de la educación, los juegos serios se utilizan como herramientas de simulación para entrenar habilidades en entornos seguros, o para presentar contenido educativo con un mayor grado de interactividad. A continuación, abordamos algunas de las características de los juegos serios como herramientas educativas, y las características que los convierten en buenos candidatos para la evaluación de aprendizaje.

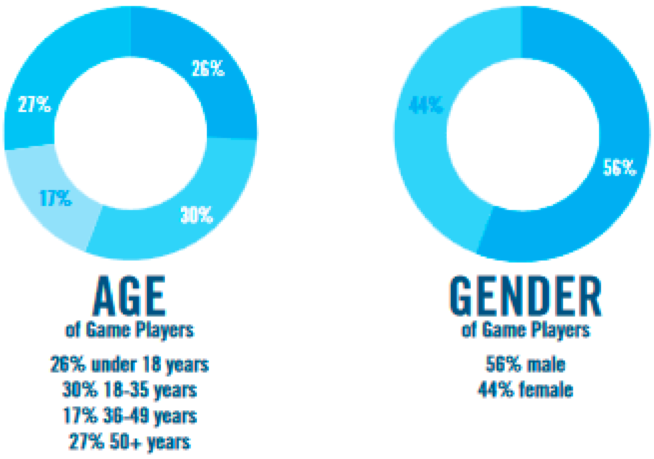


Figura 2. Proporción de jugadores por edad y género según el informe ESA 2015

2.3.1. Juegos serios como herramientas educativas

La característica diferenciadora de un juego serio respecto a cualquier otro tipo de contenido educativo es la capacidad de los alumnos de interactuar con ellos de forma reactiva (DeKanter, 2004). Esta capacidad permite a los alumnos crear, construir y experimentar en un entorno seguro, algo fundamental en todos los procesos de aprendizaje eficaces (J.P. Gee, 2005; James Paul Gee, 2003).

Esta característica básica se une al resto de bondades que los videojuegos muestran de manera intrínseca como, por ejemplo, su capacidad para motivar y mantener el interés de sus jugadores (Garris, Ahlers, & Driskell, 2002). Además, un videojuego bien diseñado es capaz de mantener captados o interesados a sus jugadores a través de lo que se conoce como “zona de flujo” (J. Chen, 2007). Éste es un proceso por el cual los videojuegos van presentando objetivos lo suficientemente desafiantes como para que el jugador quiera seguir jugando, pero lo suficientemente asequibles como para que no abandone debido a la frustración, manteniéndole así en la “zona de flujo”. Esta experiencia de la zona flujo es extrapolable a los procesos de aprendizaje.

Partiendo de estas ideas, han sido muchas las investigaciones que han querido validar los juegos serios como herramientas educativas en diferentes dominios y niveles educativos (Boyle et al., 2016; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). Por ejemplo, se han usado con resultados satisfactorios en el campo de la medicina (Baranowski et al., 2008; Berkovsky, Coombe, Freyne, Bhandari, & Baghaei, 2010; Bernardini, Porayska-Pomsta, & Smith, 2014; Evans et al., 2015), la ingeniería y la física (Hauge & Riedel, 2012; Price, 2007; Sancho, Torrente, & Fernández-Manjón, 2009; van der Graaf, Segers, & Verhoeven, 2016), enseñanza de programación (Ibrahim, Semarak, Lumpur, & Jaafar, 2010; Malliarakis et al., 2014; Resnick et al., 2009) o la literatura (Manero et al., 2013b).

A pesar de sus probadas aplicaciones, aún quedan algunos problemas por resolver. Hays es uno de los primeros en estudiar de modo más sistemático y desafiar algunas de observaciones sobre las bondades de los juegos serios a través una revisión de literatura que repasa 274 artículos (Hays, 2005). Entre sus conclusiones destaca que, aunque hay pruebas empíricas de la capacidad de los juegos serios para enseñar, las metodologías usadas en algunos experimentos, tanto para diseñar el juego como para probar su eficacia, son inconsistentes y en algunos casos incluso erróneas. Recomienda a los diseñadores de juegos que trabajen codo con codo con instructores y profesores, teniendo especial cuidado a la hora de definir las mecánicas de juego y manteniendo siempre presentes los objetivos instruccionales del juego.

Este problema aún no ha sido resuelto por completo, aunque poco a poco han ido apareciendo investigaciones que tratan de formalizar metodologías para mejorar la evaluación de juegos serios y otro contenidos digitales basados en juegos (All, Nuñez Castellar, & Van Looy, 2015, 2016).

Otro problema a resolver es la propia evaluación de la efectividad de los juegos serios. Tradicionalmente, la gran mayoría de investigaciones sobre juegos serios han seguido una aproximación experimental tradicional en su desarrollo (Calderón & Ruiz, 2015). Los experimentos son de corta duración, y el impacto en los alumnos suele medirse a partir de la comparación de pre-tests y post-tests. En ocasiones, los resultados de estos cuestionarios se ven acompañados de una sesión de debate, en la que alumnos e instructores comparten las experiencias vividas con el juego (Arafeh, Hansen, & Nichols, 2010). Algunos investigadores han apuntado que esta herramienta es poco escalable y supone problemas para los instructores (Egenfeldt-Nielsen, 2004).

En conclusión, quedan como debates abiertos la relación del aprendizaje con el diseño de los juegos serios, así como los métodos formales para su evaluación.

2.3.2. Aprendizaje y evaluación educativa

Los procesos de aprendizaje del ser humano y las teorías pedagógicas que existen detrás ellos siempre han estado ligados a la trayectoria de los juegos serios (Amory, Naicker, Vincent, & Adams, 1999; Yusoff, Crowder, Gilbert, & Wills, 2009). Para diseñar un juego capaz de enseñar es necesario tener presente algunas de las teorías relacionadas con el aprendizaje. Particularmente, la capacidad de los juegos serios de recrear mundos envolventes y reactivos en los que los estudiantes pueden realizar acciones dentro un conjunto limitado de normas, hace que sean herramientas especialmente cualificadas para la enseñanza de habilidades (Graafland, Schraagen, & Schijven, 2012).

El aprendizaje de habilidades ha sido modelado en diferentes trabajos. Fitts (Fitts, 1964) lo define como un proceso en el que el aprendiz pasa por tres fases:

- **Estado cognitivo**: el aprendiz posee en conocimiento inicial que le permite decodificar la suficiente información sobre la habilidad, obteniendo en su realización unos resultados preliminares aproximados.
- **Estado asociativo**: el aprendiz posee la capacidad suficiente como para practicar la habilidad con resultados satisfactorios, y está empezando a pulir algunas de sus componentes.
- **Estado autónomo**: el aprendiz domina la habilidad con soltura y cada nueva práctica mejora su efectividad.

Estos estados han ido siendo ampliados por nuevas teorías. Dreyfuss (Dreyfuss & Dreyfus, 1980) propone pasar de tres a cinco estados. En su teoría, añade dos estados intermedios entre el estado cognitivo y asociativo, y el asociativo y autónomo, quedando los estados en principiante, capaz, competente, experto y maestro.

La taxonomía de Bloom (Krathwohl, 2002) es un modelo que relaciona los procesos cognitivos que se dan durante el aprendizaje con los tipos de conocimiento que se van generando durante el mismo. En la intersección entre procesos cognitivos y tipos de conocimiento se encuentra la acción que relaciona a ambos y que el aprendiz debe dominar para superar el peldaño. La Figura 3 muestra una ilustración que enumera cada una de las dimensiones y explica su significado[3].

Acercando algunas de estas teorías a cuestiones más prácticas relacionadas con el aprendizaje on-line, Scalise (Scalise & Gifford, 2006) crea una tabla (Figura 4) que clasifica los tipos de ejercicios que pueden encontrarse en cursos de e-learning en cuanto a su dificultad y el grado de libertad que el alumno tiene para resolverlos. Cada uno de las actividades se clasifica bajo una acción que referencia a las acciones definidas por la taxonomía de Bloom.

[3] http://www.celt.iastate.edu/wp-content/uploads/2015/09/RevisedBloomsHandout-1.pdf

Aun siendo esta taxonomía una visión mucho más detallada de los pasos que se dan en procesos formativos, sigue siendo una extensión fundamentada en la teorías propuestas con anterioridad. Idealmente, los diseñadores de juegos serios deben tener en cuenta las acciones aquí propuestas a la hora de diseñar las mecánicas de juego (Arnab et al., 2015).

Un diseño centrado en el proceso de aprendizaje no sólo convierte a los juegos serios en herramientas educativas más efectivas, sino que permite además que sean evaluadas con mayor facilidad.

La evaluación en los videojuegos educativos puede perseguir diferentes resultados. La gran mayoría pretende evaluar resultados de aprendizaje (es decir, si los estudiantes aprendieron jugando con el juego o no), pero en ocasiones, la evaluación pretende medir otros factores como la eficiencia (por ejemplo, si un juego es más rápido que instructor enseñando una habilidad) o la motivación (si los jugadores tuvieron una experiencia positiva con el juego) (All et al., 2015).

Figura 3. Gráfico de la Taxonomía de Bloom extendida

El diseño de evaluación centrado en evidencias (ECD, de sus siglas en inglés Evidence-centered assessgment design) (Mislevy, Almond, & Lukas, 2003) basa la evaluación de resultados en las tareas que realizan los alumnos durante una actividad educativa. A

Most Constrained → *Least* Constrained

Fully Selected — *Intermediate Constraint Item Types* — *Fully Constructed*

Less Complex ↓ More Complex

1. Multiple Choice	2. Selection/ Identification	3. Reordering/ Rearrangement	4. Substitution/ Correction	5. Completion	6. Construction	7. Presentation/ Portfolio
1A. *True/False* (Haladyna, 1994c, p.54)	2A. *Multiple True/False* (Haladyna, 1994c, p.58)	3A. *Matching* (Osterlind, 1998, p.234; Haladyna, 1994c, p.50)	4A. *Interlinear* (Haladyna, 1994c, p.65)	5A. *Single Numerical Constructed* (Parshall et al, 2002, p. 87)	6A. *Open-Ended Multiple Choice* (Haladyna, 1994c, p.49)	7A. *Project* (Bennett, 1993, p.4)
1B. *Alternate Choice* (Haladyna, 1994c, p.53)	2B. *Yes/No with Explanation* (McDonald, 2002, p.110)	3B. *Categorizing* (Bennett, 1993, p.44)	4B. *Sore-Finger* (Haladyna, 1994c, p.67)	5B. *Short-Answer & Sentence Completion* (Osterlind, 1998, p.237)	6B. *Figural Constructed Response* (Parshall et al, 2002, p.87)	7B. *Demonstration, Experiment, Performance* (Bennett, 1993, p.45)
1C. *Conventional or Standard Multiple Choice* (Haladyna, 1994c, p.47)	2C. *Multiple Answer* (Parshall et al, 2002, p.2; Haladyna, 1994c, p.60)	3C. *Ranking & Sequencing* (Parshall et al, 2002, p.2)	4C. *Limited Figural Drawing* (Bennett, 1993, p.44)	5C. *Cloze-Procedure* (Osterlind, 1998, p.242)	6C. *Concept Map* (Shavelson, R. J., 2001; Chung & Baker, 1997)	7C. *Discussion, Interview* (Bennett, 1993, p.45)
1D. *Multiple Choice with New Media Distractors* (Parshall et al, 2002, p.87)	2D. *Complex Multiple Choice* (Haladyna, 1994c, p.57)	3D. *Assembling Proof* (Bennett, 1993, p.44)	4D. *Bug/Fault Correction* (Bennett, 1993, p.44)	5D. *Matrix Completion* (Embretson, S, 2002, p. 225)	6D. *Essay* (Page et al, 1995, 561-565) & *Automated Editing* (Breland et al, 2001, pp.1-64)	7D. *Diagnosis, Teaching* (Bennett, 1993, p.4)

Figura 4. Recopilación del tipo de actividades que se realizan en cursos on-line, categorizando su acción cognitiva y su dificultad

través de una instrumentalización adecuada, las interacciones de los aprendices deben servir como evidencia para la evaluación de su aprendizaje. ECD pide a los diseñadores de contenidos educativos que traten de responder las siguientes preguntas antes de comenzar con el desarrollo de cualquier contenido educativo (Shute, Leighton, Jang, & Chu, 2016):

1. ¿Qué conocimiento o habilidad quiere evaluarse?
2. ¿Qué observables medibles pueden extraerse de la actividad?
3. ¿Qué criterios y reglas se evalúan los observables?
4. ¿Qué tipo de tareas se deben desarrollar durante la actividad para obtener los observables deseados?

Es decir, los diseñadores de juegos serios deben considerar el flujo completo de uso del juego. No solo debe ser capaz de enseñar, sino que además deber estar centrado en la evaluación de los alumnos, desafiando a los estudiantes a que prueben y contrasten, con sus acciones dentro del juego, la adquisición del conocimiento impartido.

2.3.3. Diseño de juegos serios

El diseño de un juego serio es fundamental para que cumpla sus objetivos como herramienta educativa. Siendo un campo relativamente joven, muchos de los juegos serios que se han ido desarrollando han carecido del diseño adecuado para cumplir sus objetivos. Poco a poco, los investigadores han ido encontrando algunos de los elementos de diseño que los hacen más efectivos.

Dickey (Michele D. Dickey, 2006a) explora las características de los juegos masivos de rol on-line en busca de características que poder aplicar al aprendizaje en videojuegos educativos, mientras Dondlinger (Dondlinger, 2007) hace un repaso de 35 publicaciones que recogen videojuegos educativos y su diseño para extraer claves y puntos comunes en todos ellos. En su revisión, define como elementos definitorios de su efectividad la capacidad de motivar a los estudiantes, aunque no encuentra una fuente única para esta motivación. Algunos estudios la atribuyen al propio acto intrínseco de juego (Amory et al., 1999; Denis & Jouvelot, 2005), mientras que otros lo basan en la naturaleza atractiva de las narrativas que contienen (M D Dickey, 2005; Michele D. Dickey, 2006b). Precisamente Dondlinger marca en su revisión la existencia de narrativas como punto clave para la efectividad de los juegos serios, así como la existencia de objetivos y reglas consistentes que formen su universo. También destaca la capacidad de los juegos de proveer de un ciclo de retroalimentación rápido a través de diferentes medios sensoriales (imágenes, vídeos, sonidos…).

Los juegos revisados se basan en teorías de aprendizaje como el constructivismo (los humanos generan conocimiento y significado a partir de sus experiencias), construccionismo (los individuos construyen modelos mentales para entender el mundo que los rodea) y cognición situada (el saber está siempre ligado a la acción), y tienen como objetivo mejorar múltiples áreas cognitivas: capacidad de atención, concentración espacial, resolución de problemas, toma de decisiones, trabajo colaborativo, creatividad o pensamiento abstracto.

Otros trabajos intentan definir de manera más detallada los elementos que debe contener un juego serio para optimizar su efectividad. Marne (Marne, Wisdom, Huynh-Kim-Bang, & Labat, 2012) define 6 facetas que considera fundamentales en el diseño de cualquier juego serio:

- **Faceta 1**: Objetivos pedagógicos. Todo juego serio debe tener unos objetivos instruccionales claros.
- **Faceta 2**: Simulación del dominio. El juego serio debe representar de manera fidedigna el contexto en el que se desarrolla y sobre el que pretende instruir.
- **Faceta 3**: Interacciones con la simulación. El jugador posee una serie de acciones claras que le permiten interactuar con la simulación.

- **Faceta 4**: Problemas y progresión. El juego define una progresión para el jugador, en la que se enfrente a problemas de dificultad gradual.
- **Faceta 5**: Decorum. El juego posee una estética marcada que lo hace atractivo al jugador.
- **Faceta 6**: Condiciones de uso. Los diseñadores definen la aplicabilidad y contexto del juego. Es decir, cómo, cuándo, dónde y por quién debe ser utilizado.

De manera similar, Anneta (Annetta, 2010) define un marco de referencia para el diseño de juegos educativos. En él, identifica la mayoría de elementos recogidas por las 6 facetas de Marne, pero añade uno más: la enseñanza informada, es decir, los juegos serios deben ser capaces de proveer de datos de evaluación a instructores. Esta evaluación debe a hacerse a través de mecanismos embebidos que permitan al propio juego comunicar los resultados de los alumnos. Concluye que esta característica puede posicionar a los juegos serios como nuevos exámenes estandarizados.

Mientras que estos trabajos proponen un marco general con elementos a tener en cuenta para el diseño de juegos serios, otros autores han querido centrarse en el diseño de las actividades que los jugadores deben completar para lograr resultados satisfactorios.

El Modelo basado en Teoría de Actividades para Juegos Serios (Carvalho et al., 2015) aplica al diseño de juegos el marco conceptual desarrollado por la teoría de actividades (Jonassen & Rohrer-Murphy, 1999), una disciplina que estudia, desde el punto de vista de las ciencias sociales, las prácticas de los seres humanos durante procesos de desarrollo. La principal aplicación de este modelo es compartimentar los módulos que existen dentro de un juego serio, y valorar el papel que desempeñan dentro del juego como unidad completa. Esto permite a los diseñadores tener una mayor perspectiva respecto a los aspectos que cubre el juego.

El modelo de juego experimental (Kiili, 2005) propone un marco abstracto que estructura toda la actividad que debe darse dentro del juego. La Figura 5 muestra el flujo de juego que define este trabajo.

El modelo parte de objetivos educativos, desde los que crea una serie de problemas a resolver por el jugador. Para ello, deberá pasar por fases de generación de ideas, de experimentación activa y de reflexión sobre la retroalimentación ofrecida por el juego. Si observamos con detenimiento algunas de las fases contempladas por este modelo, podemos ver procesos que nos recuerdan a la teoría de aprendizaje de habilidades de Fitts (Fitts, 1964), presentada en la sección anterior.

Torrente (Torrente et al., 2014) centra esfuerzos en definir una metodología de juegos serios centrada en dominios especializados como la medicina, donde deben hacerse consideraciones específicas. Finalmente, otros autores han indagado en las implicaciones técnicas que afectan al diseño de los juegos serios. Por ejemplo, Moreno-Ger (Moreno-Ger et al., 2008) explora los requisitos necesarios para la integración de juegos serios en entornos on-line, su capacidad de adaptación y evaluación, así como consideraciones adicionales como la integración de estándares.

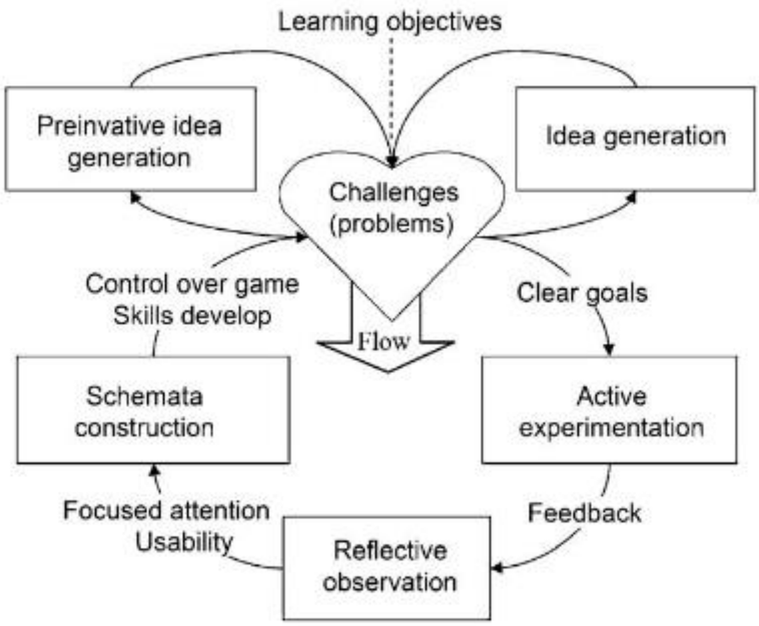


Figura 5. Modelo de juego experimental de Kiili

2.4. Análisis de datos en videojuegos

Como expusimos en la sección anterior, la gran mayoría de juegos serios son evaluados a través de experimentos con un formato tradicional, basado en cuestionarios que los alumnos rellenan antes y/o después de la experiencia con el juego (Calderón & Ruiz, 2015). Esta modalidad de recogida datos, estandarizada dentro del mundo académico, contrasta con las prácticas que se realizan dentro del mundo de los videojuegos comerciales.

La naturaleza interactiva de los videojuegos los convierte en una fuente natural de datos. Un jugador puede llegar a realizar cientos de acciones en el transcurso de una sesión de juego. Estas acciones, enlazadas con su avance dentro del juego, describen el progreso total del jugador, incluyendo detalles como los desafíos que superó con mayor destreza así como aquellos en los que encontró más dificultades.

La industria del videojuego lleva décadas recopilando datos de sus usuarios para entender cómo juegan con los productos que desarrollan. Las técnicas han ido avanzando con el paso del tiempo, y han pasado de medidas simples, como número de jugadores o duración de sesiones de juego, a visualizaciones mucho más complejas, como grafos de navegación y mapas de calor. Los juegos serios aún no han empezado a explotar todo el potencial que estas técnicas pueden ofrecer, aunque sí han empezado a realizar pequeñas incursiones.

A continuación, detallamos la evolución de estas técnicas y lo que han supuesto para el análisis de videojuegos.

2.4.1. Analíticas web

Una de las primeras herramientas utilizadas para el análisis de datos de videojuegos fueron las analíticas web (Drachen, Thurau, Togelius, & Bauckhage, 2013). Este tipo de análisis (a su vez derivado de algunos fundamentos de Inteligencia de Negocio (Rud, 2009)) es utilizado por administradores de páginas web para obtener información sobre el tráfico que visita sus páginas.

La Figura 6 muestra una visualización de datos web recogida por la herramienta Google Analytics (Plaza, 2011). En ella podemos ver datos como, por ejemplo, el número de usuarios, el número de sesiones o duración de las mismas. Estos datos suelen verse extendidos por información demográfica de los usuarios o datos de comercio electrónico (Hasan, Morris, & Probets, 2009).

La facilidad de uso de estas herramientas hace que muchos desarrolladores de videojuegos también las usen para medir el uso de sus juegos (Bondzulic, 2015; Burns & Colbert, 2013). Sin embargo, las analíticas web no logran capturar algunas de las interacciones intrínsecas de los videojuegos. Por ello, existe una disciplina más focalizada (que incorpora muchas de las técnicas de analítica web), denominada analíticas de juego.

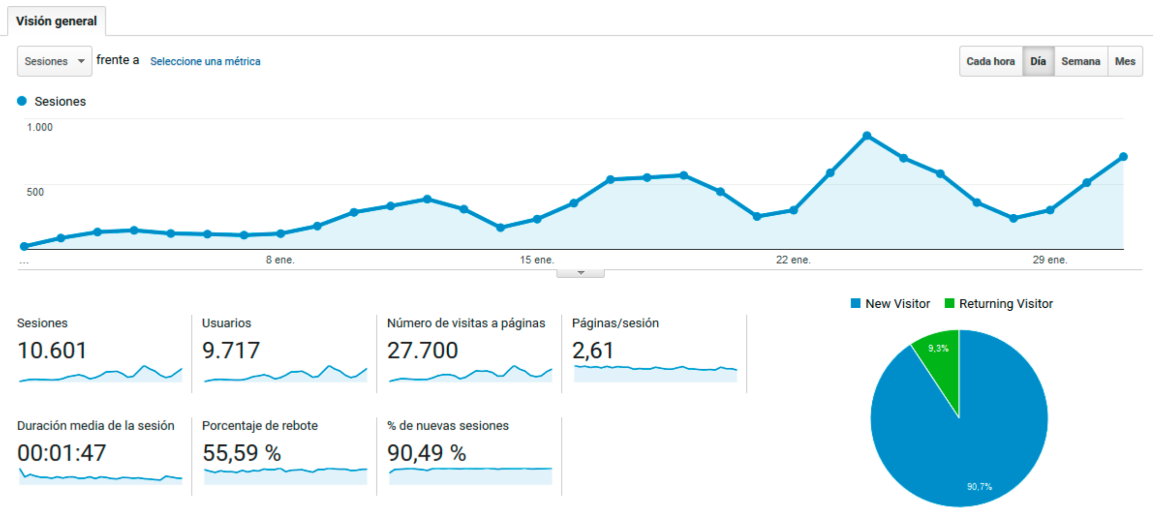


Figura 6. Visualización de analíticas web en la herramienta *Google Analytics*

2.4.2. Analíticas de juego

Las analíticas de juego comprenden una serie de técnicas, análisis y visualizaciones diseñadas para medir los diferentes aspectos del desarrollo y uso de videojuegos (El-Nasr et al., 2013).

Las métricas de juegos pueden estar destinadas a medir diferente procesos relacionados con los videojuegos (Drachen, Seif El-Nasr, & Canossa, 2015). Por ejemplo, las métricas pueden estar dirigidas a medir el rendimiento de un videojuego en el dispositivo destino (con frecuencia este proceso se denomina telemetría), con métricas como uso de memoria, CPU o FPS (*frames per second*). En ocasiones también se capturan métricas para medir el propio proceso de desarrollo software, pero las métricas que más utilizadas para medir el uso final que se hace de los videojuegos son las métricas de usuario, derivadas de la interacción directa de los jugadores (Canossa & Drachen, 2009).

La naturaleza de estas métricas suele estar ligada a géneros y mecánicas específicas de juego, aunque podemos encontrar algunas aplicadas de manera global (Drachen et al., 2015). La Figura 7 muestra una tabla que relaciona géneros con las métricas más comunes en cada uno de ellos. Existen varias estrategias de traqueo para estas métricas:

- **Eventos**: el jugador realiza una interacción con relevancia para el análisis de juego y ésta se envía a un servidor.
- **Frecuencia**: el envío de datos se realiza con una frecuencia predeterminada, siendo el desarrollador el que determina el conjunto de datos a enviar.
- **Iniciadas**: en ocasiones, el desarrollador quiere recolectar métricas que sólo necesita en momentos puntuales del desarrollo. Estas métricas son habilitadas o deshabilitadas por el desarrollador en vez de estar definidas como eventos que se envían siempre.

Muchas compañías de videojuegos han implementado sus propias soluciones software para dar soporte a estas métricas (Medler & Magerko, 2011). Skynet es la herramienta de analíticas de juego implementada por Bioware[4], desarrollada específicamente para analizar los juegos creados por la compañía (Zoeller, 2010). También existen servicios on-line que ofrecen soluciones generales para analíticas de juego. Gameanalytics.com[5] ofrece una herramienta web capaz de ofrecer datos sobre adquisición de jugadores, retención y monetización. Para integrarse con la herramienta, los desarrolladores deben incluir diferentes tipos de peticiones HTTP desde sus juegos.

[4] http://www.bioware.com/en/
[5] http://www.gameanalytics.com/

Género	Métricas
Todos	Número de sesiones de juego, tiempo jugando, divisa virtual gastada, enemigos eliminados, compleción de niveles, puntuación final, acciones por minuto, vidas perdidas, daño recibido, interacciones con dispositivos de entrada (clicks, pulsaciones, teclas…)
RPG (*Role playing games*)	Misiones completadas, evolución de avatar, uso de recursos, habilidades de avatar, estadísticas de combate, progreso de la historia, interacción con NPC, daño recibido/infligido, objetos recogidos, navegación por el mapa de juego
FPS (First person shooter)	Uso de armas, trayectorias, selección de personajes, selección de niveles, selección de vehículos, puntuación de equipos, saltos, puntos de captura/pérdidas, uso de objetos, traqueo de proyectiles, ángulo de cámara, orientación del personaje
Carreras	Pistas escogidas, vehículos escogidos, rendimiento de vehículos, proporción de carreras ganadas y perdidas, mejoras para vehículos adquiridas, velocidad media por pista, colisiones, posición del vehículo
Aventura	Progresión de la historia (basada en nodos), interacciones con NPC, recorrido del mapa, compleción de puzles, uso de objetos, progresión del personaje
Estrategia	Selección de unidades, órdenes dadas las unidades, mejoras adquiridas, ajustes de mapa, tiempo utilizado para cada tipo de tareas (p. ej., construir respecto a atacar al enemigo)

Figura 7. Tabla de las métricas más comunes agrupadas por género de juego.

En algunas ocasiones, los datos capturados no solo se utilizan para el análisis del comportamiento de usuario, sino que se incorporan a las dinámicas de juego dotándolos de una nueva dimensión (Medler, 2009). La serie de juegos Mass Effect[6] incorpora un sistema que guarda datos de usuario entre las diferentes entregas de la saga, haciendo que las decisiones que el jugador ha tomado en juegos anteriores tengan efecto sobre los siguientes juegos. Algunos juegos de carreras, como Mario Kart[7], traquean los mejores resultados de sus jugadores para que después puedan competir contra sí mismos.

El mundo académico también ha explorado las posibilidades de las analíticas de juego como vehículo para responder diferentes preguntas. TRUE (Tracking Real-Time User Experience) (J. H. Kim et al., 2008) es una herramienta enfocada al análisis de comportamiento de usuario. Su objetivo es asistir a desarrolladores y diseñadores durante la fase de prueba de los juegos. La herramienta no solo registra las interacciones de los

[6] https://www.masseffect.com/es-es
[7] http://mariokart8.nintendo.com/es/

usuarios con el juego, sino que además incorpora vídeo de las sesiones de juego e información adicional solicitada al usuario a través de cuestionarios (por ejemplo, si se divirtió jugando a alguno de los niveles).

2.4.3. Analíticas de juegos serios

Aunque las analíticas de juegos serios nacen de manera natural desde las analíticas de juego, su propósito es completamente distinto. Mientras las analíticas de juego tienen como objetivo optimizar valores como la retención de jugadores o la monetización, el objetivo principal de las analíticas de juegos serios es obtener información sobre el proceso de aprendizaje (Loh, Sheng, & Ifenthaler, 2015). Estas analíticas usan principalmente la información de las interacciones de los aprendices con el juego, y no solo en la información obtenida mediante instrumentos externos a la naturaleza del juego como son los cuestionarios.

Las analíticas de juegos serios comprenden desde los métodos de captura de datos y los elementos de diseño que los acompañan, hasta los análisis y visualizaciones más apropiados para cada tarea. A continuación, se presentan algunas aproximaciones que diversos juegos serios han tenido a este tipo de análisis.

Una aproximación común dentro del análisis de juegos serios es la evaluación basada en cuestionarios que están embebidos dentro de las narrativas de juego (Dudzinski et al., 2013; Manero et al., 2015). En estos casos, los diseñadores intentan integrar estos cuestionarios como parte de la historia (por ejemplo, a través de una conversación con un personaje), haciendo que su presentación parezca más natural para el jugador. Sin embargo, esta técnica puede provocar en los estudiantes la sensación de estar siendo examinados, lo que puede dañar su experiencia con el juego. Son varias las corrientes que abogan por lo que se conoce como "evaluación sigilosa" (stealth assessment), basada en recoger datos de aprendizaje de forma implícita y sin afectar al curso normal de juego, buscando que el jugador sea lo menos consciente posible (M. Ventura et al., 2013).

Algunas aproximaciones más alineadas con esta idea se basan en el análisis de transición de estados de jugador. El estado del jugador se representa con un conjunto limitado de variables de juego, y se observa su evolución para evaluar el éxito o fracaso de la sesión de juego. Loh (Loh, Sheng, & Li, 2015) usa esta aproximación para distinguir patrones de juego entre usuarios noveles y expertos. El estudio representa los caminos que sigue cada jugador dentro una cuadrícula. Usa como punto de partida el camino recorrido por jugadores expertos, y lo compara con el camino de jugadores noveles, en busca de diferencias. Como método de evaluación, utiliza la semejanza de un camino con los camino seguido por los usuarios expertos. Lee (Lee, Liu, & Popovic, 2014) analiza la transición entre estados de juegos para crear un modelo capaz de predecir el siguiente estado de juego dado el estado actual. Este modelo puede usarse, por ejemplo, para prevenir en tiempo real a los instructores sobre estudiantes que están teniendo problemas en el desarrollo de la actividad.

El marco teórico ADAGE (Owen, Ramirez, Salmon, & Halverson, 2014) intenta abstraer la noción de estado de juego desde un conjunto de variables a conceptos más cercanos al diseño del propio juego. Su objetivo es proveer de un mecanismo que permita la evaluación automática del recorrido realizado por un jugador. Para ello, define varias categorías de eventos de juego:

- **Unidades de juego**: representan los estados de progreso más cercanos a la mecánica de juego. Esto puede ser niveles, fases o mapas. Los eventos emitidos se relacionan con la compleción de estas unidades.
- **Logros críticos**: momentos que tienen especial relevancia en el juego, y que suponen la consecución de algún desafío. Estos eventos son la base para evaluar el aprendizaje del jugador.
- **Niveles de maestría finales**: momentos en el que el jugador debe combinar las habilidades adquiridas durante el juego y hacer uso de ellas para sobrepasar un desafío final.

ADAGE propone como evaluación la realización de análisis sencillos sobre estos eventos. Por ejemplo, una métrica de evaluación podría ser el número de logros críticos conseguidos respecto al número total de logros.

No obstante, consideramos que las analíticas de juegos serios no solo deben centrarse únicamente en el análisis o la visualización de datos, sino también deberían proveer a los profesores de herramientas que les permitan seguir el aprendizaje de sus alumnos. Por ejemplo, la aplicación para aprender idiomas Duolingo (Ye, 2014) ofrece a profesores un panel en el que puede seguir el progreso individualizado de sus alumnos en cada una de las lecciones (Figura 8[8]). The Radix Endeavor[9] es un juego masivo on-line que imparte contenidos sobre ciencia, tecnología, ingeniería y matemáticas. Entre sus características, ofrece a los profesores una visualización on-line que permite organizar las clases y mantener un seguimiento del progreso de sus alumnos.

2.5. Estándares educativos y analíticas de aprendizaje

Una de las características que los juegos serios no suelen compartir con sus equivalentes comerciales es su despliegue. Mientras que los videojuegos comerciales suelen ser componentes independientes que pueden comunicarse opcionalmente con un servidor

[8] Fuente: https://frenchmajorvsculture.wordpress.com/2015/06/25/duolingo-the-best-free-app-for-language-learning-now-has-duolingo-for-schools/
[9] https://www.radixendeavor.org/

externo, los juegos serios suelen estar integrados dentro de entornos virtuales de aprendizaje, como SGA o MOOCs (Del Blanco Aguado, Torrente, Martínez-Ortiz, & Fernández-Manjón, 2011; Freire, del Blanco, & Fernandez-Manjon, 2014). Adicionalmente, la comunicación entre videojuegos comerciales y servidores de analíticas suele realizarse con interfaces propietarias, mientras que los juegos serios deben comunicarse a través de estándares, normalmente abiertos, soportados por los entornos en los que se integran. Por tanto, el trabajo de las analíticas de juegos serios no se detiene en la búsqueda y validación de técnicas de análisis y evaluación, también necesita encontrar modelos y

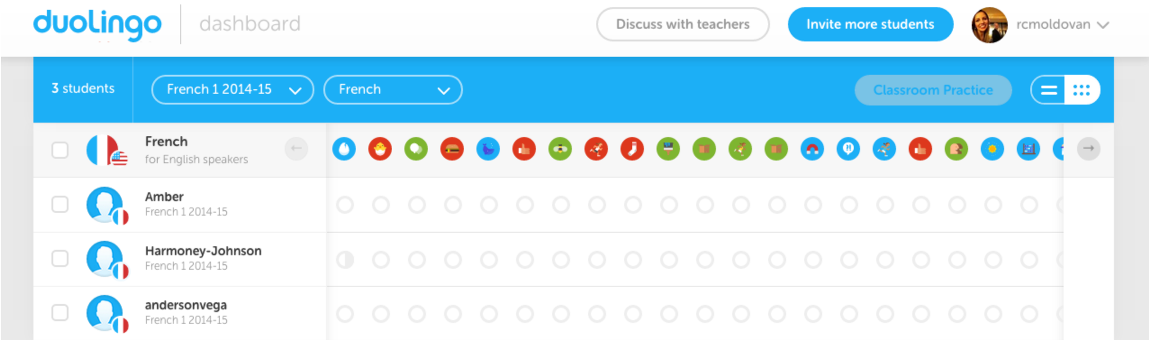


Figura 8. Panel para profesores ofrecido por la aplicación de aprendizaje de idiomas Duolingo.

representaciones estándares de los resultados que obtiene, para que estos pueden ser compartidos con otros sistemas que puedan estar interesados en ellos.

Los estándares educativos dedicados a analíticas de aprendizaje suelen modelar datos de dos tipos: estáticos, datos que no cambian con el tiempo; o dinámicos, datos que se actualizan de manera más frecuente (Griffiths, Hoel, & Cooper, 2016). Y estos datos pueden proceder de múltiples fuentes: personas, recursos, servicios, actividades educativas o evaluaciones (Cooper, 2014).

En esta sección nos centramos en aquellos estándares que modelan datos dinámicos, principalmente procedentes de personas, actividades educativas y evaluaciones, ya que son los que pueden representar con mayor cercanía la naturaleza interactiva de los juegos serios. Primero repasamos algunos de los estándares y especificaciones que han tratado de modelar aprendizaje en contenidos educativos en general, para después centramos en tres de los más adecuados para juegos serios: SCORM, IMS Caliper y Experience API.

2.5.1. Estándares de caja negra

Definimos estándares de caja negra como aquellas especificaciones que solo modelan y comunican resultados a los entornos con los que se integran, sin aportar detalles de cómo se obtuvieron esos resultados (del Blanco et al., 2013). Es decir, estos estándares comunican como datos, por ejemplo, si el estudiante superó o no la actividad, o la puntuación que obtuvieron, pero nunca detalles sobre cómo se llegó a ese resultado.

Este es el caso de la Open Badge Initiative, llevada a cabo por Mozilla (Mozilla, 2011). La idea de esta iniciativa es recopilar, a través de un repositorio de confianza, las habilidades y competencias que una persona ha ido logrando a lo largo de su carrera. En este caso, Mozilla se encarga de ser la entidad verificadora de los logros, pero la concesión de los mismos le corresponde a instituciones de terceros. Mozilla pone al servicio de instituciones educativas un protocolo estándar que permite comunicarse con el repositorio global, tanto para conceder medallas (o badges) a individuos como para comprobar las que ya poseen.

Dentro de la gran familia de estándares definidos por el IMS Global Learning Consortium[10], IMS Basic Outcomes ofrece la posibilidad de comunicar notas de resultados asociados a actividades concretas (Fontenla, Pérez, & Caeiro, 2011). La organización de Advanced Distributed Learning[11] (ADL) también ha aportado a esta familia de estándares añadiendo características de evaluación a uno de sus estándares: SCORM.

SCORM es uno de los estándares más extendidos en el despliegue y comunicación de recursos educativos dentro de entornos de aprendizaje virtuales (Capone, 2004), y ha sido utilizado en algunas investigaciones como estándar para traquear resultados en juegos serios (del Blanco et al., 2013). El estándar define un conjunto de requisitos técnicos que permite que sistemas de aprendizaje virtual puedan importar y reutilizar contenidos educativos completos, a los que define como SCO (Shareable Content Objects). Entre sus características, define una interfaz de comunicación que permite una comunicación estándar y bidireccional entre un SCO (contenido educativo) y el entorno virtual con el que se integra.

Esta interfaz de comunicación está basada en el modelo de datos CMI (Advanced Distributed Learning, 2011). Este modelo permite el envío del estado de compleción, puntuación y progreso de la actividad. También permite el envío de información extra, como comentarios en texto libre, interacciones para traquear cuestionarios y otros elementos educativos, y objetivos, para traquear puntuación y compleción de sub-tareas. La Figura 9 contiene una tabla detallando algunos de los campos del modelo CMI y su relación con los SCOs.

Como se ha comentado, la principal desventaja de SCORM y del resto de estándares basados en objetivos es que no deja rastro del proceso de evaluación que se ha seguido para conceder un resultado u otro. Las nuevas aproximaciones basadas en analíticas de

[10] https://www.imsglobal.org/
[11] https://www.adlnet.gov/

aprendizaje están buscando modelos más abiertos que permitan una mayor granularidad a la hora de representar el proceso de aprendizaje.

2.5.2. Estándares de caja blanca

Definimos estándares de caja blanca como aquellas especificaciones que, además de proveer resultados consumibles por sistemas externos, ofrecen información sobre las interacciones y el modelo de evaluación seguido para llegar a esos resultados (del Blanco et al., 2013). Muchos de estos estándares se basan en el registro de interacciones de usuarios en archivos de texto o bases de datos (logs), que son después procesados por un sistema de análisis capaz de ofrecer visualizaciones y obtener distintos tipos de resultados.

El PSLC DataShop Tutor Message Format (Koedinger, Baker, Cunningham, & Skogsholm, 2010) define un formato XML destinado a representar las interacciones realizadas por un estudiante dentro de un entorno educativo. Por ejemplo, el XML mostrado en la Figura 10 representa la carga y reproducción de un vídeo en este formato. En una línea similar, Contextualized Attention Metadata (Schmitz, Wolpers, Kirschenmann, & Niemann, 2007) ofrece un sistema para realizar el seguimiento de las interacciones de usuarios en

Elemento del modelo	**Descripción**
Datos de lanzamiento *cmi.launch_data*	Provee de datos que el SCO puede utilizar para su inicialización, por ejemplo con objeto de adaptar el contenido de la actividad
Localización *cmi.location*	Representa una localización o posición dentro de un SCO. Por ejemplo, si el SCO contuviera un libro electrónico, este campo podría utilizarse para marcar la página que el usuario está leyendo
Estado de compleción *cmi.completion_status*	Indica si el usuario ha completado el SCO o no.
Puntuaciones *cmi.score.raw* *cmi.score.min* *cmi.score.max* *cmi.score.scaled*	Identifica la puntuación numérica final del usuario en el SCO. Permite enviar los valores máximos y mínimos de la puntuación.
Estado de éxito *cmi.success_status*	Indica como el usuario completó el SCO, si de manera satisfactoria o no.
Tiempo total cmi.total_time	Alberga el tiempo que el usuario utilizó para completar el SCO.

Figura 9. Tabla con los campos del modelo CMI relacionado con su uso dentro de SCORM

entornos de aprendizaje.

Otra especificación relevante en el seguimiento de interacciones es Activity Stream (Snell et al., 2011). Este formato, con sintaxis en JSON[12], fue diseñado para representar secuencialmente actividades realizadas por un usuario en contexto concreto. Cada acción se representa por una actividad, que consta de tres elementos:

- **Actor**: usuario identificado que realiza la actividad.
- **Verbo**: la acción que se está realizando.
- **Objeto**: el objetivo que está recibiendo la acción.

Esta especificación nació con propósito general y se ha utilizado como base para representar las secuencias de interacciones que se dan redes sociales como, por ejemplo, Facebook.

```
<context_message
   context_message_id="02CE3AE5-F6D5-9177-913F-C34730F1096C"
   name="LOAD_VIDEO">
   <dataset>
      <name>Example Media Dataset</name>
      <level type="unit">
         <name>Stoichiometry</name>
         <level type="section">
            <name>What are moles?</name>
            <problem>
               <name>mymovie.flv</name>
            </problem>
         </level>
      </level>
   </dataset>
</context_message>

<tool_message
    context_message_id="02CE3AE5-F6D5-9177-913F-C34730F1096C"
    <semantic_event
        transaction_id="1F3A9B23-9164-DD83-EBB2-1589FD38D4B3"
        name="VIDEO_ACTION" />
    <event_descriptor>
        <selection>
        _level0.VideoPlayerInstance1.sliderButtonName
        </selection>
        <selection type="media_file">mymovie.flv</selection>
        <selection type="clip_length">00:08:00.0</input>
        <action>play</action>
        <input type="time">00:02:34.2</input>
    </event_descriptor>

</tool_message>
```

Figura 10. Representación de una carga y reproducción de vídeo en el *PSLC DataShop Tutor Message Format*

[12] http://www.w3schools.com/js/js_json_intro.asp

Dentro del mundo de la educación, IMS y ADL han utilizado Activity Stream como base para desarrollar sus estándares de seguimiento de actividad en entornos educativos, con IMS Caliper y Experience API respectivamente. Por su especial relevancia dentro del ámbito educativo, pasamos a profundizar en las características de cada uno de ellos.

2.5.2.1. IMS Caliper

Caliper Analytics es una especificación desarrollada por IMS, cuyo objetivo es establecer un canal de comunicación abierto y continuo entre actividades educativas y entornos educativos (IMS Global Consortium, 2016).

Los tipos de actividades soportados por Caliper, así como los eventos e interacciones que se pueden comunicar desde ellos están predefinidos por la especificación, que además están agrupados por perfiles de métricas (metric profiles). Por ejemplo, Caliper define un perfil de métrica denominado "Media", destinado a representar las interacciones de usuarios con elementos multimedia (por ejemplo, reproducir un vídeo). La especificación también define perfiles específicos para evaluación.

La especificación Sensor API define la representación estándar de cada uno de los eventos de los perfiles de métricas. Su estructura se basa también en Activity Stream y contiene, entre otros atributos: actor, acción y objeto. Tanto las acciones como los objetos posibles vienen predefinidos por el perfil. La Figura 11 muestra la estructura del sistema y cómo se relacionan los diferentes perfiles con la especificación Sensor API.

La principal limitación de esta especificación es la imposibilidad de extender los perfiles de métricas para soportar nuevos tipos de contenido educativos (como juegos serios). La especificación aún está en desarrollo, y esto podría cambiar en el futuro.

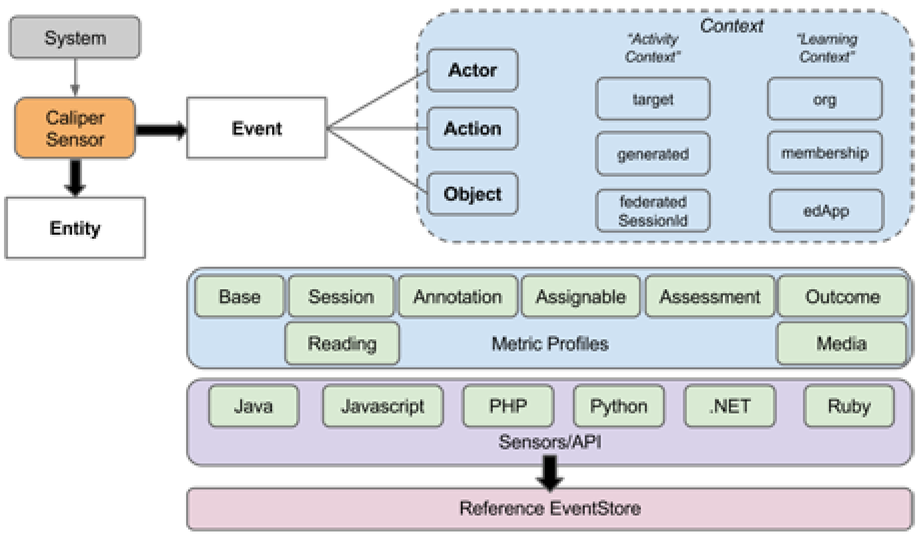


Figura 11. Arquitectura de IMS Caliper (https://www.imsglobal.org/activity/caliperram)

2.5.2.1. Experience API

Experience API (xAPI) es una especificación desarrollada por una comunidad abierta liderada por ADL (ADL Initiative, 2016). Al igual que IMS Caliper, su objetivo es definir un modelo de datos que permita la comunicación de la actividad desde un contenido educativo a un entorno virtual.

La unidad de seguimiento definida es la frase (o statement), y también se compone de actor, verbo y objeto. xAPI añade otros atributos que pueden resultar interesantes para la evaluación de actividades, como resultado, que contiene las puntuaciones o consecuencias de la frase; contexto, que contiene el entorno educativo en el que se desarrolla la actividad; o autoridad, que identifica a la autoridad que certifica la validez de una frase.

Al contrario que Caliper, xAPI no predetermina los verbos y objetos que puede traquear. En cambio, ofrece una estructura totalmente configurable y extensible, que los desarrolladores pueden adaptar a sus necesidades. De hecho, esta extensión se puede hacer mediante una comunidad abierta de desarrollo en la que cualquier parte interesada puede participar.

Sin embargo, su oferta en el análisis es más limitada. Mientras que la aproximación cerrada de Caliper se debe al afán de proveer un sistema de traqueo completo, desde la captura de datos hasta su análisis, xAPI solo ofrece soporte para modelar el envío de datos y una solución de almacenamiento para albergar frases (la base de datos que se denomina Learning Records Store). Deberán ser los desarrolladores los que implementen sus propios procesados de datos y análisis para completar el ciclo completo de analíticas de aprendizaje.

2.6. Conclusiones

Este capítulo concluye con un resumen esquemático de las principales conclusiones extraídas tras el análisis de dominio:

1. **Las analíticas de aprendizaje han empezado a utilizarse con éxito en contenidos educativos on-line, pero aún no han sido aplicadas en profundidad a los juegos serios.**

   Como se describió en la sección 2.1, las analíticas de aprendizaje han empezado a ser aplicadas a múltiples medios y contextos para comprender y mejorar los procesos de aprendizaje: desde sistemas clásicos como los tutores inteligentes, hasta sistemas e-learning más modernos como los MOOCs. Sin embargo, como se repasa en la sección 2.3, las aproximaciones que se han hecho sobre juegos serios todavía no son maduras y no han hecho un uso completo de las ventajas que ofrecen dichas analíticas.

2. **Los juegos serios son un contenido educativo probado, pero aún existe discusión sobre la mejor aproximación para diseñar y evaluar su efectividad.**

   En la sección 2.2 se repasó la relación que existe entre los juegos serios y algunas teorías de aprendizaje, así como las aproximaciones que múltiples investigaciones hacen al diseño de juegos considerando estas teorías. Algunos investigadores afirman que estas aproximaciones en ocasiones son incompletas o erróneas, y dificultan la evaluación del aprendizaje en los juegos serios, destacando la necesidad de seguir buscando metodologías generales para su validación. La naturaleza interactiva de los juegos serios los convierte en candidatos perfectos para las analíticas de aprendizaje. Su integración con ellas puede abrir caminos y nuevas metodologías de diseño y evaluación.

3. **Las analíticas de juego serios han empezado a despegar, pero aún necesitan de la estructuración y validación presente en las analíticas de juego o las analíticas web.**

   En la sección 2.3 se exploraron las disciplinas de análisis más cercanas a los juegos serios, en concreto, las analíticas web y las analíticas de juegos. Estas disciplinas están alcanzando un punto de madurez del que aún distan las analíticas de juegos serios. Muchos de sus principios pueden ser reutilizados o adaptados, pero las analíticas de juegos serios necesitan seguir avanzando en su propio camino, analizando un aspecto que las otras disciplinas no miden: el aprendizaje.

4. **Se está produciendo un cambio de paradigma en los estándares educativos encargados de comunicar evaluación de aprendizaje. Los juegos serios deben adaptarse a este nuevo paradigma.**

   En la sección 2.3 se muestran algunos de los estándares educativos utilizados en la actualidad para la comunicación de resultados desde recursos educativos. Existen muchas iniciativas pero actualmente se tiende a modelos de “caja blanca”, basados en recoger y trazar con más detalle las interacciones de los estudiantes con los contenidos educativos. Los juegos serios, como contenido educativo que suele desplegarse en integración con otros entornos, deben hacer uso de estos estándares para comunicar su actividad. Es necesario explorar la manera de aplicar estándares a ese proceso para obtener todos los beneficios que ofrecen los juegos serios.

Capítulo 3. Objetivos y planteamiento del trabajo

En el capítulo 1 se presentó el objetivo principal de esta tesis junto con un esbozo de las tareas a realizar. Este capítulo construye, a partir del estudio del dominio realizado en el capítulo 2, el alcance total del objetivo de la tesis, así como los sub-objetivos que la componen.

A continuación, cada uno de los sub-objetivos se desglosa en pasos a completar, que son detallados como planteamiento del trabajo de esta tesis.

3.1. Objetivos de la tesis

Como se define su título, el objetivo de esta tesis es *mejorar la evaluación de juegos serios a través de analíticas de aprendizaje*. De una manera extendida, podemos detallar este objetivo como:

> Proponer un marco teórico y práctico que permita mejorar la evaluación de juegos serios, utilizando como fuente principal de datos las interacciones realizadas por los usuarios con el juego. Los principales beneficiarios de esta evaluación serán los profesores y los instructores. El alcance de esta mejora será planteado para todo el ciclo de vida de un juego serio, desde su incepción y diseño, pasando por su análisis haciendo uso de analíticas de aprendizaje, y hasta su despliegue final en entornos educativos.

Antes de continuar, debemos limitar el alcance de algunos de los términos contenidos en esta definición:

- **Marco teórico y práctico:** el trabajo de esta tesis no quiere quedarse en el marco teórico. A partir del modelo conceptual quiere desarrollar software que pueda ser utilizado en casos de usos reales, y que sea ese uso que el valide o invalide las hipótesis de esta tesis. Esta validación se hará con juegos serios reales y una infraestructura de captura de datos.
- **Evaluación de juegos serios**: como se expuso en el capítulo 2, la evaluación de juegos serios puede tratar de obtener el impacto en diferentes aspectos de los usuarios que los usan, como la motivación, la eficiencia o el aprendizaje. En el contexto de esta tesis, evaluar un juego serio siempre se centrará en el *aprendizaje*, es decir, en la capacidad de un juego serio de "medir" o determinar el nivel de conocimiento de un estudiante a partir de su interacción con el propio juego.
- **Profesores e instructores como principales beneficiarios**: en la sección 2.1 nombrábamos los diferentes actores que podían beneficiarse de la aplicación de analíticas de aprendizaje (principalmente, estudiantes, profesores e personal

administrativo/instituciones). Los resultados de evaluación que persigue esta tesis son en pos de mejorar la tarea docente de profesores e instructores, es decir, la captura de datos, análisis y visualizaciones que se desarrollarán estarán enfocadas a simplificar el seguimiento de los alumnos por parte del profesorado. Aunque una vez capturados estos datos también se pueden utilizar en beneficio de los alumnos o de los administradores educativos

- **Ciclo de vida de un juego serio**: son muchos los aspectos envueltos en el desarrollo y uso de un juego serio. El uso de analíticas de aprendizaje afecta a todos los pasos de su ciclo de su vida, y el objetivo de esta tesis es proponer metodologías y técnicas que consideren los pasos más importantes de este ciclo: diseño, despliegue y evaluación.

Los juegos serios son la pieza angular de esta tesis. Uno de los principales desafíos de este trabajo es integrar y probar sus avances en juegos serios reales, que puedan ofrecer una validación de los avances logrados. En general, el coste de desarrollar cualquier videojuego puede y suele ser muy alto. En su implementación se requiere trabajo de desarrollo software, pero también la creación de recursos artísticos (imágenes, animaciones, sonidos, vídeos…). En el caso de los juegos serios, a este trabajo hay que sumar la validación educativa, que suele requerir la colaboración con un equipo heterogéneo en el que hay también instructores y profesores.

Para mantener la cantidad de trabajo necesaria para el desarrollo de juegos serios dentro de límites abordables en un trabajo de tesis, se consideran dos opciones:

- **Integrar resultados con juegos serios ya existentes o desarrollados por el grupo de investigación**: el desarrollo de este trabajo se hace bajo el amparo del grupo de investigación e-UCM. El trabajo de este grupo en diferentes proyectos nacionales e internacionales conlleva, en ocasiones, la creación de juegos serios. Los avances de esta tesis pueden ser en integrados en estos juegos, siempre y cuando no supongan un riesgo para el trabajo final requerido por los proyectos.
- **Desarrollar juegos serios asequibles desde cero**: puesto que los objetivos de esta tesis abarcan todo el ciclo de vida de un juego serio, se hace necesario el desarrollo de, al menos, un juego serio que recoja todos los resultados que se recojan en ella. Para mitigar dependencias externas, el contenido educativo de este juego deberá ser una materia que el doctorando (y el director) domine, para que él mismo pueda validar la aproximación educativa. También se considera la posibilidad de desarrollar algún juego con contenido mínimo que sirva de prueba conceptual para mostrar ciertos avances.

Además del desarrollo de juegos serios, se hace necesario el uso de una arquitectura software que sirva de entorno de despliegue, recogida de datos y análisis. Esta arquitectura debe cumplir todos los requisitos del proceso de analíticas de aprendizaje descrito en la sección 2.1, y será la principal proveedora de resultados en juegos serios.

Además de cubrir los procesos básicos de análisis, la plataforma debe cumplir los siguientes requisitos para satisfacer las necesidades de esta tesis:

1. La plataforma debe estar enfocada (a permitir una adaptación que la enfoque) a las analíticas de aprendizaje.
2. El código sobre el que se ejecute debe estar disponible y ser abierto, para que se puedan realizar las modificaciones que se requieran para el desarrollo del trabajo.
3. La propiedad de los datos recogidos por la plataforma debe ser del grupo de investigación en el que se enmarca la tesis puesto que el trabajo incluye la realización de experimentos con usuarios reales.

Además en todo momento en el diseño experimental se debe aplicar un escrupuloso respeto de los aspectos éticos y de protección de datos y más en aquellos casos en los que los usuarios son menores de edad.

En la sección 2.3 se presentaban algunas plataformas que pueden cubrir algunos de los pasos de este proceso. Lamentablemente, ninguna de ellas cumple las tres características requeridas para el trabajo, así que se hace necesario el trabajo de una plataforma de análisis de datos desde cero.

La cantidad de recursos necesarios para desarrollar una plataforma de análisis de datos completa es alta. Por tanto, el desarrollo de la plataforma dará prioridad a los módulos fundamentales del proceso de análisis (captura, análisis y visualización) creando implementaciones mínimas pero robustas para otras características más secundarias en el contexto de esta tesis (como escalabilidad o seguridad).

Finalmente, además de la creación de juegos serios y el desarrollo de una plataforma de análisis, se plantea como objetivo de esta tesis desarrollar una especificación abierta que permita capturar las interacciones y logros reportados por los juegos. Esta especificación abierta aumenta su reutilización ya que permite que la solución planteada sea generalizable y aplicable a otros casos. Siguiendo con el espíritu eminentemente práctico de esta tesis, además de un modelo abstracto, se marca como objetivo la implementación completa de la especificación sobre un estándar actual que lo permita. Como validación, se incluirá esta especificación en al menos un juego serio.

En la filosofía del grupo de investigación, y cumpliendo con el carácter abierto que todo proyecto de investigación debería seguir, todo el software creado dentro del ámbito de esta tesis será publicado con licencias libres para que otros puedan reutilizarlo y adaptarlo.

3.2. Planteamiento del trabajo

El trabajo planteado en la sección anterior es amplio y concierne tareas de muy distintos tipos. Se resume en 3 objetivos:

1. Diseñar e implementar una arquitectura de análisis enfocada a las analíticas de aprendizaje centrado en juegos serios.
2. Proponer una metodología de diseño de juegos serios que facilite su evaluación a la vez que estructure su contenido.
3. Diseñar un modelo para captura de datos en juegos serios, y formalizarlo a través de un estándar de comunicación.

En vez de abordar todos los objetivos como un solo bloque, se propone una aproximación iterativa como planteamiento de trabajo. Cada una de esas iteraciones puede tener una o varias tareas. Podemos clasificar las tareas necesarias para completar esta tesis en 4 categorías:

1. **Desarrollo de juegos serio / integración en juego serio**: Esta tarea incluye el trabajo de desarrollo de un juego serio, partiendo desde cero o desde una versión anterior a la que se añaden nuevas características. Esta tarea también considera la integración de características de análisis en un juego desarrollado por un tercero, o la integración de un juego con la plataforma de análisis.
2. **Desarrollo de plataforma de análisis**: Esta tarea incluye las labores de desarrollo necesarias en la plataforma de análisis para dar soporte a experimentos con juegos serios.
3. **Realización de experimentos y validación de resultados**: Esta tarea requiere la realización de un experimento con usuarios reales, que valide avances significativos en un juego serio y/o en la plataforma de desarrollo.
4. **Desarrollo del estándar de comunicación**: Esta tarea comprende el trabajo necesario para el desarrollo del estándar que representará los resultados de juegos serios. Como este paso requiere una cierta madurez del estado de la tesis (principalmente, un modelo de datos sólido y validado), será abordado en la fase final.

En la Figura 12 se muestra una figura que esquematiza el proceso de trabajo. El proceso de iteración principal constará de un desarrollo paralelo de juego serio y la plataforma de análisis, seguido por un experimento con usuarios reales. En la fase final, se desarrollará el estándar de comunicación y se añadirá a un juego serio.

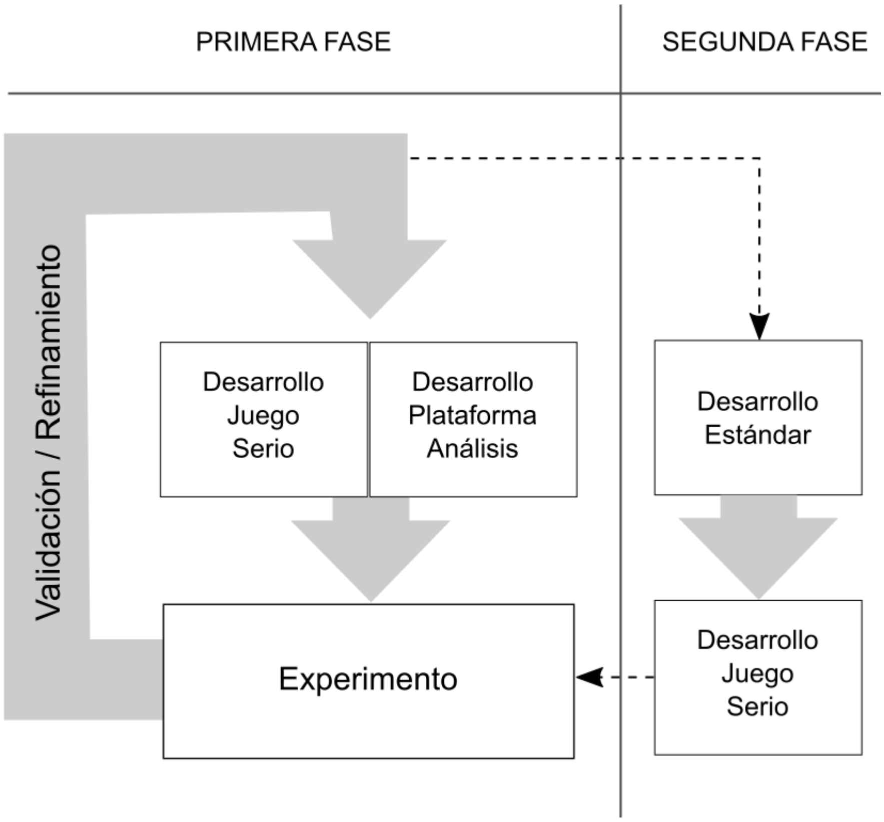


Figura 12. Planteamiento iterativo del trabajo para el desarrollo de la tesis

Capítulo 4. Discusión integradora y contribuciones

En este capítulo se presentan las contribuciones de estas tesis, disgregadas de los artículos enumerados en el capítulo 6. La presentación de los resultados se hace en base a los objetivos expuestos en el capítulo 3, y no al orden cronológico en el que se fueron sucediendo. El capítulo contiene una sección por cada objetivo planteado, más una sección final que detalla los casos de estudio que validaron los resultados.

Si se quiera seguir el orden cronológico de las iteraciones definidas para el proceso de trabajo, el orden es el siguiente:

- Primera iteración: secciones 4.1.1, 4.3.1, 4.4.1.
- Segunda iteración: secciones 4.1.2, 4.2.1, 4.3.2, 4.4.2.
- Tercera iteración: secciones: 4.1.3, 4.2.2, 4.4.3, 4.4.5.
- Fase final: secciones 4.3.3, 4.4.4.

4.1. Arquitectura para analíticas de aprendizaje centrada en juegos serios

4.1.1. Arquitectura inicial para analíticas de aprendizaje

La esquematización inicial de la plataforma para el análisis del aprendizaje se presenta en *A framework to improve evaluation in educational games* (sección 6.1). Esta publicación plantea los pasos que, idealmente, toda plataforma de análisis de aprendizaje debería contemplar en su implementación:

- **Seleccionar:** De entre todos los datos disponibles en el contenido educativo, escoger aquellos que sean más relevantes.
- **Capturar:** Recolectar y almacenar los datos seleccionados.
- **Agregar y generar informes:** Limpiar y ordenar los datos capturados, y generar visualizaciones con ellos.
- **Evaluar:** Comprender los datos agregados y convertirlos en información de evaluación de estudiantes.
- **Adaptar:** Utilizar la información obtenida para adaptar el contenido educativo a las necesidades del estudiante.
- **Usar y refinar:** Utilizar la información obtenida para generar intervenciones que mejoren el proceso educativo en su conjunto.

- **Compartir:** Compartir con aquellos sistemas que puedan estar interesados la información obtenida por el sistema.

Cada uno de estos pasos se ve acompañado de una pieza de modelo que define qué datos procesar y cómo actuar con ellos. La Figura 13 muestra una esquematización del sistema completo. Como se observa, todo el proceso de análisis nace un motor de juegos, encargado de enviar las señales de interacción desde los juegos serios.

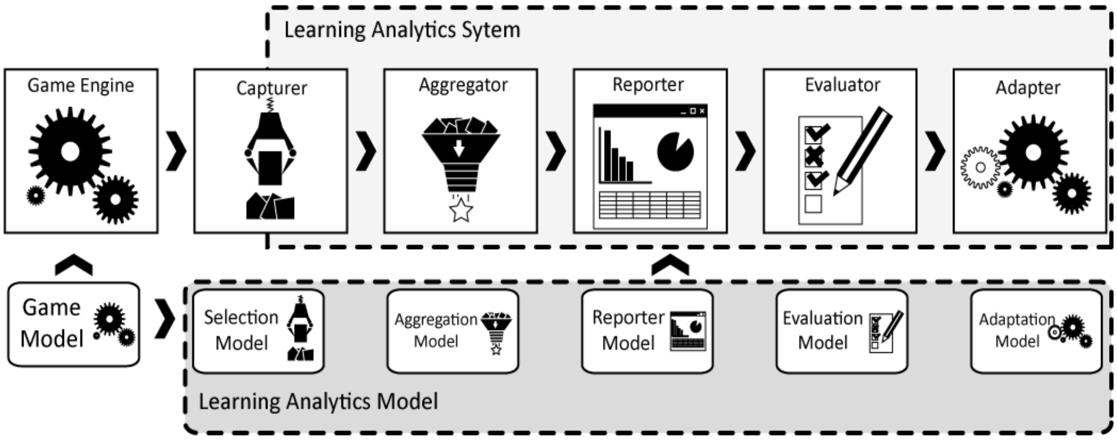


Figura 13. Planteamiento iterativo del trabajo para el desarrollo de la tesis

En esta primera aproximación, se hace especial énfasis en esa integración con un motor de juegos, que comprende los dos primeros pasos del análisis: selección y captura.

Como se mencionó en el capítulo 3, una de las estrategias para obtener juegos serios de esta tesis es integrar el sistema de analíticas con juegos que se fueran desarrollando en el grupo de investigación e-UCM. En el artículo, se hace una propuesta de implementación de que integra un proceso de captura de datos en el motor de juegos educativos e-Adventure (Torrente et al., 2010). La plataforma para desarrollo de juegos e-Adventure se compone de dos módulos: un editor, que permite la creación de juegos de aventuras (*point & click)* sin necesidad de programación; y un motor, que procesa los modelos de juegos creados en el editor y los ejecuta en forma de juego.

La herramienta e-Adventure era la utilizada por los investigadores del grupo para crear nuevos juegos serios. La integración nativa de un proceso de captura de interacciones en el motor de juegos implicaba que todo nuevo juego implementado con e-Adventure soportaría la captura de interacciones por defecto. La sección 4.3.1 muestra con más detalle el modelo de captura de datos.

Como resultado de esta primera aproximación se obtuvo:

1. Diseño completo de la plataforma de análisis de aprendizaje para juegos serios.
2. Implementación mínima de un servicio para dos de los procesos: selección y captura.
3. Integración de captura en el motor de juegos e-Adventure.

4.1.2. Arquitectura refinada para analíticas de aprendizaje

El desarrollo de la plataforma de análisis continúa con una inmersión más detallada en los pasos del proceso de análisis, revisando el trabajo realizado en los pasos ya implementados (selección y captura) y entrando en mayor detalle en aquellos que aún estaban por implementar.

En *Application of Learning Analytics in educational videogames* (sección 7.3), se refina el modelo de captura con eventos, seleccionando aquellos con una mayor significación de aprendizaje. El modelo se muestra con más detalle en la sección 4.3.2.

Este artículo también estudia las posibles aplicaciones de evaluar los resultados en un juego serio. Define tres categorías:

- **Evaluación del artefacto de juego:** Los datos de interacción capturados son utilizados para evaluar la validez del juego desde un punto de vista lúdico, de la misma manera que lo harían las analíticas de juego. Esta evaluación se plantea para la primera fase del desarrollo, cuando se tiene una primera versión del juego y es necesario detectar sus errores o flaquezas antes de lanzarlo al público objetivo.
- **Evaluación *del* aprendizaje:** Esta la evaluación objetivo de esta tesis, que consiste en medir el nivel de conocimiento de los estudiantes en un área concreta a partir de sus interacciones con un juego serio.
- **Evaluación *para* aprendizaje:** En esta evaluación, la medición de nivel de conocimiento no se utiliza para informar a los profesores, sino para afectar el propio transcurso del juego (por ejemplo, adaptando su dificultad).

A partir de este momento se focaliza el desarrollo en la evaluación del aprendizaje, moviendo a un segundo plano las otras dos.

El artículo también muestra algunas de las visualizaciones iniciales implementadas para la plataforma de análisis. La Figura 14 muestra una visualización animada representando el grafo de escenas de un juego serio. En la animación se muestra, con tiempo acelerado, como cada uno de los jugadores va desplazándose por las escenas del juego.

En esta segunda iteración se intentó explorar las posibles visualizaciones que podían implementarse con los datos existentes, pero sin tener en cuenta su objetivo final, es decir, si los profesores e instructores podían beneficiarse de ellas o no.

Como resultado de este refinamiento se obtuvo:

1. Diseño de evaluación para analíticas de aprendizaje de juegos serios: se descarta evaluación para aprendizaje (adaptación) y se centra en la evaluación del aprendizaje. La evaluación del artefacto es compatible en algunos aspectos con la evaluación del aprendizaje, así que se mantiene, pero en un segundo plano.
2. Implementación de visualizaciones e informes para la plataforma de análisis

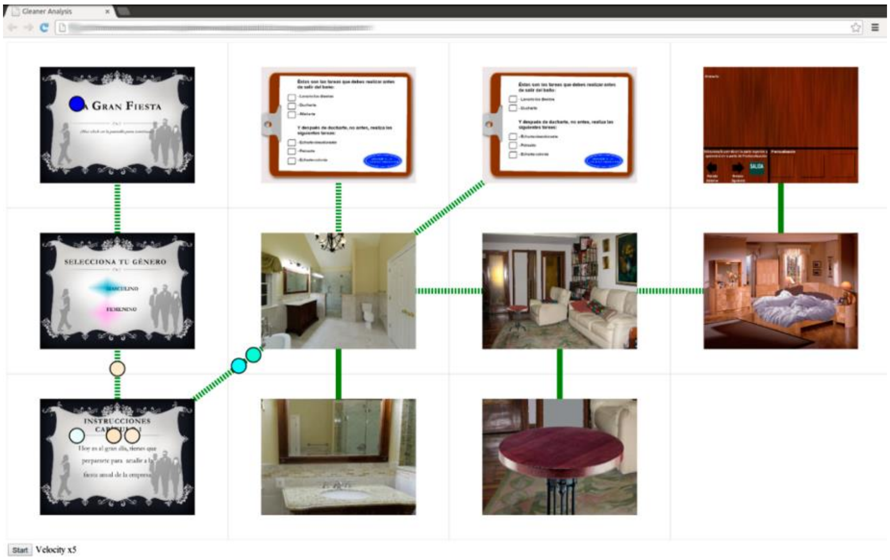


Figura 14. Visualización animada que muestra como cada uno de los jugadores, representados por círculos, van desplazándose por las distintas fases de un juego serio.

4.1.3. Versión final de la arquitectura

En la primera iteración de la arquitectura se definió un marco completo para el análisis. En la segunda iteración se focalizaron los aspectos de evaluación y se experimentó con múltiples visualizaciones. En la versión final se obtiene una arquitectura enfocada a la evaluación del aprendizaje centrada en el profesor. Parte de esta iteración final se recoge en *Applying learning analytics to simplify serios games deployment in the classroom* (sección 7.4).

Este artículo presenta una plataforma de análisis que ofrece las siguientes características para asistir a desarrolladores y profesores en el proceso completo de despliegue de un juego serio:

1. Los desarrolladores disponen de una librería de seguimiento que pueden integrar con sus juegos para comunicar interacciones a la plataforma de análisis.
2. La plataforma de análisis permite mostrar resultados de evaluación en tiempo real. Es decir, si un profesor despliega un juego serio en una de sus clases, la plataforma le permite observar cómo evolucionan los resultados de sus alumnos. También si están cometiendo errores y su intervención es necesaria.

3. La plataforma de análisis recoge los resultados finales de evaluación de todos los alumnos. Esto permite al profesor realizar un análisis de la clase después de la sesión de juego.

La implementación final de la plataforma de análisis recibe el nombre de *Gleaner*, y su código está disponible en https://github.com/e-ucm/gleaner

4.2. Metodología de diseño de juegos serios

4.2.1. Elementos de juego para estructuración de contenido educativo

En *Building a scalable game engine to teach computer science languages* (sección 7.1.), se identifican, a través de un caso de estudio que se expondrá con más detalle en la sección 4.4.2, consideraciones de diseño y elementos de juego útiles para la estructuración de la entrega del contenido educativo:

- Los estudiantes deben practicar la habilidad o conocimiento que el juego serio pretende impartir. Una mera experiencia contemplativa, donde el juego solo expone conocimiento, no es suficiente para el aprendizaje.
- El juego debe promover la reflexión del estudiante, proceso necesario para el afianzamiento de conocimientos. Por ello, deben evitarse presiones de tipo temporal que puedan limitar esta reflexión.
- Los niveles del juego deben presentar el contenido educativo de manera incremental. Los conceptos deben ser introducidos de manera individual, permitiendo al estudiante entender y practicar cada uno de ellos antes de pasar al siguiente.
- Cada nivel debe tener un objetivo claro e idealmente deben existir varios caminos para lograr ese objetivo. Esto fomenta la creatividad de los estudiantes a la hora de aplicar el conocimiento adquirido.
- El uso de puntuaciones y otros elementos competitivos es necesario para ofrecer una sensación de progreso al estudiante, así como fomentar una sana competición entre compañeros.

Estas consideraciones se aplican a un juego serio cuya efectividad se valida a través de un experimento tradicional, basado en cuestionarios antes y después de la experiencia con el juego. El experimento tiene resultados positivos, mostrando aprendizaje en cada uno de los grupos con los que se utilizó el juego.

Las contribuciones obtenidas a partir de este trabajo son:

1. Identificación de consideraciones de diseños y elementos de juego para la estructuración de aprendizaje, aplicadas a un juego serio real.

2. Validación tradicional de un juego serio basado en los principios de diseño propuesto. Esta validación basada en cuestionarios sirve como punto de partida para la aplicación de analíticas de aprendizaje.

4.2.2. Metodología para evaluación de aprendizaje

En el artículo *A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes* (sección 7.5) se formalizan las directrices reflejadas en la sección anterior en una metodología de desarrollo de juegos serios completo. La Figura 15 muestra el diagrama que contempla todo el proceso, que abarca desde la incepción inicial del juego, hasta su desarrollo, despliegue y evaluación final.

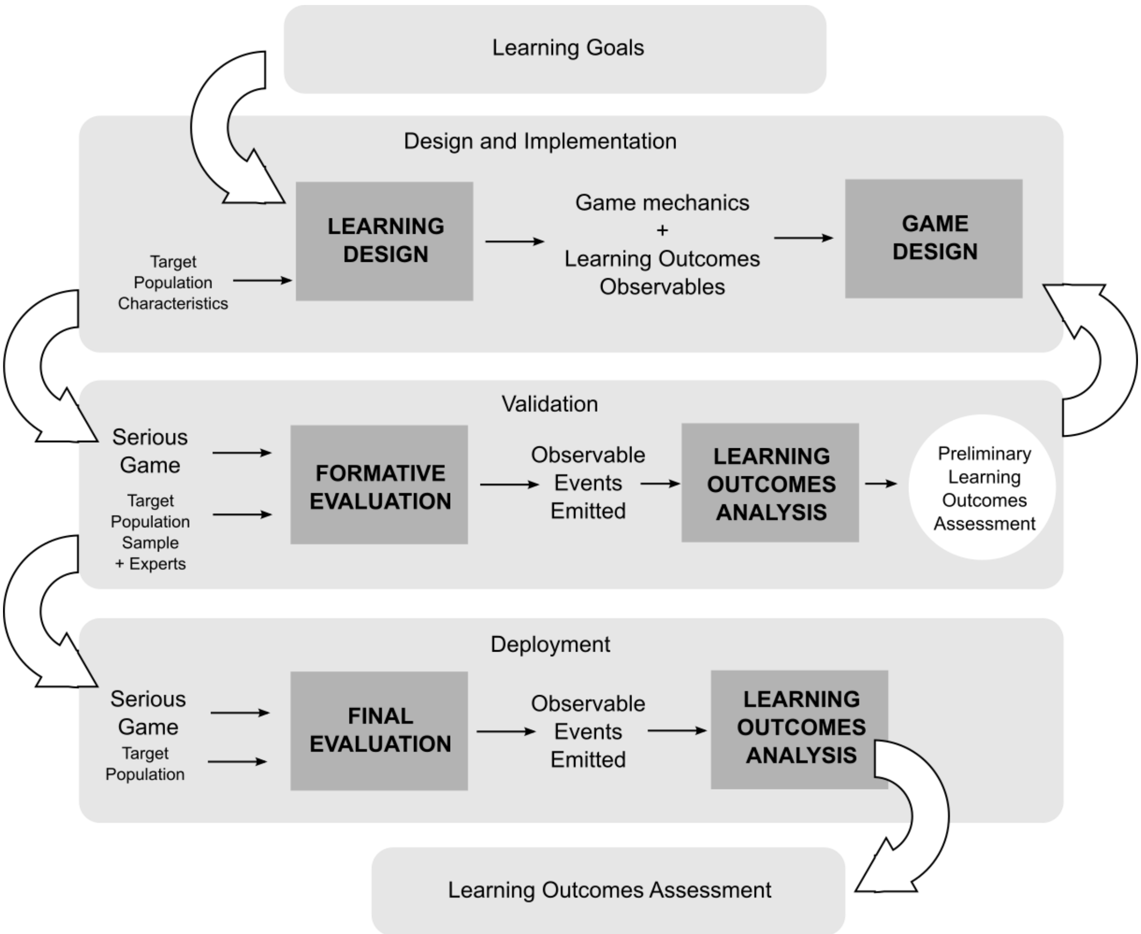


Figura 15. Visión completa de la metodología para diseñar, desarrollar y desplegar un juego serio.

El proceso comienza con el establecimiento de los objetivos instruccionales del juego. Éstos dan lugar a un diseño educativo, que además tiene en cuenta el público objetivo al que irá dirigido el juego. A partir de ir se establecen las mecánicas de juego y los observables que serán necesarios para medir conocimiento, formando el diseño de juego. Se realiza una primera implementación del juego serio, que pasa por una evaluación formativa. Esta evaluación la realizan expertos en el campo de conocimiento que el juego imparte, más una pequeña muestra del público objetivo, y su intencionalidad es descubrir errores (técnicos o instruccionales) y refinar el juego. En esta evaluación formativa también se capturan señales de interacción y se realizan las primeras evaluaciones de aprendizaje, valorando su idoneidad y utilidad para los instructores. Con todos estos resultados se reevalúan todos los aspectos del juego serio y se itera.

Una vez el juego serio ha sido validado, se despliega por completo para el público objetivo. Los estudiantes interactúan con él, y el sistema de análisis empieza a recibir resultados de aprendizaje que pueden ser consultados por los instructores.

Además del proceso general de desarrollo y despliegue de un juego serio, en este artículo también se define un patrón de juego que estructura como deben introducirse nuevos conceptos educativos, atendiendo a las necesidades técnicas requeridas para la evaluación de aprendizaje. La Figura 16 muestra este patrón.

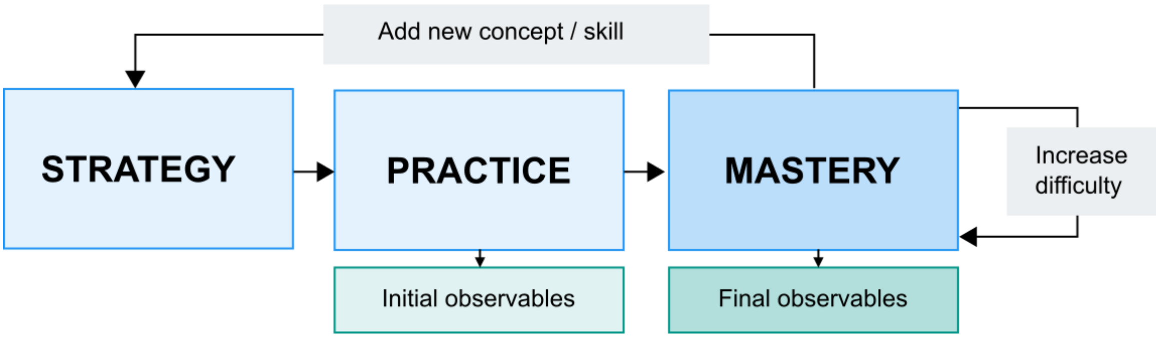


Figura 16. Fases del patrón de diseño de juego para la introducción de conceptos. Durante la fase de práctica *(practice)* y maestría *(mastery)* se emite observables para la evaluación de aprendizaje.

Este patrón divide la introducción en el juego de cada nuevo concepto en tres fases:

- **Estrategia:** En esta fase el juego debe ofrecer información básica sobre el nuevo concepto o habilidad, en qué consiste y cómo puede empezar a practicarlo.
- **Práctica:** En esta fase el juego propone una serie de desafíos iniciales en los que el jugador debe utilizar el nuevo concepto o habilidad. Todos los aspectos necesarios para este habilidad son expuestos durante esta fase, que se desarrollará en un entorno seguro donde el jugador puede fallar. Esta fase emitirá interacciones que serán recogidas para hacer una evaluación inicial del conocimiento del jugador.
- **Maestría:** En esta fase el juego propone una serie de desafíos con un extra de dificultad, en el que jugador debe demostrar, sin ayuda del juego, que domina el concepto o habilidad adquirido durante la fase de práctica. Esta fase emitirá interacciones que serán utilizadas para hacer una evaluación final del conocimiento del jugador.

El patrón define dos momentos de evaluación para el estudiante, durante la fase de práctica y durante la fase de maestría. Estas evaluaciones nos ofrece dos medidas: una con el conocimiento inicial del jugador (CI), y otra con su conocimiento final (CF). Estas dos medidas nos permiten hacer una evaluación de aprendizaje individualizada para cada estudiante, que pueden ser clasificados en cuatro categorías:

- **Aprendices**: si CI es bajo y CF es alto, significa que el jugador cometió errores en la fase de práctica, pero que aprendió de ellos y demostró su aprendizaje en la fase de maestría. Idealmente, si el juego tiene un buen diseño y se aplica a un público objetivo adecuado, la mayoría de estudiantes deberían caer en esta categoría.
- **Maestros**: si ambos CI y CF son altos, significa que el jugador sobrepasó todos los escollos propuestos por el juego sin mucha dificultad. En este caso, no se puede afirmar que el estudiante aprendiera con el juego serio, aunque los resultados abalan que posee los conocimientos que el juego pretende impartir.
- **No-aprendices**: si ambos CI y CF son bajos, significa que el jugador encontró problemas durante todo el desarrollo del juego, y que no aprendió nada con él.
- **Atípico**: si CI es alto y CF es bajo, significa que el jugador sobrepasó la fase de práctica sin problema, pero que no logró completar con éxito la fase de maestría. Si muchos jugadores caen en esta categoría podría indicar un diseño de niveles poco balanceado.

De este trabajo extraemos las siguientes contribuciones.

1. Un modelo general para el desarrollo de juegos serios enfocado a evaluación.
2. Un patrón de diseño de juegos que permite la clasificación de jugadores por su perfil de aprendizaje.

4.3. Modelo de datos para analíticas de aprendizaje

4.3.1. Captura de datos ingenua

La primera aproximación al modelado de datos a capturar en un juego serio se hace en *A framework to improve evaluation in educational games* (sección 7.2). En este artículo se hace una primera categorización de todos los eventos coleccionables de un juego serio:

- **Eventos crudos de interacción**: comprende todos los tipos de eventos producidos por los dispositivos de entrada del juego, tales como clicks de ratón, pulsaciones de teclado, controles de joystick, etc.
- **Evolución del estado de juego**: comprende todas aquellas variables que definen el estado de juego y su evolución en el tiempo.
- **Eventos lógicos**: comprende todos aquellos eventos que pueden tener especial relevancia para el avance del juego, como completar una fase, perder una vida, completar un objetivo, etc.

De esta primera aproximación obtenemos una primera definición, a grandes rasgos, de los tipos de eventos que pueden recogerse de un juego serio.

4.3.2. Captura de eventos significativos

En *Application of Learning Analytics in educational videogames* (sección 7.3), partiendo de las trazas genéricas propuestas en el trabajo anterior, se propone un conjunto de trazas "universales". Se usa el término universal para denotar su aplicabilidad a cualquier juego serio. Los tipos de trazas definidos son:

1. **De juego**: eventos que tenga que ver con el flujo de ejecución global del juego. Estos pueden ser:
    a. **Empezar**: se produce cada vez que el jugador comienza una nueva sesión de juego.
    b. **Abandonar**: se produce cuando el jugador abandona en el juego sin haberlo completado.
    c. **Completar**: se produce cuando el jugador completa el juego hasta su final.
2. **De nivel**: eventos relacionados con el progreso de los jugadores dentro de las fases y niveles definidos para el juego.
    a. **Empezar**: el jugador comienza un nuevo nivel. El evento contiene un identificador único para el nivel.
    b. **Completar**: el jugador completa un nivel. El evento contiene un identificador único para el nivel.
3. **De variables:** eventos que contengan la evolución de los valores de las variables significativas del juego, entendiendo por significativas aquellas variables que tienen especial relevancia a la hora de caracterizar el estado de juego (como puntuación, número de intentos restantes, etc.).
    a. **Actualizar**: una variable significativa ha actualizado su valor por algún evento producido dentro del juego. El evento contiene un identificador único de variable y su nuevo valor.
4. **De interacción:** eventos directamente producidos por la interacción de los jugadores con los dispositivos de entrada del juego serio.
    a. **Interacción:** un evento de interacción con todos los datos relevantes del evento, por ejemplo, el botón pulsado en ratón/teclado/joystick.

La propuesta es que cada vez que se produzca uno de estos eventos dentro del juego, se emita una señal con los datos apropiados, añadiendo información extra que pueda resultar relevante para los futuros análisis, como por ejemplo una marca de tiempo.

La principal contribución de este trabajo es la definición de un modelo común sienta las bases de un conjunto predefinido de eventos. Si varios juegos serios empiezan a emitir señales en el mismo formato y con el mismo tipo de datos, aquellas visualizaciones y análisis que se desarrollen para ellos podrán compartirse en juegos.

4.3.3. Estandarización

Una vez el modelo ha sido madurado a través de su validación por casos de uso (sección 4.4), el siguiente paso es formalizarlo y ponerlo al uso de terceros a través de su especificación formal. Para ello, se utiliza el estándar más adecuado tras valorar las especificaciones recopiladas en la sección 2.4.

En *Applying standards to systematize learning analytics in serious games* (sección 7.7), primero se define, de manera abstracta, los usos y atributos de que cada posible tipo de evento emitido por un juego serio, y después se presenta la implementación del modelo sobre la especificación xAPI. Este trabajo se realiza en cooperación con ADL, propulsora de la especificación xAPI, y culmina con la publicación del perfil oficial para juegos serios de xAPI[13].

La Figura 17 muestra uno de los eventos definidos por la especificación JSON. En este caso, la *frase* representa un usuario identificado lanzando un juego serio. Éste sería uno de los primeros eventos emitidos en una sesión de juego normal. Los posibles verbos, objetos y extensiones definidos para xAPI en este perfil de juegos serios pueden encontrarse tanto en la sección 6.7, dentro del artículo, como en la web del perfil.

```
{
  "actor": {
    "name": "John Doe",
    "mbox": "mailto:johndoe@example.com"
  },
  "verb": {
    "id": "http://adlnet.gov/expapi/verbs/initialized",
    "display": { "en-US": "initialized" }
  },
  "object": {
    "id": "http://rage.e-ucm.com/activities/Countrix",
    "definition": {
      "name": { "en-US": "Countrix Serious Game" },
      "type:" "https://w3id.org/xapi/seriousgames/activities/serious-game"
    }
  }
}
```

Figura 17. Representación en JSON de un evento de iniciación de juego serio en formato xAPI.

La principal contribución de este trabajo es el desarrollo de perfil de juegos serios para xAPI, y su publicación como parte oficial del repositorio de perfiles de ADL.

4.4. Casos de estudio

4.4.1. La gran fiesta y Donaciones

La gran fiesta es un juego serio desarrollado por Ángel del Blanco, como parte del trabajo realizado por el grupo e-UCM para el proyecto EDUWAI[14]. *La gran fiesta* es un juego en

[13] http://xapi.e-ucm.es/vocab/seriousgames
[14] http://www.asi-soft.com/?q=es/content/eduwai

primera persona, cuyo objetivo educativo de este juego es mostrar a personas con dificultades en su desarrollo cognitivo algunos de los hábitos saludables que deben mantener en su vida diaria, por ejemplo, relacionados con la higiene o con las relaciones sociales. En la narrativa de juego, el jugador debe prepararse y asistir a una fiesta en su empresa.

Este juego fue desarrollado con la plataforma e-Adventure, y fue uno de los primeros en beneficiarse de la integración nativa de captura de datos explicada en la sección 4.1.1. En este caso, la captura de datos era especialmente relevante, porque dado el perfil de sus jugadores, resultaba más dificultoso recoger datos de su experiencia a través de cuestionarios.

En *Application of Learning Analytics in Educational Videogames (*sección 7.3) se detalla el experimento realizado con este juego serio, en el que participaron 19 usuarios. Como primer experimento dentro del marco de esta tesis, sirvió para probar la implementación inicial del sistema de analíticas. También sirvió para generar algunas visualizaciones que mostraban como los usuarios interactuaban con el juego. La Figura 18 muestra una de esas visualizaciones. En ella se muestra un mapa de calor donde se acumulan los clicks que los jugadores realizaban sobre los elementos de una de las escenas de juego. El objetivo de esta visualización era confirmar que los jugadores localizaban los elementos interactivos de la escena.

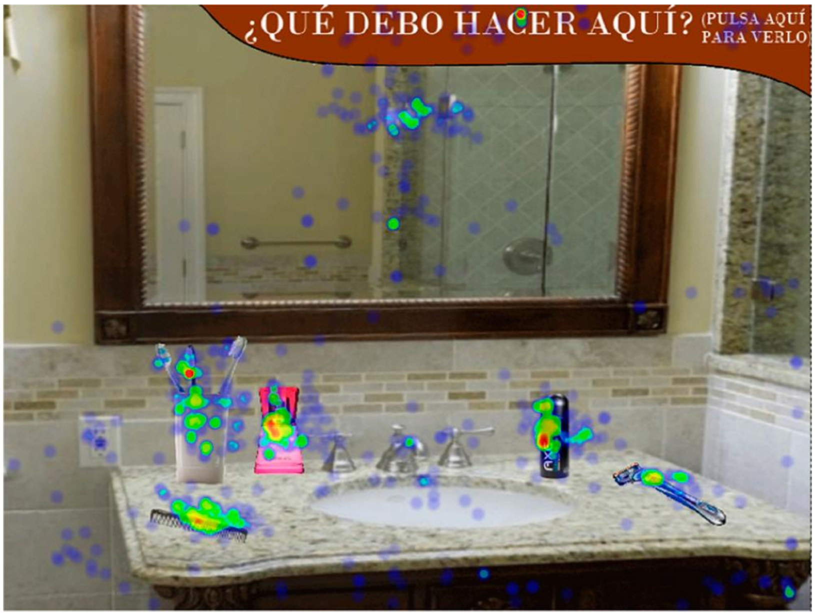


Figura 18. Mapa de calor que muestra una vista agregada de todos los clicks realizados por los jugadores de *La gran fiesta* en una de sus escenas.

Donaciones es un juego serio implementado por Blanca Borro, también desarrollado con e-Adventure (Borro Escribano, Del Blanco, Torrente, Borro Mate, & Fernandez Manjon, 2015). El objetivo principal de este juego es entrenar a miembros de la Organización Nacional de Trasplantes (ONT) sobre el proceso que se debe seguir en la donación de un órgano. En este caso, el jugador se ponía en la piel de un médico dentro de una unidad de trasplantes, y era instruido y después testeado sobre los pasos del proceso de donación.

Al igual que *La gran fiesta,* fue uno de los primeros juegos en utilizar la integración nativa de captura de datos en e-Adventure, y con él se realizó un experimento con 150 usuarios. Este experimento también sirvió para probar la implementación inicial del sistema de analíticas, y las lecciones aprendidas con él se recogen en *Learning Analytics and Educational Games: Lessons Learned from Practical Experience* (sección 7.6).

4.4.2. *Lost in Space*

Building a Scalable Game Engine to Teach Computer Science Languages (sección 7.1) recoge el proceso de diseño y desarrollo del juego serio *Lost in Space.* Este juego fue desarrollado desde cero por el doctorando. Como el trabajo ya contaba con dos juegos desarrollados en e-Adventure con mecánicas de juego simliares (juego de aventura *point & click* en primera persona), se decidió crear un juego de un género completamente distinto para ampliar la aplicación del trabajo realizado.

En este caso se escogió el género de puzle, y se desarrolló un juego cuyo objetivo era enseñar lenguajes instruccionales a estudiantes de un curso de tecnologías web. La narrativa coloca al jugador a los mandos de nave especial que debe llevar hasta la salida del nivel evitando diferentes tipos de obstáculos. Para ello, debe introducir comandos en una consola utilizando un lenguaje instruccional (por ejemplo, XML). Estos comandos se traducen en órdenes que guían a la nave por el nivel. La Figura 19 muestra un esquema de uno de los niveles a superar en el juego.

El juego *Lost in Space* se utilizó en varios experimentos. En *Building Scalable Game Engine to Teach Computer Science Languages* se cuenta un experimento realizado con alumnos de dos perfiles: estudiantes de ingeniería y estudiantes de ciencias sociales. El juego se utilizó con 36 alumnos y se probó, mediante el uso de cuestionarios antes y después del experimento, que el juego enseñaba a ambos perfiles de alumnos. La implementación de este juego también ayudó a clarificar a algunos aspectos necesarios en el diseño de juego serios (sección 4.2.1).

Posteriormente, el juego se instrumentalizó para la captura de datos y se utilizó en experimentos subsecuentes. Uno de ellos, contado con detalle en *Applying learning analytics to simplify serious games deployment in the classroom* (sección 7.4), consistió en sustituir una clase teórica sobre el lenguaje de marcado XML por el juego. El juego, integrado con la plataforma de analíticas de aprendizaje, proveyó al instructor de métricas y alertas del progreso de sus alumnos con el juego (sección 4.1.3).

El código completo del juego puede encontrarse en el siguiente repositorio: https://github.com/anserran/lostinspace/commits/master

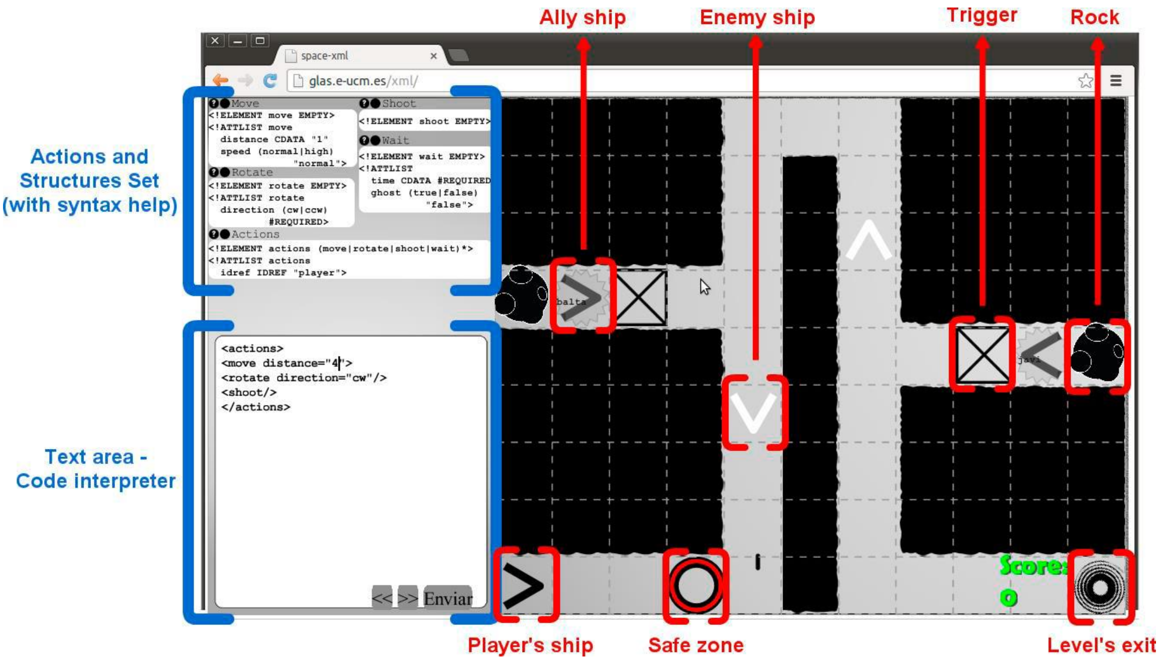


Figura 19. Esquema de uno de los niveles que los jugadores deben resolver en el juego serio *Lost in Space.*

4.4.3. La Dama Boba

La Dama Boba es un juego serio implementado por Borja Manero haciendo uso de la plataforma e-Adventure (Manero, Fernández-Vara, & Fernández-Manjón, 2013a), inspirado en la obra de teatro homónima. El juego formó parte de un experimento que involucró a 7 instituciones educativas y más de 350 de alumnos (Manero et al., 2015). En *La Dama Boba* el jugador encarna a uno de los protagonistas de la obra de teatro, y su misión es ir resolviendo puzles que lo van guiando por el argumento de la obra. El objetivo de este juego serio es doble: por un lado, pretende incrementar el interés de alumnos de secundaria por el teatro clásico; y por otro, impartir varios conceptos relacionados con la poesía, la literatura y la gramática.

La Dama Boba aplica la metodología descrita en la sección 4.2.2, y en el artículo *A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes* (sección 7.5) se muestran los resultados de evaluación de aprendizaje obtenidos por este método, en consonancia con la plataforma de analíticas de aprendizaje descrita en la sección 4.1.

La metodología clasificó a la gran mayoría de participantes como *maestros*, es decir, la mayoría de alumnos obtuvo muy buenos resultados tanto en la fase de práctica como en la fase de maestría. Esto pudo deberse a que la captura de señales realizada durante la fase de práctica no fue la suficientemente amplia, y el juego falló en capturar el conocimiento inicial de los estudiantes. Este factor se vio propiciado por la necesidad de que el juego pudiera completarse en 40 minutos o menos, factor necesario para su despliegue en la duración de una clase de secundaria normal y que obligó a acortar la fase de práctica.

4.4.4. *Countrix*

El juego *Countrix* se presenta en *Applying standards to systematize learning analytics in serious games* como demo técnica del uso de la especificación de xAPI. *Countrix* es un juego serio, también desarrollado desde cero por el doctorando, de preguntas y respuestas sobre Geografía. El jugador comienza cada sesión de juego con un tiempo límite. En ese tiempo debe contestar al mayor número de preguntas posibles. La Figura 20 muestra dos capturas de pantalla del juego en acción.

El objetivo de este juego era tener un ejemplo de juego serio sencillo que hiciera uso de todo el vocabulario esbozado el perfil xAPI de juegos serios. El juego, además, cuenta con un visor de *frases* que permite ver todos los eventos xAPI emitidos durante la partida.

El código completo del juego puede descargarse del siguiente repositorio: https://github.com/e-ucm/countrix.

El juego también está disponible para su descarga e instalación en dispositivos Android: https://play.google.com/store/apps/details?id=com.anserran.countrix.

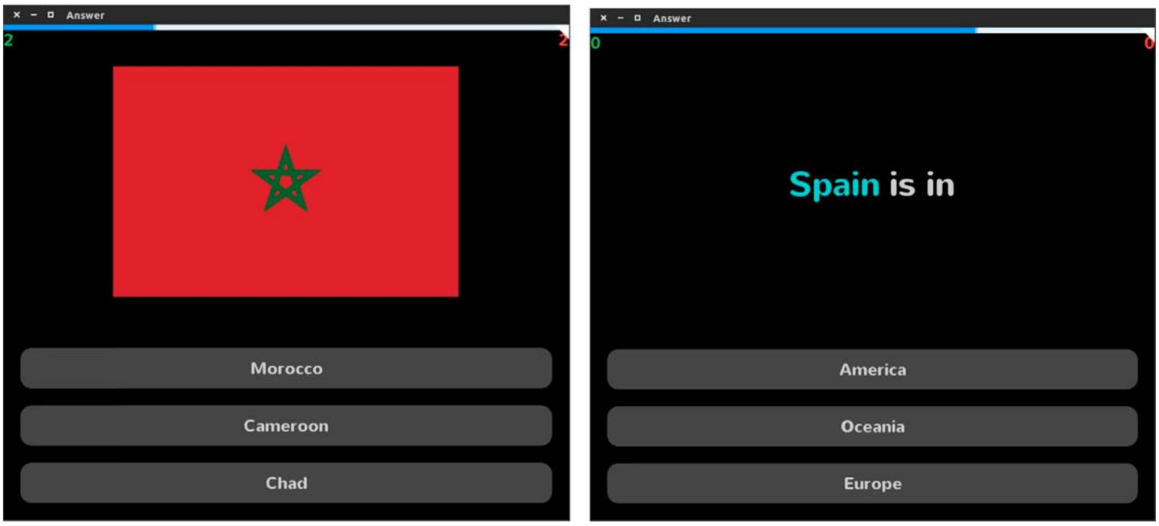


Figura 20. Captura de pantalla de dos tipos de preguntas que muestra el juego serio *Countrix*.

4.4.5. Conclusiones prácticas de los casos de estudio

Para finalizar, el artículo *Learning Analytics and Educational Games: Lessons Learned from Practical Experience* (sección 7.6) recoge algunas de las practicidades descubiertas durante el desarrollo de los casos de estudios recogidos en esta sección. Este artículo repasa, a modo de resumen, algunas de las cosas a tener en cuenta en el despliegue de juegos serios con analíticas de aprendizaje. Las más destacadas son:

- **Análisis a dos niveles**: siempre que sea posible y se esté realizando un experimento de validación de un juego serio, conviene suplementar la captura de datos del sistema de analíticas con aproximaciones más tradicionales como cuestionarios escritos. Este permite tener dos fuentes de datos independientes a la hora de evaluar cualquier tipo de resultado, así como comparar resultados.
- **Duración del experimento**: uno de las practicidades que nos encontramos a la hora de realizar experimentos en diferentes centros educativos es que su duración no podía prolongarse más 50 minutos (duración estándar de clases tanto educación universitaria como educación secundaria). Eso conllevaba planear el diseño del juego serio a una duración menor de 50 minutos, sin olvidar dejar tiempo para dar la instrucciones necesarias a los alumnos (no se ponían a jugar en el minuto 0) y para posibles cuestionarios antes y/o después del experimento.
- **Maximizar el control para reducir errores**: Los investigadores deben asegurarse de que controlan todos los aspectos envueltos en el desarrollo del experimento. En algunos casos, las instituciones educativas o los profesores pueden afectar de manera intencionada al desarrollo del experimento. Por ejemplo, algunos profesores podían tener la sensación de que los resultados de su clase podían reflejar su trabajo como profesor, e intentaban a ayudar a sus alumnos durante los experimentos, lo que desvirtuaba sus resultados.
- **Plan de emergencia para captura de datos**: en algunos centros educativos la conexión a Internet era lenta o tenía cortes intermitentes. Esto ponía en riesgo la integridad de las interacciones enviadas desde los juegos al servidor de analíticas. Por ello, en todos los experimentos se hacía un envío de trazas doble: por un lado, se enviaban las trazas al servidor y por otro se hacía una copia local de las trazas, a la que se recurría si la conexión a Iternet fallaba.

Capítulo 5. Conclusiones y trabajo futuro

En este capítulo se presenta, a modo de resumen, las principales contribuciones del trabajo realizado en esta tesis. Se concluye con una sección de trabajo futuro, en el que se repasan las líneas de investigación que esta tesis deja abiertas.

5.1. Conclusiones y principales contribuciones

La principal contribución de esta tesis es la propuesta de un marco de trabajo general que permite aplicar analíticas de aprendizaje a los juegos serios. Este marco se acotó en una tarea concreta: crear una arquitectura de analíticas de aprendizaje centrada en asistir a profesores e instructores en la evaluación de aprendizaje haciendo uso de juegos serios. Dentro de este ámbito, el trabajo se ha centrado principalmente en tres aspectos:

1. Diseño de juegos serios: estructuración del aprendizaje, diseño de mecánicas de juego y requisitos para enlazar y aplicar el diseño con las analíticas de aprendizaje.
2. Arquitectura de análisis: procesos necesarios para habilitar las analíticas de juegos serios. Además se ha trabajado en definir un modelo de evaluación en juegos serios.
3. Modelado de datos de captura: dentro de todo el proceso de análisis, se ha hecho especial hincapié en el paso de selección y captura de datos, definiendo y formalizando un modelo de eventos a capturar en juegos serios, base para habilitar el proceso de análisis del aprendizaje.

El trabajo desarrollado para esta tesis ha creado diferentes contribuciones al campo de los juegos serios y las analíticas de aprendizaje, que pasamos a resumir en las siguientes secciones.

5.1.1. Estudio del dominio

En el capítulo 2 presentamos un amplio resumen del estado del dominio implicado en este trabajo de tesis. Uno de los principales desafíos residía en la revisión de múltiples dominios colindantes y relacionados con los objetivos de la tesis. Algunos de ellos no habían intersectado con frecuencia en la literatura.

Por un lado, revisamos las analíticas de aprendizaje. Un campo emergente que se había aplicado a múltiples contenidos educativos, pero que aún no había hecho muchas incursiones en el mundo de los juegos serios. De esta revisión se extrajeron las siguientes conclusiones:

- El proceso de analíticas de aprendizaje sigue los mismos pasos que cualquier otro tipo de análisis: captura, análisis, obtención de resultados y acción.
- La diferencia está, entonces, en el objetivo de aplicación. Las analíticas de aprendizajes podían ayudar a estudiantes, profesores o estamentos

institucionales, y sus análisis siempre estaban enfocados a mejorar el proceso educativo en alguno de sus aspectos.

- Tras revisar el trabajo realizado con diferentes contenidos educativos, la gran mayoría de ellos estaban dedicados a evaluar y predecir los resultados de alumnos.

A partir de estas conclusiones se decidió centrar los esfuerzos de la tesis en la evaluación de resultados centrada en profesores e instructores. Las razones para no centrarse en el resto de posibilidades también se basaron en el estudio del dominio. Por un lado, la evaluación de resultados era lo más relevante para profesores. Esta evaluación también podría extenderse a alumnos (cubriendo así, a 2 de los 3 tipos de actores envueltos en el proceso de analíticas de aprendizaje). Debido al carácter focalizado de los juegos serios, parecía improbable que sus resultados pudieran ser utilizados de manera directa por los estamentos administrativos e institucionales de las entidades educativas, por eso se les descartó como público objetivo de los resultados de esta tesis.

También fue necesario hacer un repaso de cómo se estaba estructurando la impartición de contenido educativo dentro de los juegos serios. Se descubrieron múltiples publicaciones definiendo marcos de trabajo para el diseño de juegos serios, sin embargo, muy pocos de ellos tenían en cuenta los requisitos para la evaluación de aprendizaje a través de la captura de eventos.

Con este repaso se detectó la necesidad de desarrollar una metodología de diseño de juegos serios que tuviera en cuenta la evaluación de resultados a partir de interacciones de usuario. La metodología debía ir más allá del propio del diseño de juego, también debía contemplar cómo integrarse con una plataforma de análisis de datos.

En el estudio del dominio también se repasaron las disciplinas de análisis de datos más relacionadas con juegos serios y analíticas de aprendizaje. Este repaso sirvió para fundamentar algunas de las ideas que serían implementadas después. También se revisaron las plataformas de análisis de datos disponibles en su momento. De este repaso se concluyó que ninguna cumplía las características necesarias para desarrollar los objetivos de la tesis, y se decidió construir una plataforma desde cero.

Finalmente, se quería evitar que todo el trabajo realizado fuera un desarrollo aislado que empezara y acabara en el marco de esta tesis. Por ello, se decidió explorar la posibilidad de implementar algunos de sus aspectos del trabajo con estándares ampliamente utilizados en el mundo educativo. Así, se estudiaron algunos de los estándares disponibles en su momento y se decidió utilizar una especificación que cubriera al menos el proceso de captura de datos. Tras la revisión, xAPI destacó por su flexibilidad y su carácter abierto, y decidió utilizarse para formalizar la captura de datos.

5.1.2. Plataforma para el análisis de aprendizaje en juegos serios

Uno de los principales retos de esta tesis era el diseño y construcción de una plataforma de análisis que diera soporte a los experimentos desarrollados en el marco de la propia tesis. Idealmente, esta plataforma tendría la suficiente generalidad como para poder ser utilizada por terceros.

Esta tesis define un marco teórico, basado en una arquitectura software, para la aplicación de analíticas de aprendizaje a juegos serios. Analiza el proceso de analíticas de aprendizaje desde cero y adapta cada una de las etapas de este proceso para aplicarlo a juegos serios.

Además del trabajo teórico, que ha sido repasado con más detalle en la sección 4.1, queremos destacar también como contribución la parte más técnica de este desarrollo, ya que se ha creado una implementación de referencia.

La plataforma que implementa todo el marco teórico propuesto recibió el nombre de *Gleaner (Game LEarning ANalytics for Educational Research)*. Esta es una plataforma potente y escalable que utiliza tecnologías como NodeJS[15], MongoDB[16], Apache Kafka[17] o Apache Storm[18] para construir un servicio capaz de recibir, analizar y visualizar interacciones procedentes de juegos serios.

De manera complementaria, se ha integrado un rastreador (tracker) en el motor de juegos e-Adventure. Esto permite que cualquier juego desarrollado con esta herramienta pueda tener comunicación directa con Gleaner. Además, también se ha implementado una librería, disponible para Java, Android y JavaScript[19], que permite que cualquier juego pueda también comunicar sus interacciones a Gleaner.

Finalmente, como validación final de la plataforma, queremos destacar que ha sido utilizada como arquitectura e implementación de referencia para analíticas de aprendizaje y juegos serios en el proyecto de financiación europea FP7 NoE GALA (Hauge et al., 2014b).

5.1.3. Metodología de diseño de juegos serios centrada en analíticas de aprendizaje

La capacidad educativa de los juegos serios ha sido probada de manera empírica por numerosos experimentos encontrados en la literatura. Sin embargo, aún existen dudas sobre los procesos de aprendizaje que se dan dentro de los juegos serios, y cuál es la mejor manera de evaluarlos.

---

[15] https://nodejs.org/es/
[16] https://www.mongodb.com/es
[17] https://kafka.apache.org/
[18] http://storm.apache.org/
[19] https://github.com/e-ucm/libgdx-tracker

El trabajo de esta tesis ha intentado aportar algo de luz a esta cuestión pendiente aplicando analíticas de aprendizaje. Los videojuegos son medios altamente interactivos por naturaleza. Sea cual sea el aprendizaje que se desprende del uso de un juego serio, el proceso debe estar reflejado en las decisiones e interacciones que realizan los estudiantes con ellos. Estos procesos ya han sido estudiados dentro de la industria del videojuego, así que parece lógico aplicarlos a los juegos serios.

Esta tesis ha propuesto una metodología que formaliza el proceso de diseño, implementación y despliegue de un juego serio, enfocado a evaluación del aprendizaje. Para esa metodología ha tomado principios de las analíticas de juego (como la insistencia en la captura de datos directamente desde el juego), pero también de las aproximaciones a la validación de contenidos más académica (como la validación de los objetivos instruccionales o las evaluaciones formativas).

5.1.4. Modelo de datos para captura de eventos en juegos serios

En esta tesis se ha hecho un esfuerzo para que terceros puedan utilizar algunos de los resultados obtenidos de esta tesis de manera directa. Uno de esos esfuerzos se ha centrado en la modelación del proceso de captura de datos, que ha culminado con una especificación completa del modelo utilizando Experience API.

En esta tesis se ha propuesto un modelo de captura de datos para juegos serios, centrado en los eventos más significativos que pueden darse una sesión de juego. En el contexto de este modelo, entendemos significativo como aquellos eventos que puedan tener relevancia para la evaluación del aprendizaje.

Los eventos planteados tienen una voluntad generalista, es decir, están diseñados para que cualquier juego serio pueda utilizarlos, sea cual sea su contenido educativo, su género o sus mecánicas de juego.

Una vez se validó la utilidad del modelo a través de su uso en varios casos de uso, se pasó a la formalización del mismo haciendo eso de la especificación xAPI (hasta el momento, se había utilizado un formato JSON propio para representar las trazas de juego).

Para el desarrollo de esta implementación se contactó con la institución ADL, que es el ente dependiente del Departamento de Defensa de los Estados Unidos encargados de la especificación, y logramos su colaboración en la elaboración de un perfil xAPI de juegos serios (ADL fueron los responsables de la creación de SCORM que es el estándar mas utilizado en e-learning).

Como contribución final de esta línea de investigación se ha obtenido un perfil oficial de xAPI para juegos serios, así como la creación de una comunidad de prácticas, centrada en juegos serios, encabezada por miembros del grupo e-UCM y bajo el ala de las comunidades de práctica de ADL y xAPI[20]. Este trabajo se ha mencionado como uno de los relevantes en xAPI[21].

5.1.5. Validación por casos de estudio

Finalmente, en esta tesis se ha hecho un énfasis especial en que todos los avances teóricos que se fueran realizando estuvieran abalados por su correspondiente caso de estudio. Por lo tanto, los resultados de este trabajo se han visto integrados con diferentes juegos serios.

La importancia de estos casos de estudio no residía tanto en el número total de juegos serios integrados con el sistema de analíticas, sino en lo que cada uno de ellos aportaba como novedad al total del trabajo. Uno de los elementos más definitorios de los videojuegos es cada uno crea su propio mundo, con diferentes estéticas, géneros y mecánicas de juego. Por ello, se intentó buscar juegos cuyo contexto, dominio de conocimiento, género y nivel de educación variara, para así cubrir un espectro más amplio.

La tabla contenida en la Figura 21 muestra un resumen de todos los juegos con los que se ha integrado el trabajo de esta tesis, así como su dominio, género, plataforma de desarrollo y nivel educativo.

Título	**Dominio educativo**	**Género / Mecánicas**	**Plataforma de desarrollo**	**Nivel educativo**
La gran fiesta	Ciencias sociales	Aventura gráfica en primera persona	e-Adventure	Para personas con problemas en su desarrollo cognitivo (Down)
Donaciones	Medicina	Aventura gráfica en primera persona / Pregunta y respuesta	e-Adventure	Para doctores que están ejerciendo
Lost in Space	Ingeniería / Programación	Puzle	Desarrollo nativo en HTML5	Estudiantes universitarios
La Dama Boba	Literatura / Teatro	Aventura gráfica en tercera persona / Puzzle / Preguntas y respuestas	e-Adventure	Estudiantes de secundaria
Countrix	Geografía	Preguntas y respuestas	Desarrollo nativo en Android	Todos.

Figura 21. Tabla de juegos serios integrados con el trabajo desarrollado en esta tesis.

[20] https://www.adlnet.gov/serious-games-cop/
[21] http://www.adlnet.gov/xapi-year-in-review/

5.2. Trabajo futuro

El trabajo de esta tesis empezó desde una aproximación general que pretendía aunar analíticas de aprendizaje y juegos serios. Conforme el estudio del domino y el trabajo fue avanzando, se identificaron numerosas posibilidades que habrían de quedar fuera para mantener el ámbito de la tesis lo más focalizado posible.

Afortunadamente, el trabajo de esta tesis no acaba con este doctorando. Algunas de sus resultados han servido como base para la concesión de proyectos de investigación a nivel europeo, tales como H2020 RAGE y H2020 BEACON. Estos proyectos continuarán y ampliarán parte del trabajo desarrollado, dando lugar a nuevas líneas de investigación.

Algunas de las líneas más importantes (tanto técnicas como teóricas) y que quedan abiertas para el futuro son:

- Compleción y finalización de Gleaner: aunque este proyecto deja una versión muy avanzada con un conjunto de características suficiente, queda pendiente su culminación como un producto software completo, que tenga en cuenta consideraciones adicionales que tuvieron que ser minimizadas durante el desarrollo de la tesis, como por ejemplo las relacionadas con seguridad y escalabilidad. Este aspecto ya está en desarrollo en el grupo e-UCM ya que se va a utilizar dentro de los pilotos de gran dimensión de los proyectos H2020 RAGE y BEACONING.
- Estudiar el efecto de juegos serios de larga duración a través de las analíticas de aprendizaje: esta tesis se ha centrado en juegos serios que se aplican una única vez durante un tiempo relativamente corto (máximo de 1 hora). Aunque esto ha servido para sentar las bases del análisis de resultados, aún queda explorar como la evaluación de aprendizaje se extiende a videojuegos con los que estudiantes deban interactuar múltiples veces a lo largo del tiempo.
- Trabajar más con profesores para entender mejor sus necesidades: Lamentablemente, en los casos de uso propuestos en esta tesis, los investigadores representaban papeles múltiples. Eran diseñadores, desarrolladores e instructores. Una colaboración más cercana con profesores y sus necesidades dentro de una clase pueden ayudar a identificar nuevos tipos de visualizaciones e intervenciones que sean más útiles para ellos.
- Avanzar en la definición de un modelo que tenga en cuenta las mecánicas de juego: el modelo de captura de datos actual tiene un carácter generalista. El siguiente paso es extenderlo para capturar las diferentes mecánicas de juego que pueden existir. Esto dará lugar a nuevos tipos de eventos, pero también a nuevos tipos de análisis y visualizaciones. Este aspecto ya está parcialmente en desarrollo ya que

hay un trabajo de fin de master en el que se está extendiendo el modelo para juegos geoposicionados.

- Extender la integración de analíticas de aprendizajes y juegos serios en diversos entornos de aprendizaje: La integración de captura con xAPI es un primer paso hacia la integración total de juegos serios en entornos virtuales de aprendizaje, pero aún quedan otras tareas relacionadas con el despliegue y con la integración de los juegos en contextos educativos más amplios.

Capítulo 6. Artículos presentados

A continuación se incluyen los artículos editados que se aportan como parte de esta tesis doctoral.

6.1. Building a Scalable Game Engine to Teach Computer Science Languages

6.1.1. Cita completa

Ángel Serrano-Laguna, Javier Torrente, Borja Manero, Baltasar Fernández-Manjón (2015), ***Building a Scalable Game Engine to Teach Computer Science Languages***, IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, Vol. 10, Issue 4, pp. 253-261. DOI: 10.1109/RITA.2015.2486386

6.1.2. Resumen original de la contribución

Everyday more people are interested in learning Computer Science (CS), either to improve their skill set to apply for new jobs or just for personal growth. The sector of the population looking for instruction on these subjects has increased and diversified. We need new tools that appeal to this wider audience, and game-based learning is one of the most promising approaches at the moment.

There is a need for more scalable game-based instruction paradigms, that can be easily adapted to different levels of complexity and content related to CS (different programming languages, different programming paradigms…). Throughout this paper we present a flexible and scalable architecture to create videogames for learning CS languages. The architecture is based on the idea that students control the game using small pieces of text written in some CS language. The keys of the scalability of our approach are: 1) it separates the CS language used to write the programs from the game design; and (2) the game model provides a system of levels that allows incremental learning of CS language structures.

As validation and implementation of our approach, we developed “Lost in Space”, an educational videogame to teach the XML markup language. In this game, the player travels through several levels guiding a spaceship introducing small pieces of XML in a text console. Players can move and rotate the ship among other power-ups that get unblocked as he or she advances in the game.

The game was tested with undergraduate students from computer science and social sciences, by comparing it with traditional instruction (i.e. a teacher with a slides presentation). Students who played the game were much more engaged than those who attended the lecture, showing a more active attitude throughout the whole experience and also spent more time practicing after class. Findings also suggest that the game was effective for instruction regardless of the background of the students. However, the

educational gain observed with the game-based instructional approach, although effective, was not significantly higher than traditional instruction. We think that our approach is adequate to introduce CS languages in general, as well as new programming languages, and seems to be more appealing to new comers than traditional instruction.

Building a Scalable Game Engine to Teach Computer Science Languages

Ángel Serrano-Laguna, Javier Torrente, Borja Manero, Baltasar Fernández-Manjón
ISIA Department, Complutense University of Madrid, Spain
({angel.serrano | jtorrente | balta} @fdi.ucm.es; borja@sip.ucm.es).

***Abstract*—Everyday more people are interested in learning Computer Science (CS), either to improve their skill set to apply for new jobs or just for personal growth. The sector of the population looking for instruction on these subjects has increased and diversified. We need new tools that appeal to this wider audience, and game-based learning is one of the most promising approaches at the moment.**

There is a need for more scalable game-based instruction paradigms, that can be easily adapted to different levels of complexity and content related to CS (different programming languages, different programming paradigms…). Throughout this paper we present a flexible and scalable architecture to create videogames for learning CS languages. The architecture is based on the idea that students control the game using small pieces of text written in some CS language. The keys of the scalability of our approach are: 1) it separates the CS language used to write the programs from the game design; and (2) the game model provides a system of levels that allows incremental learning of CS language structures.

As validation and implementation of our approach, we developed "Lost in Space", an educational videogame to teach the XML markup language. In this game, the player travels through several levels guiding a spaceship introducing small pieces of XML in a text console. Players can move and rotate the ship among other power-ups that get unblocked as he or she advances in the game.

The game was tested with undergraduate students from computer science and social sciences, by comparing it with traditional instruction (i.e. a teacher with a slides presentation). Students who played the game were much more engaged than those who attended the lecture, showing a more active attitude throughout the whole experience and also spent more time practicing after class. Findings also suggest that the game was effective for instruction regardless of the background of the students. However, the educational gain observed with the game-based instructional approach, although effective, was not significantly higher than traditional instruction. We think that our approach is adequate to introduce CS languages in general, as well as new programming languages, and seems to be more appealing to new comers than traditional instruction.

***Index terms*—Application software, educational technology, software architecture, computer programming, educational videogames**

I. Introduction

VIDEOGAMES have been used as successful educational tools in many different fields and topics: engineering and problem solving [1], learning foreign languages [2], health [3] and even theatre [4]. The idea of using videogames to support the learning of computer programming is not new, as it dates back almost to the origins of instructional games [5]. However, this interest seems to have reached a peak recently, with numerous advocators bringing the potential of digital game technology to the field of computer programming education, and new educational programs trying to introduce programming in early educational stages [6].

First, the popularity of digital games can attract talented people to computer science and software engineering, professions that are essential in today's economy. Second, digital games can help smooth the learning curve for novice programmers by providing a highly visual and motivating environment. Finally, learning programming since childhood can foster development of high level thinking and problem-solving skills.

As a consequence, there are numerous initiatives dedicated to facilitating the use of game technology for learning computer programming. Most of them target kids, although other software targets college students. The drawback of these systems is that they are hard to scale, and are usually devised for a specific target audience, educational goal and/or programming language, which makes it difficult to repurpose and reuse the software in different settings.

In this paper, we present a flexible and scalable game engine to create games for learning programming that could be used by students with different backgrounds, as an extended version of the paper published in [7]. The game engine facilitates educational game development by providing game mechanics that are appropriate for learning programming. It is scalable because (1) it separates the programming language being learnt from the game design, allowing for reuse of the games with different programming languages; and (2) it provides a system of levels that allows incremental learning of programming structures. The engine has been used to develop the game *Lost in Space,* which supports the learning of several programming languages. We have conducted two case studies with college students from different backgrounds to evaluate the effectiveness of the game for learning computer programming. The first case study involved a group of computer science students, with previous knowledge about programming. The second case study was conducted with social sciences students, with no programming background.

II. Related Work

Game technology has been long used to support learning programming. One of the preferred approaches is to use activities related to game design and development [8], [9]. For example, in [10] researchers use the state-machine nature of board games to simplify the introduction of some programming concepts.

In other approaches, students create simple program snippets that control characters and objects in a game environment, usually through a visual or simplified programming language. These snippets can have several purposes: beat a level in a game, create interactive content or animate a simple scene with several characters. The visual condition of the results obtained facilitates the feedback and gets students rapidly engaged in programming.

Multiple tools have been developed around this paradigm, such as GameMaker [11], Alice [12], Greenfoot [13], Microsoft's Kodu [14], Scratch [15] and other self-built systems [16]. The literature is full of experiences where this paradigm has been successfully applied. For example, De Kereki [17] reports effective use of Scratch as a motivational tool in an introductory programming course. In other example, Chen and Cheng [18] use videogame development as the core activity of their programming introduction course.

Other studies have explored the activity of playing digital games in programming courses, which is an approach more similar to ours. In many of these games students do not write programs: the videogames are used as mere containers of theoretical content about programming, in which the only goal is to pursue an increase in students' motivation and interest for the activity. However, the lack of active programming is a limitation of the approach, since writing programs is an essential activity to learn programming.

Several examples exist in the research using this approach. In [19] authors report the use of a role-playing game where quests are directly related to programming concepts (e.g., answering questions about some programming language feature). In a similar approach, the use of a game environment provided a significant increase in student motivation towards the subject [20] and student performance. In [21] authors propose two mini-games: a typing game and a fill-in-the-blank game to practice Java syntax, aiming to improve players' basic Java skills. In [22] a game is used to teach C with crosswords puzzles and duck shot games. Finally, in [23] a game is used to teach C++ concepts.

There are also educational games where students need to write little programs to move on in the game or achieve a specific goal. In [24] authors present an augmented reality game that uses cards as instructions to create shapes. In this game, players must combine several cards that represent basic instructions to create a target shape using augmented reality. In [25] authors present a 3D game where players' avatars are controlled using a subset of the Logo instructional language. Similarly, one of the mini-games proposed in [21] invites players to introduce commands to guide the main character to a target point using a simple programming language. A hybrid approach is described in [26], where students are asked to program an algorithm in order to beat a game-based challenge. The game runs this algorithm and gives the student a final score, based on the algorithm effectiveness.

The main drawback of these approaches is that the games developed are hard to reuse and to scale. Most of them cover specific languages and specific programming concepts, and makes it difficult to adapt them to cover new concepts or new programming languages to fit other target audiences. In the following sections, we present our game engine which tries to tackle some of these drawbacks.

III. Game Engine

In this section, we present the game model proposed for our approach, and the engine architecture that supports it.

A. Game Model

The engine architecture is built upon a game model that defines the main high level features for our approach. This model is defined through the following key ideas:

1) Students must code in the game.

Writing code is essential to learn programming, thus, students must create programs to overcome the different levels in the game. This enables a learning-by-doing approach.

2) Promote reflection, avoid time pressure

We propose a slow paced gameplay, based on a turn-based strategy where the time has no influence on the final score. We want to avoid game mechanics that require fast reaction of the players. This allows students to take their time to think out a good solution, which is good practice in programming and promotes reflection.

3) Separate input and game mechanics

The actions that are available in the game must be simple enough so that a valid syntactic construction, representing the action, can be built with as many programming languages as possible. This adds scalability as it allows for the reuse of a game for learning different languages. The game defines a set of actions that users can perform in the game (e.g. move, rotate, shoot, etc.). This action set is linked to a set of programming structures (e.g. procedures, loops, conditions), which enables players to generate more complex behaviors in the game. An interpreter is configured to translate orders formulated in the target programming language (e.g. Java, C++, Python, etc.) onto these game sets of actions and structures.

4) Level structure

It enables incremental introduction of new concepts and programming constructs as the student becomes more skilled, facilitating a balanced level of challenge that keeps the student engaged and prevents frustration [27].

5) A clear goal is set up for each level

Clear goals are a desirable feature for any good video game [28], and it also facilitates writing programs. Puzzles and levels will be designed in such a way that multiple solutions are possible, so players can find creative solutions to them.

6) Use scores to promote competition between peers and to provide a sense of progress

Scores are used as simple metaphors to engage students and at the same time to reflect user performance. Scores help users to compare their results to others.

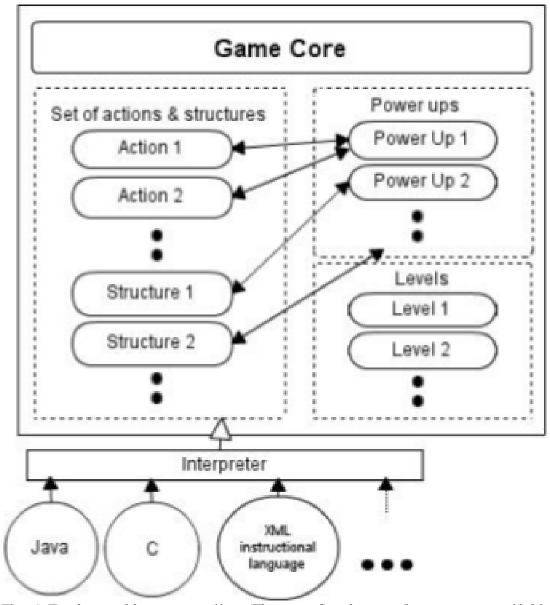


Fig. 1.Engine architecture outline. The set of actions and structures available is directly related to the power-ups unlocked in the game levels. An interpreter translates the programs to the set of actions and structures.

B. Engine Architecture

In this section, we present the engine architecture that supports the model presented in the previous section. Fig. 1 shows the main components of the architecture. These components are:

1) Game core

The game core provides basic functionality for the game. It includes, among other components, a system of game rules, a physics engine and a rendering engine. Ideally, the game core should be "fixed", i.e., it should not vary when adding new programming languages.

2) Set of actions and structures

This component contains the set of finite actions and control structures that players can use within the game. This set can be extended to fulfill new needs derived from the inclusion of new programming languages.

3) Interpreter

The interpreter translates the programming code typed by the students into game actions and structures, making the programming language to interact with the game exchangeable. For every new language that needs to be introduced, it is required to create a new interpreter able to translate it.

4) Levels

Every level is composed of logic blocks. Every block has its own logic and behavior within the game (defined in the game core). To make them extensible, levels are defined separately, in a proprietary descriptive format understandable by the game engine. Levels can be created or adapted to meet any specific needs and to adjust the duration of the game (e.g. it can be interesting to have several versions of the game with different durations, to fit one-hour lectures, two-hour lectures, etc.)

Levels are short, easy at first, and their complexity increases through the game. The game contains power-ups, each of them representing an action or a structure of the defined set. Ideally, every new power-up allows the player to perform a new action that is required to beat the current level, and all of them are evenly spread throughout all the levels in the game.

On the player side, these power-ups unlock new programming structures/elements that the players use (and learn) by writing new programs.

This mechanism is scalable, allowing designers to add new power-ups and puzzles every time the game needs to be expanded to add new concepts or programming structures.

All levels are defined in an easy-to-read format that is technology agnostic. For example, Lost in Space uses the XML format for its level definition (Fig. 2).

```
<phase>
  <wall x="0" y="0" height="9" width="5" />
  <wall x="4" y="0" width="9" height="3" />
  <wall x="6" y="2" width="7" height="8" />
  <player x="5" y="4" />
  <exit x="5" y="6" />
</phase>
```

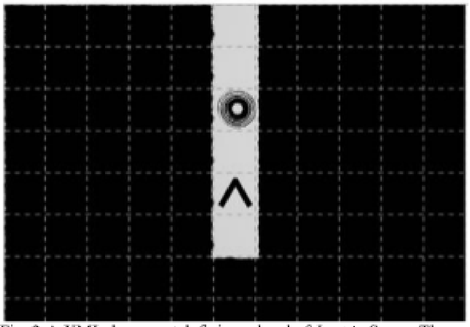


Fig. 2.A XML document defining a level of *Lost in Space*. The output level is shown below.

Each tag represents one of the blocks present in the level grid. The example defines three types of game elements: walls (3), player (1) and exit (1). The behavior and logic of each block is defined in the game core.

IV. Lost in Space

Lost in Space[1] is built upon the game model and architecture presented in section III. The game was developed to present new programming languages to students from different fields and different background knowledge. Fig. 3 shows a screenshot of the game, with the main elements highlighted.

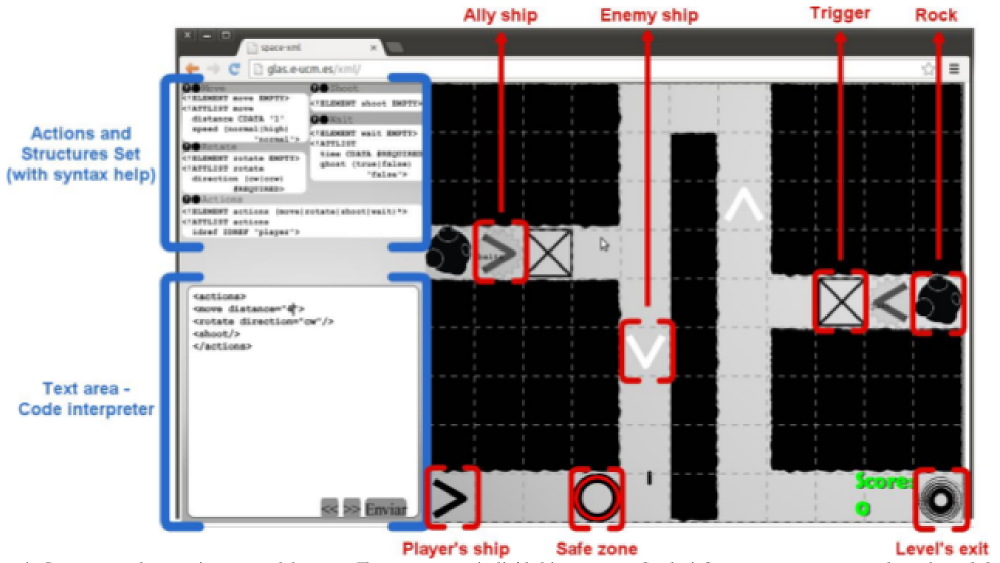


Fig. 3.Lost in Space screenshot running on a web browser. The game screen is divided in two parts. On the left part, a text area to type the code, and above, the available set of structures and actions. On the right, the current level, formed by different game elements.

The game screen is divided in two parts. The left side contains two elements: the code interpreter text area (bottom), where players must introduce their code snippets; and a help window (top) where syntax clues about the unlocked powers collected are shown.

The right side shows the current level. The goal for each level is simple: drive the player's spaceship to the current level exit (a wormhole), eluding any obstacles that may be laid out between them. These include spaceships that can be *allied* (the player can write instructions to control them and use them to help beat the level) or *enemies* (they will try to destroy the player's spaceship and their behavior cannot be controlled by the user), obstacles (*rocks* and *walls*), *safe zones* (where the player cannot be hit by enemy fire) and *triggers*, which release actions in the game (e.g. movement of obstacles, shooting, open walls, etc.). If the player's spaceship is hit by a shot or collides with an obstacle or another ship, it is destroyed and the level starts again.

To complete the levels, the player counts with several atomic operations. These operations affect the main ship and the allied ships, and are unlocked in the course of the game, every time a new power-up is collected. In the last level the player can use a total of 5 instructions:

- *Move*: translates the spaceship a number of units (squares in the grid) in the direction it is facing.
- *Rotate*: changes the direction of the spaceship 90 degrees, clockwise or counter-clock wise.
- *Shoot*: shoot a missile that can destroy enemies and rocks.
- *Wait:* inserts a waiting time between two actions.
- *Disappear*: makes the spaceship invincible for a short period of time.

The gameplay flow goes as follows: the player writes a program and submits it. The code is analyzed and interpreted by the game. If the syntax is correct, it generates a set of actions that are executed in the game, modifying the state of the player's spaceship. Otherwise it reports the error back to the player. As the player advances in the game, power-ups appears at certain levels. Once collected, power-ups are unlocked, and appears at the power-ups bar, on the top left side of the screen.

Table I provides some examples of program snippets to interact with the game. As the table shows, the game supports two target programming languages: Java and an XML-based programming language. Although XML is a markup language and not a programming language, it provides a well-defined syntax which allows building programming or declarative languages on top of it. For example, some technologies like *Ant* or *Maven* use XML-based languages to define procedural behaviors.

[1] Available at *http://bit.ly/lostinspacexml* at date August 20, 2015. Source code available in *https://github.com/anserran/lostinspace*.

TABLE I
EXAMPLES INPUT PROGRAMS FOR THE GAME

Effect in the game	Java	XML
Move ship 4 spaces	ship.move(4);	<move distance="4"/>

Make ally shoot	if
	(ship.getId().equals("ally"){
	ship.shoot();
	}
		<actions idref="ally">
		<shoot/>
		</actions>

Shoot 4 times	for (int i = 0; i<4; i++){
	shoot();
	}
		<actions repeat="4">
		<shoot/>
		</actions>

The game has thirteen levels, and its estimated completion time is 50-60 minutes, since its original use was for a one-hour lecture. The game architecture easily allows for adding new levels and mechanics, as well as new programming languages. To add new levels, it is only necessary to create the XML documents defining the new levels. To add support for a new programming language, for example, Python, it is only necessary to develop a new interpreter able to parse Python code.

V. CASE STUDY I

In this section we describe the first experience using *Lost in Space* to learn XML syntax in Computer Science settings, taking advantage of the XML-based programming language the game supports. In this case study, we used the game to completely replace a lecture about the XML markup language with half of the students of a classroom. The goal of this case study was to compare game-based instruction using *Lost in Space* with traditional instruction based on lectures (slides presentation with a teacher). The null and alternative hypotheses are defined as follows:

- H0: The effect of instruction is the same regardless of the approach (game-based vs. traditional) used.
- HA: The effect of instruction is different depending on the approach used.

A. Methods, Participants and Settings

Students were randomly and evenly distributed into either experimental group or control group–A and B respectively–. Students in the experimental group –3 females and 14 males– attended a gameplay session with the game *Lost in Space,* while students in the control group –3 females and 11 males– attended a lecture supported by a PowerPoint presentation and given by an experienced teacher. Both sessions lasted one hour and covered the same contents about XML and programming. Students in the experimental group did not get explicit guidance by any instructor. They only received some basic instructions to access the web page where the game was hosted. Students playing the game were told to upload a screen capture with the final score to the course learning management system (Moodle) as a way to encourage competition.

Identical pre and post tests were conducted to compare obtained scores in both groups. Each test had two exercises scoring from 0 to 10 each (20 is the maximum total final score for the test).

In the first exercise students were asked to identify syntactic errors in an XML document. Marking a correct error increased the total score in the exercise. Marking a non-existing error decreased the score.

In the second exercise students were asked to write a small XML document conformant to a given Document Type Definition (DTD). In order to score the exercise, three things where evaluated: 1) the ability of the student constructing valid pieces of XML, 2) the compliance of the document with the DTD, and 3) the completeness of the written document, using as reference the document the exercise asked for.

Tests were anonymous –a unique untraceable code was used to pair pre and post tests for each student–, and students had 15 minutes to complete each test. At the end of the session, students in both groups were also asked to rate the educational experience using a 5-point Likert scale.

B. Results

Fig. 4 compares results of pre and post tests for groups A (experimental) and B (control). Table II shows a more detailed summary of the final results, broken down by questions.

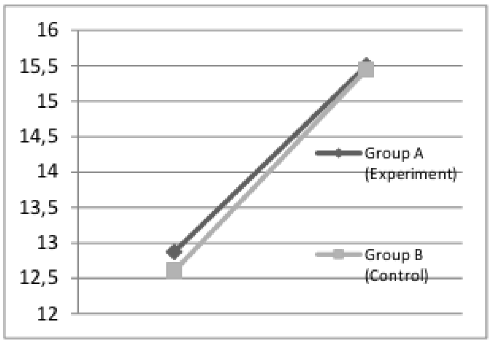


Fig. 4. Score results from Group A and Group B from pretest and posttest

In the pretest, students scored 12.87 and 12.60 on average in groups A and B respectively; this difference was not statistically significant using an unpaired T-Test ($p=0.84$) and an Independent-Samples Mann-Whitney U-Test ($p=0.968$). Therefore we conclude that both groups had equivalent initial knowledge on the subject.

TABLE II
RESULTS FROM CASE STUDY I

Exercise	Pre / Post	Group A	Group B
Q1 (over 10)	Pre test	7.38 ± 1.56	7.12 ± 1.81
	Post test	8.29 ± 1.17	8.35 ± 1.63
	Gain (Post-Pre)	0.91 ± 1.30 **(+9.1%)**	1.23 ± 1.96 (+12.3%)
Q2 (over 10)	Pre test	5.49 ± 2.30	5.47 ± 2.66
	Post test	7.20 ± 2.26	7.09 ± 2.16
	Gain (Post-Pre)	1.71 ± 2.49 **(+17.1 %)**	1.61 ± 1.65 (+16.15 %
Total =Q1+Q2 (over 20)	Pre test	12.87 ± 3.05	12.60 ± 4.08
	Post test	15.50 ± 2.87	15.44 ± 3.57
	Total gain	2.62 ± 2.91 **(+13.1 %)**	2.83 ± 2.99 (+14.5 %)

Both groups scored higher in the post test. Group A registered a score increase (post-pre) of 13.1% and group B an increase of 14.5% on average. Differences between post and pretests were found to be statistically significant in both groups after running a paired T-Test for each group ($p=0.002$ for group A, $p=0.003$ for group B) and a related-samples Wilcoxon

Signed Rank test (0.002 and 0.009 respectively).

Differences in the post tests across groups were found not statistically significant after running an unpaired T-Test (p=0.960) and a Mann-Whitney U-Test (p=0.858), indicating that both groups ended up with a similar knowledge level. As a consequence, we fail to reject the null hypothesis.

We also observe that the score increase was higher in the second exercise (writing an XML document) than in the first one (finding errors in an XML document) for both groups. However, in none of the groups this difference was found to be statistically significant (Group A: T-Test p=0.237, Related-Samples Wilcoxon Signed Rank Test p=0.344; Group B: T-Test p=0.530, Wilcoxon Signed Rank Test p=0.706).

There are also differences in self-reported satisfaction. Figure 5 shows the results for Group A and Group B. It is high in both groups (Group A = 4.76 and Group B = 3.92), although it is higher in Group A (game group), with more than 76% of top score (5/5) responses.

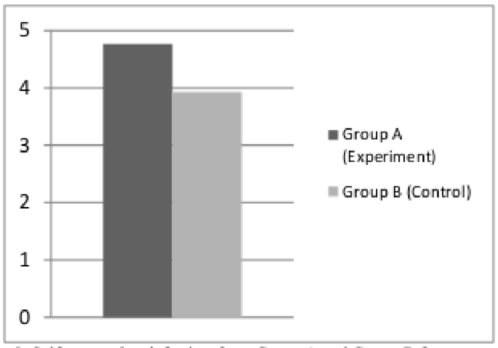


Fig. 5. Self-reported satisfaction form Group A and Group B from posttest

C. Discussion

Results show that students' scores increased both after traditional instruction and game-based instruction. Post test scores were equally high for both (near 15 over 20 on avg.), showing little not significant differences, which suggests that both approaches are appropriate for teaching/learning XML.

The two approaches were similar in content (i.e., in both sessions the educational content covered was similar, in fact, game content was designed around the slides presentation); however, teaching approaches were different. During instruction, the teacher explicitly presented the XML foundations –using a PowerPoint presentation–, in the game, XML content was implicit, i.e., the game did not present any formal content to the player with text or any other direct explanation: the students build their own knowledge through active play.

The average effect of instruction (13-14%) is not very high. However, this may be the consequence of two causes: the reduced exposure to instruction (less than 50 minutes) and the high pretest scores of the participants (around 12.5 points over 20).

There was no difference in the pretest score between groups, which allows us to discard any potential bias introduced during the randomization process.

Finally, students in the experimental group were more satisfied with the experience than the students in the control group. This is not because the satisfaction in the control group was low but a consequence of the outstanding satisfaction rates achieved in the experimental group. This finding is consistent with researchers' observations during the sessions, who noticed deep engagement in students in the experimental group. Researchers observed abnormally frequent interaction between peers, who vigorously competed to get the highest possible score. Students also kept playing after the class.

If we consider all factors, the different engagement observed in the two groups may suggest that overall *Lost in Space* was a better instructional approach, although it did not yield better results than traditional instruction in terms of knowledge acquisition.

VI. Case Study II

We designed a second case study to explore whether the effects of using the game for instruction –knowledge increase and high motivation– with a different student population are consistent to those observed in the previous experiment. This will help us discuss on the scalability of the game model proposed by analyzing the size of the potential target audience that could use the game to learn programming.

In the second study we replicated the gameplay session (instruction delivered to Group A) described in section VI with college students enrolled in a social sciences degree, who had no previous programming background (Group C) and therefore were expected to obtain a significantly lower pretest score. The null and alternative hypotheses are described as follows:

- H0: The score increase factors (post-pre) of the two game-based instruction groups (A vs. C) with different programming backgrounds are equal.
- HA: The score increase factors are not equal.

A. Method, Participants and Settings

Group C was made up of 13 students (5 males and 8 females) from a Degree in Information and Documentation (social sciences), who had not received previous instruction in computer programming and had no technical background.

We replicated the experimental design described in section VI, using the score increase (obtained as the difference in score obtained between pretest and posttest) as the independent variable that estimates the knowledge gain about XML markup language.

B. Results

Figure 6 shows a high-level view of the results for Group C compared to Group A, while Table III provides insight on these results. These data indicate that the initial knowledge of students is lower than in groups A and B (who had computer programming background) as initially expected. The difference between pretest scores was found statistically significant after running an unpaired T-Test (p=0.001) and an Independent-Samples Mann-Whitney U-Test (p=0.002).

Students' score was higher in the post test than in the pre-test, showing an average increment of 10.95%. This difference was found statistically significant after running a paired-samples T-Test (p=0.004) and a Related-Samples Wilcoxon Signed Rank Test (p=0.011), which suggests that the game was also effective for group C.

Compared to group A, both groups showed similar total score increments, being slightly higher for group A (2.62 ± 2.91) than for group C (2.19 ± 2.91 for group C). This difference was not found statistically significant after running an unpaired T-Test (p=0.650) and an Independent-Samples Mann-Whitney U-Test (p=0.662), allowing us to retain the null hypothesis. It suggests that the effect of game-based instruction was similar for all students regardless their previous computer programming background.

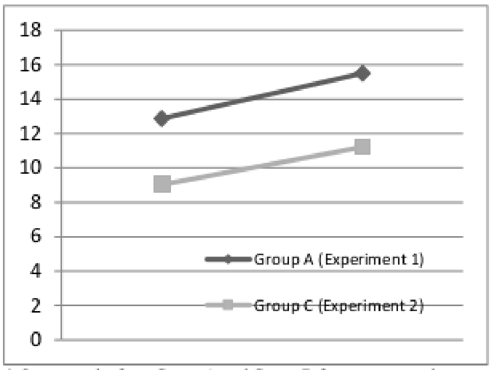


Fig. 6. Score results form Group A and Group B from pretest and posttest

TABLE III
RESULTS FROM CASE STUDY II

Exercise	Pre / Post	Group C	Group A
Q1 (over 10)	Pre test	6.10 ± 1.76	7.38 ± 1.56
	Post test	5.98 ± 1.84	8.29 ± 1.17
	Gain (Post-Pre)	-0.13 ± 1.30 **(-1.3%)**	0.91 ± 1.30 **(+9.1%)**
Q2 (over 10)	Pre test	2.91 ± 2.03	5.49 ± 2.30
	Post test	5.24 ± 2.51	7.20 ± 2.26
	Gain (Post-Pre)	2.32 ± 2.22 **(+23.2 %)**	1.71 ± 2.49 **(+17.1 %)**
Total =Q1+Q2 (over 20)	Pre test	9.02 ± 2.63	12.87 ± 3.05
	Post test	11.22 ± 3.73	15.50 ± 2.87
	Total gain	2.19 ± 2.91 **(+10.95 %)**	2.62 ± 2.91 **(+13.1 %)**

However, results are not totally equivalent in both groups, as differences across exercises are significant for group C. In this group, students did not improve their score for the first exercise (-1.3% increase on average), while their performance in the second exercise increased a 23.2% on average.

Finally, Figure 7 shows the average self-reported satisfaction of Group C, along with the average satisfaction of Group A and Group B. Students valued the experience similarly to Group A, with an average score of 4.86 (over 5).

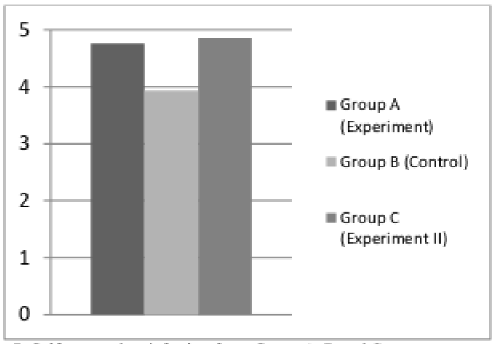


Fig. 7. Self-reported satisfaction from Group A, B and C

C. Discussion

Group C started from a lower level than Group A, which is likely a consequence of their lack of programming background. However, data suggest that students in group C increased their knowledge after playing the game in a similar way to group A. However, an interesting finding that deserves further discussion is that students in group C only improved scores for the second exercise.

Both exercises had different mechanics. In the first exercise students had to identify syntactic and semantic errors in a fragment of an XML document. To complete this exercise, they had to *be aware* of the syntactic and semantic rules to form valid XML documents, i.e., they needed to understand the concept of markup/programming language, and that all of them have a syntax in its foundation. In the second exercise students had to write a short XML document. In this case, students needed to have the procedural skills to write XML documents. While students have to apply syntactic and semantic rules to reach a valid solution, they do not need to be aware of what these rules look like - they just need to apply them.

That may be explained by the programming background of group A, which allowed them to infer the syntactic and semantic rules behind XML after practicing with the game, even if these rules were never explicitly presented to them. In contrast, students in group C were not able to make that inference on their own, probably as a consequence of their lack of computer programming background.

Finally, group C reported the highest satisfaction of the three groups involved in the experiment. According to the observations made by the instructors present during the experiment, the reasons for this difference are the following: 1) group C did not have any similar past experience; and 2) their low exposure to videogames in their daily life increased the 'surprise effect' in the session, which lead to greater satisfaction, in contrast to group A, which was probably more exposed to videogames due to their CS background.

VII. CONCLUSIONS AND FUTURE WORK

In this paper we discussed how educational videogames can help address some of the challenges related to computer programming instruction, and we made an extensive review on how videogames and learning programming are coupled together in the literature.

Building upon the extensive literature on this topic, we identified the need to find more scalable and flexible game-based instruction paradigms, given the increasing interest in providing computer programming instruction to a wider sector of the population, a population that not only includes computer science students but also college students of different disciplines and even kids.

As a response, we propose a flexible game architecture, supported by an extensible game engine and game model, to generate fun, engaging and entertaining games that can be extended to cover different languages, suit diverse target audiences and fill different classrooms time spans.

We developed a game using this game engine (*Lost in Space*), which was used for XML markup language instruction in two different college settings (computer science and social sciences respectively). In these experiences the game was well accepted by the students, and we also observed that they deeply engaged in gameplay. Moreover, data collected suggest that they learned with this kind of game-based instruction in a similar way to traditional instructional methods regardless of their background. However, a potential limitation of the approach for students with no computer programming background was identified.

The data collected suggests that students from social science were not able to infer the syntax of the language on their own, as they do not have any previous programming background and syntactic rules were never explicitly provided to them. However, students with programming background were able to make that inference. Instructors willing to use this approach should take this finding into consideration and design a strategy to help students to construct the explicit representation of the knowledge acquired, using debriefing sessions or closer tutoring, for example.

However, it seems reasonable to assert that, whatever the background of the target audience, our approach can be a rewarding and entertaining way to learn and teach programming.

We think that our approach is adequate to introduce computer programming in general as well as new programming languages in particular. And, even if it was initially designed as an additional strategy to support learning, it also can be used as the main educational resource in some contexts.

Finally, we would like to point out that this approach can have several extra advantages if we also consider the game engine as an assessment tool.

We deployed *Lost in Space* in a web server with some basic user tracking. This helped us discover helped us discover that some of the students kept playing several days after the experiment was over. This gave us the idea that a more advanced tracking of the gameplay of each player can help us to assess and improve the learning process [29].

The building blocks of the game engine– mainly levels and power-ups– mark clear milestones that we can track and assess to have a deeper insight of how players are learning programming. This will also help us to improve the game engine and game model.

ACKNOWLEDGMENT

Special thanks to Ricardo García-Mata from the UCM statistical service for his assistance with the data analysis. Also thanks to all beta testers from the e-UCM group who helped us to improve and polish the game, and to all the students that participated in the experiments.

This research has been partially funded by Regional Government of Madrid (eMadrid S2013/ICE-2715), by the Complutense University of Madrid (GR3/14-921340), by the Ministry of Education, Culture and Sport of Spain (TIN2013-46149-C2-1-R, FPU AP2012-4310) and by the European Commission (SEGAN 519332-LLP-1-2011-1-PT-KA3-KA3NW, H2020-RAGE 644187).

REFERENCES

[1] [J. A. Betz, "*Computer Games: Increase Learning in an Interactive Multidisciplinary Environment*," J. Educ. Technol. Syst., vol. 24, no. 2, pp. 195–205, 1996.

[2] W. L. Johnson and N. Wang, "Experience with serious games for learning foreign languages and cultures," in *Proceedings of the SimTecT Conference, 2007.*

[3] T. Baranowski, R. Buday, D. I. Thompson, and J. Baranowski, "*Playing for Real. Video Games and Stories for Health-Related Behavior Change,*" American Journal of Preventive Medicine, vol. 34. 2008.

[4] B. Manero, C. Fernández-Vara, and B. Fernández-Manjón, "*E-learning a Escena : De La Dama Boba a Juego Serio,*" IEEE-RITA, vol. 1, no. 1, pp. 51–58, 2013.

[5] T. W. Malone and M. R. Lepper, "*Making learning fun: A taxonomy of intrinsic motivations for learning,*" R. E. Snow and M. J. Farr, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1987, pp. 223–253.

[6] P. Doerschuk, J. Liu, and J. Mann, "An INSPIRED game programming academy for high school students," *2012 Front. Educ. Conf. Proc.*, pp. 1–6, Oct. 2012.

[7] Á. Serrano-Laguna, J. Torrente, B. Manero, and B. Fernandez-Manjon, "A game Engine to Learn Computer Science Languages," in *IEEE Frontiers in Education Conference, 2014*, pp. 765–771.

[8] K. Doss, V. Juarez, D. Vincent, P. Doerschuk, and J. Liu, "Work in progress — A survey of popular game creation platforms used for computing education," *2011 Front. Educ. Conf.*, pp. F1H–1–F1H–2, Oct. 2011.

[9] T. Barik, M. Everett, R. E. Cardona-Rivera, D. L. Roberts, and E. F. Gehringer, "A community college blended learning classroom experience through Artificial Intelligence in Games," in *2013 IEEE Frontiers in Education Conference (FIE)*, 2013, pp. 1525–1531.

[10] J. Heliotis, I. Bezakova, and S. Strout, "Programming board game strategies in CS2," in *2013 IEEE Frontiers in Education Conference (FIE)*, 2013, pp. 4–5.

[11] J. Robertson and J. Good, "*Story creation in virtual game worlds*," Commun. ACM, vol. 48, no. 1, pp. 61–65, 2005.

[12] A. E. Rais, S. Sulaiman, and S. M. Syed-Mohamad, "Game-based approach and its feasibility to support the learning of object-oriented concepts and programming," in *2011 Malaysian Conference in Software Engineering*, 2011, pp. 307–312.

[13] P. Doerschuk, V. Juarez, J. Liu, D. Vincent, K. Doss, and J. Mann, "Introducing programming concepts through video game creation," in *2013 IEEE Frontiers in Education Conference (FIE)*, 2013, no. 0634288, pp. 523–529.

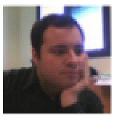
[14] M. B. MacLaurin, "*The design of kodu: a tiny visual programming language for children on the Xbox 360,*" ACM Sigplan Not., vol. 46, no. 1, pp. 241–245, 2011.

[15] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, and Y. Kafai, "*Scratch : Programming for all,*" Commun. ACM, vol. 52, no. 11, pp. 60–67, 2009.

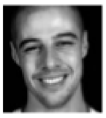
[16] S. Arakawa and S. Yukita, "An Effective Agile Teaching Environment for Java Programming Courses," *Proceedings. Front. Educ. 36th Annu. Conf.*, pp. 13–18, 2006.

[17] I. F. de Kereki, "Scratch: Applications in Computer Science 1," in *2008 38th Annual Frontiers in Education Conference*, 2008, pp. T3B–7–T3B–11.
[18] W.-K. Chen and Y. C. Cheng, "*Teaching Object-Oriented Programming Laboratory With Computer Game Programming.*" IEEE Trans. Educ., vol. 50, no. 3, pp. 197–203, Aug. 2007.
[19] M. Chang, "Web-Based Multiplayer Online Role Playing Game (MORPG) for Assessing Students' Java Programming Knowledge and Skills," in *2010 Third IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, 2010*, pp. 103–107.
[20] P. Sancho, J. Torrente, and B. Fernández-Manjón, "Do Multi-User Virtual Environments Really Enhance Student's Motivation in Engineering Education?" IEEE digital library, San Antonio, TX, USA, 2009.
[21] T. Mitamura, Y. Suzuki, and T. Oohori, "Serious games for learning programming languages," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 1812–1817.
[22] R. Ibrahim, J. Semarak, K. Lumpur, and A. Jaafar, "Using educational games in learning introductory programming: A pilot study on students' perceptions," in *2010 International Symposium on Information Technology*, 2010, pp. 1–5.
[23] S. H. A. Hamid and L. Y. Fung, "Learn Programming by Using Mobile Edutainment Game Approach," in *2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07)*, 2007, pp. 170–172.
[24] N. Masso and L. Grace, "Shapemaker: A game-based introduction to programming," in *2011 16th International Conference on Computer Games (CGAMES)*, 2011, pp. 168–171.
[25] I. Paliokas, C. Arapidis, and M. Mpimpitsos, "PlayLOGO 3D: A 3D Interactive Video Game for Early Programming Education: Let LOGO Be a Game," in *2011 Third International Conference on Games and Virtual Worlds for Serious Applications,* 2011, pp. 24–31.
[26] H. C. Jiau, J. C. Chen, and K.-F. Ssu, *"Enhancing Self-Motivation in Learning Programming Using Game-Based Simulation and Metrics*," IEEE Trans. Educ., vol. 52, no. 4, pp. 555–562, Nov. 2009.
[27] J. Chen, "*Flow in games (and everything else),*" Commun. ACM, vol. 50, no. 4, pp. 31–34, 2007.
[28] J. P. Gee, "*What video games have to teach us about learning and literacy,*" Comput. Entertain., vol. 1, no. 1, p. 20, Oct. 2003.
[29] Á. Serrano-Laguna and B. Fernández-Manjón, "Applying learning analytics to simplify serious games deployment in the classroom," in *IEEE Global Engineering Education Conference, EDUCON*, 2014, pp. 872–877.

Ángel Serrano Laguna finished his Master Degree in Computer Science at the UCM in 2012. Currently he is working as a full-time researcher in the e-UCM research group, while working on his PhD. His research focuses in the design and implementation of Serious Games and the creation of tools to ease its introduction in the classroom, as well as the study of Learning Analytics techniques to assess those games.



Javier Torrente, PhD, MSc, obtained his PhD in computer science at the UCM in 2014. His PhD focused on how to reduce the cost of making digital games more accessible for people with a disability. Currently he works as a full-time researcher for University College London (UK), and he was formerly a member of the e-UCM research group at UCM, Madrid, Spain. He has published more than 70 research papers in academic journals and conferences in the fields of serious games, HCI and accessibility.



Borja Manero Iglesias received a Bachelor in Physics (major in Computer Science) from the Universidad Complutense of Madrid (1999) and in dramatic arts by Réplika School of Theater (2007). Currently, he is a lecturer in the Department of Software Engineering and Artificial Intelligence of the Complutense University in Madrid, Spain. His research interests focus on educational games and their use on theater and cinema.

Baltasar Fernández-Manjón, IEEE Senior Member, is full professor in Computer Science at UCM and director of the e-UCM e-learning research group. His research interest is focused on applications of ICT in education and in serious games and educational simulations applied to different domains (e.g. medicine, culture, languages). He is also working in the application of e-learning standards to the integration of those technologies in e-learning systems.

6.2. A framework to improve evaluation in educational games

6.2.1. Cita completa

Ángel Serrano-Laguna, Eugenio Jorge Marchiori, Ángel del Blanco, Javier Torrente, Baltasar Fernández-Manjón (2012), ***A framework to improve evaluation in educational games,*** Global Engineering Education Conference (EDUCON), 2012 IEEE, 17-20 abril 2012. DOI: 10.1109/EDUCON.2012.6201154

6.2.2. Resumen original de la contribución

The evaluation process is key for educator's acceptance of any educational action. The evaluation is challenging in most cases but especially when educational games are used. In educational games if in-game evaluation exist it is usually based on a series of simple goals and whether these goals are achieved (i.e. assessment). But we consider that evaluation can be improved by taking advantage of in-game interaction, such as the user behavior during the game and the type and number of interactions performed by the user while playing. In this paper, we propose an evaluation framework for educational games based on in-game interaction data. We discuss how user interaction data is collected in the most automatic and seamless way possible, how to analyze the data to extract relevant information, and how to present this information in a usable way to educators so they achieve the maximum benefit from the experience. The evaluation framework is implemented as part of the eAdventure educational platform, where it can be used both to improve upon traditional basic assessment methods (i.e. goals, scores & reports) and to provide information to help improve interaction with games (e.g. discovery strategies).

A framework to improve evaluation in educational games

Ángel Serrano, Eugenio J. Marchiori, Ángel del Blanco, Javier Torrente, Baltasar Fernández-Manjón
Department of Artificial Intelligence and Software Engineering
Complutense University
Madrid, Spain
aserrano@e-ucm.es, {emarchiori, angel.dba, jtorrente, balta}@fdi.ucm.es

Abstract—**The evaluation process is key for educator's acceptance of any educational action. The evaluation is challenging in most cases but especially when educational games are used. In educational games if in-game evaluation exist it is usually based on a series of simple goals and whether these goals are achieved (i.e. assessment). But we consider that evaluation can be improved by taking advantage of in-game interaction, such as the user behavior during the game and the type and number of interactions performed by the user while playing. In this paper, we propose an evaluation framework for educational games based on in-game interaction data. We discuss how user interaction data is collected in the most automatic and seamless way possible, how to analyze the data to extract relevant information, and how to present this information in a usable way to educators so they achieve the maximum benefit from the experience. The evaluation framework is implemented as part of the eAdventure educational platform, where it can be used both to improve upon traditional basic assessment methods (i.e. goals, scores & reports) and to provide information to help improve interaction with games (e.g. discovery strategies).**

Learning Analytics; Educational video games; framework proposal; case study;

I. INTRODUCTION

In traditional education, either in higher education or in other levels, the main evaluation method is based on written final exams [1]. This method, as some authors have pointed out [2], presents a series of problems. These problems are related not only to the student evaluation, but also to the evaluation of the educational action itself: the amount of data available is limited, and it is usually restricted to students and educators subjective perceptions (e.g. through polls about the past courses). Other metrics, mostly based on exam grades, might not give enough information about the educational action, or whether it was a success or a failure and why. Moreover, these data usually become available when the action is finished or when it is too late to make an intervention, improvement or correction in the ongoing action.

With the emergence of the Web, on-line educational resources have grown exponentially. Many institutions now use LMS (Learning Management System) to organize their courses, to allow students to communicate among themselves and with teachers, and to improve access to educational resources [3]. Still, despite all of these on-line resources, evaluation is still usually performed using traditional methods. Most of the content presented in LMS is finally evaluated through written exams in classrooms, or through online tests or exams.

However, there is a whole new body of data, derived from the student interaction with on-line educational resources. These data can be collected and analyzed not only to improve the evaluation methods, but also to obtain real-time feedback about the progress of any educational action, enabling educators to predict results and react to that progress.

The field studying the use and analysis of this kind of data is known as Learning Analytics [4]. This new field advocates of capturing all the data derived from interaction with on-line educational resources, and analyzing it to assess students, predict future events and act consequently to refine educational actions. These ideas have been successfully applied in other disciplines, like Business Intelligence, a well-extended set of techniques for analyzing business data to support better business decision-making [5], or Web Analytics, where internet data are collected in order to understand and optimize web usage [6].

Currently, LMS are the main target for Learning Analytics systems. Projects like SNAPP [7] or LOCO-Analyst [8] offer statistics about the interactions made by students inside an LMS. SNAPP is focused on analyzing forum activity and creating network diagrams of all interactions among students. From this, it infers which are the most active students in a class, and those students who are "disconnected" or "at risk", among other features. LOCO-Analyst is based on student interaction with learning content (e.g. number of views, time spent with every resource) that is used to infer conclusions about learning content characteristics (e.g. difficulty or importance).

Though LMS activity reports can contribute to better understand students' interactions with content and resources, educational games represent an ideal environment to capture more detailed and diversified student interaction. In the last few years, Game Analytics are being used to let developers know about how players interact with their games. One of their main purposes is identifying where and why a player got stuck during the game[1], so game developers can try to smooth this hardness, to avoid player frustration and thus keep him

[1] http://www.gamasutra.com/view/feature/6155/hot_failure_tuning_gameplay_with_.php

engaged and playing [9]. All these ideas applied to educational games, combined with other Business Intelligence and Web Analytics techniques and guided by the Learning Analytics process, are addressed in this paper.

First, we board the main steps in the Learning Analytics process, and how these steps can be particularized in educational games, proposing a theoretical Learning Analytic Model and a Learning Analytic System. Then, we propose an implementation upon the educational game platform eAdventure and a use case to deploy the system, and finally some ideas and thoughts about the whole process.

II. Learning Analytics Steps in Educational Games

Authors agree that Learning Analytics process [2] begins with selecting the most relevant student data to be captured. Once it is captured, data must be aggregated and transformed into reportable information (for example, using charts or other visual representations). With this information, the educator should be able to judge how the student used the educational resource. Some authors call this step *predict*, since information is converted into knowledge, and knowledge enables predictions. However, in our approach we will use this step to assess the student, and for now on, we will name this step as *assess*. Assessment information can be used, under certain conditions, to dynamically assist the student, and to refine the educational resource, based on the students results. Finally, all the knowledge acquired can be shared with others whom could benefit from it. Table 1 represents a scheme with all the steps and their descriptions.

TABLE I. Learning Analytics Steps

Step	Unit produced	Description
Select	Data	Choose the basis data to be captured
Capture	Data	Collect the selected data
Aggregate & Report	Information	Sort out captured data and convert it in information
Assess	Knowledge	Understand reported information and convert it into knowledge, assess students
Use	Knowledge	Adapt the system based on assessment
Use & refine	Knowledge	improve educational action
Share	Knowledge	Share knowledge for the benefit of others

To support all these steps, we propose a system based on a Learning Analytics Model (LAM) holding all the information required for every step, and a Learning Analytics System (LAS) endued with all the processing power required by the model.

In this approach, focused on educational games, we consider the LAS as a separate system from the game engine, but both are communicated. The LAS also has access to the game model and the LAM (Fig. 1). The LAM is constructed by a set of models, which are directly related to the different modules contained by the LAS (Fig. 2).

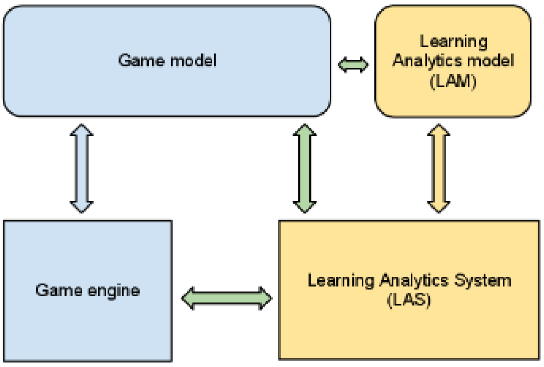


Figure 1. Relation between the different components involved in the Learning Analytics process. The LAM is dependent on the game model and the LAS. LAS is aware of the LAM and the game model, and can communicate with the game engine.

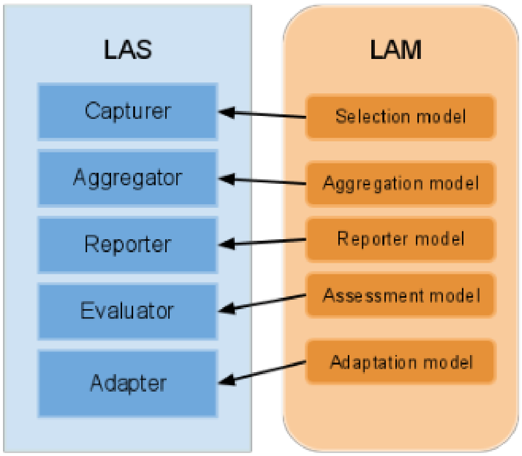


Figure 2. The LAS consists of a series modules that take part in the different steps of the Learning Analytics process. The LAM holds models for those steps which requires defining a model to work. The LAS uses the LAM to process all the data captured and generated.

In this section the learning analytics steps are described in general but also for educational games in particular, building the LAM and the LAS upon them. As starting point for this definition we consider the available data in games and how to adapt these data to the theoretical model in order to extract the maximum information possible from this media.

A. Select and capture

First, the data to be captured by the LAS are selected. These data will be the raw material that will feed the steps that follow. The data selection criteria are lead by the educational resource objectives and some constrictions such as technical limitations and privacy policies must be taken into account. In educational resources, meaningful data can be selected from personal information about the student (e.g. age, gender, etc.), academic information and any other data provided by the resource context.

While in static resources (e.g. PDF files), the only extractable data are the number of views and the time spent with them, the interactive nature of educational games provides a whole new type of data that can be selected:

- *GUI events performed by the student during the game*: mouse clicks, keys pressed, and other events (joystick movements, buttons interactions), depending on the input method. Not only the event itself can be recorded, but also the time when it occurred and whether it was performed over a target (e.g. some click over a game object). These events can provide clues about the student behavior during the game (e.g. if all GUI events were captured the LAS would be able to recreate the complete game play).

- *Game state evolution*: the game state is a set of variables and their values that specify a concrete status in the game instance. The evolution of variables through time describes the development of the different goals of the game. Depending on the case, the whole game state evolution could be recorded, or it could be recorded only in some points (e.g. when a phase ends, or a goal is achieved).

- *Logic events*: a logic event is anything that moves the game-flow forward. Changing the value of a variable, finishing a phase, launching a cut-scene (i.e. a slide-show or video), losing a life, achieving a defined goal, etc. Some logic events, and their timestamps, can be directly related to the student progress in the game, and thus be relevant for the assessment.

Selectable data are limited by the technologies used to deploy the games: Not every piece of data here proposed will be available in every game platform. These selectable data, then, are platform-dependent and must be defined in the LAM's *selection model*. To avoid unnecessary data capture, every game model should define, among all selectable data, the final data to be captured according to their own purposes.

Once the data are selected, the framework requires a way to capture it. The technology involved will be very important to establish how the data are collected. Access to different internal parts of the game engine is required to capture some of the information. This implies, for instance, that such model cannot be generally applied to commercial games provided as black boxes.

Another issue is the moment when the captured data are passed to the LAS to begin processing. The simplest way is to store all the data locally and send it back to the LAS when the game is finished. Data could be sent in certain significant moments, like when the student ends a phase or achieves a goal. Moreover, all the data could be sent to the LAS as they are being captured. Last two approaches enable real-time assessment that can be used to assist the student during the game. Depending on the needs, all these data might go through a filter in order to make it anonymous.

B. Aggregate and Report

The captured data must be organized in such a manner that it can be shown in human readable formats, like tables or graphics. A more meaningful report can be done if the LAM contains, in the *aggregation model*, semantic rules to interpret all the received data. For example, the system could relate a raw event (e.g. a variable taking a particular value) with a meaningful feat (e.g. the player completed a goal). These semantic rules can be based on the game engine, where some events can have an implicit meaning (e.g. an engine where pressing escape key always brings up the menu) or on the game itself (e.g. if the game variable "hits" is "8", the phase is completed).

Semantic rules can be expressed like conditions producing new data to be reported: when a condition (based on GUI events, a logic event or a concrete game state) is met a new unit of data, defined in the *aggregation model*, is generated. E.g. when in the game state, the variable *score* equals to 10, and the variable *gold* equals to *15*, the LAS *aggregator* produces a logic event "Goal 1 completed". The *reporter* could then treat this event as it would with any other logic event.

The LAS' *aggregator* needs to be endued with mechanisms capable of understating and processing these kinds of rules (Fig. 3).

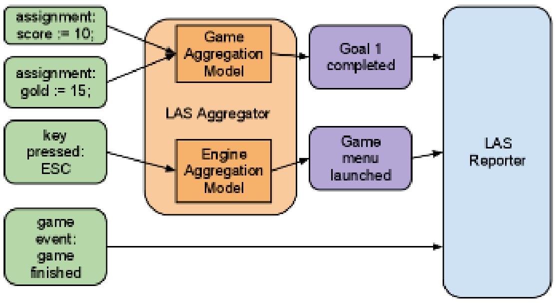


Figure 3. Raw events are passed through the semantic rules contained by the LAM, and converted to more meaningful data. Events can be grouped and simplified through semantic rules. Some events, such as the game finishing, have enough meaning and do not need to be interpreted by any rule.

After aggregation, information can be reported in common ways, such as tables or charts, but also, we can take advantage from the inherent characteristics of games to report information with new representations. For example, "heat maps" could be created for every phase, in which the heat can measure the amount of times the player clicked in every point of the phase, or the places where the player was defeated. If there is enough information of user interaction, an animation recreating how the student played a game phase could be shown.

The *reporter model* contains which information must be reported and which representations must be used. Common reports can be defined at engine level (e.g. heat maps for every phase can be common for all games), as well as reports at game level, holding important information in that particular game.

These reports can be even richer if data from different students are aggregated. Average results can spot which goals took more time or the places where most of the students failed.

C. Assess and Use

The information and the reports generated until now can give an overview of how the students are using an educational resource. However, this information should have some practical consequences to be really useful. It is the moment to transform the information received into knowledge. In the educational game context, all the information reported is processed to assess the student in this step.

Games are organized around goals. In educational games, these goals should be based on the success in some educational aspects. In our context, and based on the concepts of the selection and aggregation process, we could have several types of goals, represented by:

- A GUI event or a series of GUI events performed by the student, over a game object or in total.
- A concrete game state, fulfilled fully or partially.
- A variable taking a defined value.
- The launch of a particular logic event.

These classes of goals are platform-dependant, and should be defined by the LAM's *assessment model*.

Compound goals can be defined based on these simple goals. An educational game can define all the necessary goals to cover all the educational aspects that are to be learned by players of the game. Based on these goals and with the reported information the game can be used to assess the student.

This assessment can have two applications: one, just to measure the success of the student in the game, and act accordingly (e.g. enabling the student to access to new educational resources), or, if the captured data are being passed to the LAS during game time, dynamic adaptation through real-time assessment (e.g. if the student got stuck in some point, the system will offer him help). Rules for this assistance are contained by the *adaptation model*, and are processed by the adapter, which is able to communicate with the game engine to perform the adaptation.

Assessment and dynamic adaptation could be more sophisticated. As some authors pointed out [10], propagating information through complex structures, like Bayesian networks, can help to determine what is going on in virtual simulations, and better decide what adaptation profile to choose.

However, our approach pretends to be based on easy principles and stay accessible for as many educators as possible. Complex structures, like Bayesian networks, are normally out of reach for most educators.

D. Refine and share

With all the accumulated knowledge from previous steps, an educator can know about the global results (assessed in the previous step) of the educational action and can identify which educational goals were not achieved as expected. Thus, educators can refine the educational resources to improve the results or readjust their expectations.

In educational games, those game goals that were not solved as expected can be detected. Aggregated data from several students can ease this job, pointing out, in average, which goals made more trouble to students. From here, the game could be modified or even redesigned to facilitate its accomplishment. This does not directly imply making the game simpler or shorting the educational goals, it could be enough to smooth the learning curve in the game, or adding some extra help in game points especially hard. Maybe, learning analytics conclusions showed that the student did not get stuck, but stop playing the game after a while, indicating that it was not engaging enough.

Finally, the LAS can share all the knowledge obtained with other systems. These systems cover from LMS to institution administrative systems. Even making the data public can be an option in some cases. In order to be able to share data, some considerations such as privacy policies, what knowledge is shared or which standards are used in the communication, need to be taken into account.

III. IMPLEMENTATION PROPOSAL: EADVENTURE

eAdventure is an educational game platform developed and maintained by the <e-UCM> research group at the Complutense University of Madrid for the last 5 years. This platform includes a game engine and an easy-to-use editor, targeted at educators. eAdventure is currently undergoing the development of the 2.0 version, where new features are being added. Some of these include support for multiple platforms [11] and an easy to use narrative representation of games [12]. Moreover, we propose to implement the framework presented in this paper in this new version of the system.

eAdventure games are composed of scenes, which can represent from a simple scenario in an adventure game where the player's avatar moves to a more complex slide-show, going through an array of mini-games and other content. These scenes are always composed of simpler parts referred to as scene elements, each of which will usually have a graphical representation, a position in the screen, behaviors, etc. The current scene and the status of elements in the screen are defined as the game state. It is the flow from one scene to another, behaviors of the scene elements and effects (changing current scene, showing text, launching videos, assigning values to variables) that make up a game, by continually changing the game state until a final state is reached.

The LAS is implemented on a server. It is initialized with the Game Model, (containing the Adventure LAM) and the Engine LAM. The LAS has several modules to satisfy the requirements for every step of the Learning Analytics process (Fig. 4). The relation between the modules and the steps is detailed below.

A. Select

Given that eAdventure is intended to be a general game engine, including its own editor, our proposal tries to make selectable the biggest amount of data, letting to the game designer choose between all the available options. The eAdventure Learning Analytics Model define three units of selectable data:

- *LAGUIEvent*: represents a detailed GUI interaction. It holds the GUI event (mouse action, drag & drop, keyboard action) with its properties (mouse button, key pressed) and the target scene element, if exists.
- *LALogicEvent*: represents the launching of a game effect. It holds the generic effect data and additional information about the concrete instantiation (e.g., in the changing scene effect, the initial scene and the final scene).
- *LAGameState*: represents a game state in a certain moment. It contains a map holding all the game variables associated with their current value.

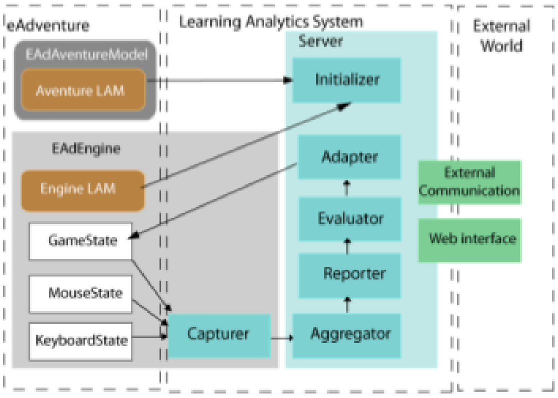


Figure 4. General organization for the LAS integrated with eAdventure. The LAS is mostly deployed in a server, an can communicate with the game engine. A web interface and external communication with other systems is offered as well.

Every of these units has associated a timestamp, representing the moment the event occurred since the game was started.

In the editor, we select where and when these data must be captured. For example, we can mark which type of GUI events (mouse, keyboard) we want to capture for every scene element (*LAGUIEvent*), and there is a special option to record all the GUI interactions performed in the game, allowing the recreation of the whole game play. It is possible also to capture those game effects that have relevance in the game flow *(LALogicEvent)* and, if desired, produce a *LAGameState* with the current game state. *LAGameStates* can be configured to be automatically generated periodically, when a game condition is met or when the game ends.

All these options are added to the *selection model* as part of the game's LAM, which will be used by the LAS' *data capturer*.

B. Capture

To capture all these data it is required a *data capturer* with access to all the relevant parts of the eAdventure game engine. The game engine has three main elements that are involved in this process: the *input listener*, which processes all GUI input from the user; the *game state*, which stores all the values for all the variables in the game at any given point in time; and the *effect handler*, which processes all in-game events (such as scene changes). All these elements communicate any relevant change to the *data capturer*, which then captures this information according to the current game *selection model*. The captured data are instances of the selectable data units presented before.

All these collected data are sent to the LAS *aggregator* (Fig. 5). Due to the multi-platform nature of the eAdventure game engine, different implementations take care of the communication with the *aggregator*. The *data capturer* can be configured to send out the captured data when the game is finished, a scene change happens, a defined condition is met or in real time.

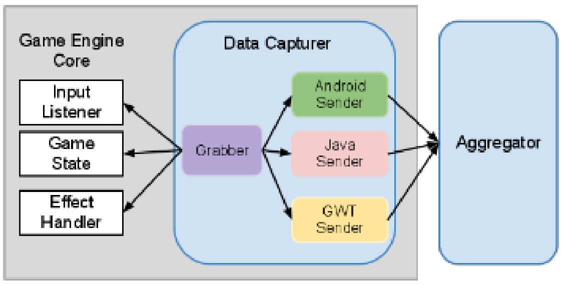


Figure 5. The *data capturer* is compound by two elements: a grabber connected with all the elements producing significant events in the game engine, and a sender managing communication with the aggregator.

C. Aggregate

The data sent are received by the *aggregator*, who makes a first data processing based on the semantic rules defined by the *aggregation model*, contained by the Adventure LAM, converting the basic units into more semantic pieces of data. This new units can be defined through the editor, as well as the rules of conversion from basic units (presented in the II.c section).

Aggregator also groups all the GUI events by type and scene element, filters redundancies and stores all the data in the LAS database.

Previous versions of eAdventure provided a basic mechanism for data aggregation and report generation. This basic mechanism allowed for information to be written in a textual report based on the values of variables in the game. This way, the game author could define a set of rules that would, for instance, write in the report that the player failed to complete a goal if a variable to indicate this was given a certain value. [13]

The same sort of data aggregation can be performed with the new LAS system. The rules in this case use a syntax that establishes the meaning of the data that were captured during the game to generate a report. The most basic reports will only include basic textual information, such as "Goal X was completed at time Y". However, more complex information can be aggregated to generate detailed information about the goal, such as "Goal X was completed by time Y, after Z attempts where the player failed to solve problems A, B, C, etc."

D. Report

The reporter represents in a web format all the information stored in the database. Among many others, the reports can be:

- A table relating scenes with the total time spent in any of them.
- Heat-maps showing where the player is hovering with the mouse most (Fig. 6).
- Screen capture recreated from game states.
- Game animations built from captured GUI events.
- Graphics showing the evolution of chosen variables in time.
- Tables with direct queries to the database.

All these data can be shown for every student or for groups of students. The system can be extended to add new reports from the stored data.

Figure 6. Heat map showing the concentration of left mouse clicks in a scene. Main heat zones are situated in interactive elements of the scene.

E. Assess

As established by the theoretical approach, this step is when the student is assessed. The assessment model contains all the goals established for the game. Goals can be defined in the editor as variables taking certain values at given times. The *evaluator* takes these goals and checks them against the stored information.

The accomplishment of these goals can be viewed through the reporter, and can be sent to the adapter to enable dynamic adaptation or to be shared with external systems.

F. Use

When the data are being captured in real time, dynamic adaptation can be used in the game. Adaptation rules are defined in the adaptation model. These rules can be defined in the editor and contain:

- *An effect*, which is considered the adaptation event and could be any eAdventure game effect (showing a text, changing a variable's value, launching a video...)
- *A condition*, establishing when the effect should be launched. Conditions can be the general conditions offered to create game logic in eAdventure games, or conditions based on goal accomplishment (e.g. a goal is not completed when the time for doing it expires).

The adapter takes the current game state and the goals information offered by the evaluator to check adaptation model conditions. When a condition is met, it communicates to the game engine the effect to be launched.

G. Refine

To support the refine task the LAS offers, through the web LAS *reporter*, information about the individual goals. This allows, for instance, the goals that lead to the worst performance to be identified. How the performance of students can be improved based on this is up to the game designer.

However, to ease this task, results obtained for the games' goals can be compared through time (checking if results improved after student played several times the game) and between different versions of the game.

H. Share

Nowadays, the selection and adhesion to standards for the content interoperability is an essential matter in the development of e-Learning contents. Current e-learning standards, like SCORM [14], are not prepared to communicate all the information collected by our LAS with other systems. For this reason, the best way of taking advantage of the full potential of our approach is to develop specific ad-hoc communication solutions for the systems that take into account all these data (e.g. a Moodle plugin). This idea can also be carried out in the eAdventure activity in LAMS [13], where all the information can be gathered and shown to educators, and use them to modify the lesson flow in an automatic or monitored way.

In the near future, it will be feasible to implement our ideas in compliance with next generation standards. For example, one the last initiatives leaded by the IMS Global Consortium, the IMS Learning Tool for Interoperability (IMS LTI), goes in that direction. This specification allows for the execution of learning tools hosted in external servers. Until other promising standards mature [15], our LAS is able to export all the information contained in the database along with the LAM required to interpret it into a exchangeable XML-based format.

IV. USE CASE: BASIC MATH GAME

We propose using the framework described in this paper in a basic math game targeted at school children. This game covers basic addition, subtraction and multiplication concepts, challenging the students to solve different problems within a given time-frame. This game is intended to provide increasing difficulty in the challenges presented to the students (e.g. the number of digits in the numbers involved is increased gradually).

The main goal of the game is sub-divided into simple goals: learn to operate with numbers of 1, 2 and 3 digits. Each time one of the goals is met, the game is adapted to provide the next level of challenge. Moreover, to provide help to the students and limit the chances that students could get stuck in any particular level, a help button is always accessible. The use of the help button is part of the information selected to be collected by the system.

In the *selection model* we marked the help button to capture all left-clicks performed over it. We also added to the model all the keyboard interactions, since these will be the input method used by the students to give their answers.

In the *aggregation model* a rule is added to convert every left-click over the help button into an "ask for help" event. Another rule is added to transform a sequence of numbers typed in the keyboard, followed by an "enter" into an "answer given" event, which its value is the introduced number and if the answer was correct.

The *reporter* shows the number of times the "ask for help" event occurred and the number of right answers. For the wrong answers, the invalid value introduced by the student is also shown. Average time for every operation is also displayed.

In the *assessment model* three goals are added: a percentage of all the given answers, after a minimum number of operations, must be correct (without using the help button) for operations with 1, 2 and 3 digits.

In the *adaptation model* an effect that changes the difficulty level (operations with 1, 2 or 3 digits) is added when the goal for the current event is achieved.

V. FINAL REMARKS

Learning Analytics, unlike Business Intelligence or Web Analytics, is still an emerging field that has a great potential. In this paper we tried to focus on a single target (i.e. educational games) in order to develop concrete methodologies trying to clarify some of the steps involved and better define the whole process. But, we think most of the ideas proposed for educational games can be extended to analyze data from other types of interactive educational resources as well, if more information about how they are being used becomes available.

The processing and logic involved in Learning Analytics can be used for other purposes different than education. The proposed LAS can help games in tasks like debugging the game design (statistics could show game points with no return), and testing (the LAS could spot if the user is playing the game as it was specified in the design).

Finally, it is important to note that the ultimate goal of Learning Analytics is to improve educational actions. We believe that learning analytics can help in establishing the educational value of games that use it. It is also important to take into account that monitoring students presents several ethics problems and privacy issues. Therefore transparency must guide all the design decisions.

REFERENCES

[1] L. Communiqué, Towards the European higher education area: Responding to Challenges in a Globalised World. Techreport. 2007.

[2] T. Elias, "Learning Analytics : definitions , processes and potential ." Learning, 23. vol: 6, issue: 4, pp. 134-148. 2003

[3] M. F. Paulsen, "Experiences with learning management systems in 113 European institutions." Educational Technology & Society. 2003.

[4] M. Brown, "Learning analytics". Learning Circuits Retrieved March, vol. 2, issue 1, pp. 1-4. 2010.

[5] S. Williams and N. Williams, "The business value of business intelligence". Intelligence, vol. 8 issue 301, pp. 30-39. The Data Warehouse Institute. 2003.

[6] M. Hassler, "Web analytics." Redline Heidelberg, vol. 3, issue 1, pp. 1-14. 2010 doi:10.1080/19322900802660292

[7] S. Dawson, A. Bakharia, E. Heathcote, "SNAPP : Realising the affordances of real-time SNA within networked learning environments." Learning. pp 125-133. 2010.

[8] J. Jovanovic, D. Gasevic, C. Brooks, V. Devedzic, M. Hatala, "LOCO-Analyst: a Tool for raising teachers' awareness in online learning environments". The 2nd European Conference on Technology Enhanced Learning, Crete, Greece, 2007, pp. 112-126.

[9] K. Gilleade, A. Dix. "Using frustration in the design of adaptive Videogames". Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology ACE 04 vol: 74, pp: 228-232. 2004

[10] V. Shute and J. M. Spector, "SCORM 2.0 white paper: stealth assessment in virtual worlds". Learning. pp. 1-10. 2008

[11] E. J. Marchiori, Á. Serrano, J. Torrente, I. Martínez-Ortiz, B. Fernández-Manjón, "Extensible multi-platform educational game framework." In press, The 10th International Conference on Web-based Learning. 2011

[12] E. J. Marchiori, J. Torrente, Á. del Blanco, P. Moreno-Ger, P. Sancho, B. Fernández-Manjón, "A narrative metaphor to facilitate educational game authoring". Computers & Education vol 58 pp. 590–599. 2012. (doi: 10.1016/j.compedu.2011.09.017). 2011

[13] Á. del Blanco, J. Torrente, E. J. Marchiori, I. Martínez-Ortiz, P. Moreno-Ger, B. Fernández-Manjón, "Easing assessment of game-based learning with <e-Adventure> and LAMS". In proceedings of the ACM International Workshop on Multimedia Technologies for Distance Learning (MTDL 2010), pp 25-30. In Conjunction with the ACM International Conference on Multimedia, October 2010, Firenze, Italy. . 2010

[14] Á. del Blanco, J. Torrente, I. Martínez-Ortiz, B. Fernández-Manjón, "Análisis del Uso del Estándar SCORM para la Integración de Juegos Educativos". IEEE-RITA Vol. 6, Núm. 3, pp. 118-127, Ago. 2011.

[15] D. Dagger, A. O'Connor, S. Lawless, E. Walsh and V. P. Wade. "Service-oriented e-learning platforms - From monolithic systems to flexible services". IEEE Internet Computing, vol. 11, issue 3, pp. 28-35. IEEE Computer Society. 2007 doi:10.1109/MIC.2007.70

6.3. Application of Learning Analytics in educational videogames

6.3.1. Cita completa

Ángel Serrano-Laguna, Javier Torrente, Pablo Moreno-Ger, Baltasar Fernández-Manjón (2014), ***Application of Learning Analytics in educational videogames,*** Revista Entertainment Computing, Vol. 5, Issue 4, pp. 313-322. DOI: http://dx.doi.org/10.1016/j.entcom.2014.02.003

6.3.2. Resumen original de la contribución

Assessment of learning contents, learning progress and learning gain is essential in all learning experiences. New technologies promote the use of new types of contents like educational videogames. They are highly interactive compared to more traditional activities and they can be a powerful source of data for all forms of assessment. In this paper, we discuss how to apply Learning Analytics (LA) with assessment purposes, studying how students interact with games. One of the biggest barriers for this approach is the variety of videogames, with many genres and types. This makes it difficult to create a comprehensive LA model for educational games that can be generally applied. In order to maintain manageable costs, we propose a two-step approach to apply LA: we first identify simple generic traces and reports that could be applied to any kind of game, and then build game-specific assessment rules based on combinations of these generic traces. This process aims to achieve a balance between the complexity and reusability of the approach, resulting in more scalable LA models for game-based learning. We also test this approach in two preliminary case studies where we explore the use of these techniques to cover different forms of assessment.

Entertainment Computing 5 (2014) 313–322

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

Application of Learning Analytics in educational videogames ☆

Ángel Serrano-Laguna *, Javier Torrente, Pablo Moreno-Ger, Baltasar Fernández-Manjón

Complutense University of Madrid, C/Profesor José García Santesmases s/n, 28040 Madrid, Spain

A R T I C L E I N F O

Article history:
Received 31 May 2013
Revised 19 January 2014
Accepted 6 February 2014
Available online 13 April 2014

Keywords:
Learning Analytics
Educational games
Educational data mining
Assessment
Reporting

A B S T R A C T

Assessment of learning contents, learning progress and learning gain is essential in all learning experiences. New technologies promote the use of new types of contents like educational videogames. They are highly interactive compared to more traditional activities and they can be a powerful source of data for all forms of assessment. In this paper, we discuss how to apply Learning Analytics (LA) with assessment purposes, studying how students interact with games. One of the biggest barriers for this approach is the variety of videogames, with many genres and types. This makes it difficult to create a comprehensive LA model for educational games that can be generally applied. In order to maintain manageable costs, we propose a two-step approach to apply LA: we first identify simple generic traces and reports that could be applied to any kind of game, and then build game-specific assessment rules based on combinations of these generic traces. This process aims to achieve a balance between the complexity and reusability of the approach, resulting in more scalable LA models for game-based learning. We also test this approach in two preliminary case studies where we explore the use of these techniques to cover different forms of assessment.

1. Introduction

Teachers of all educational levels and areas of knowledge are increasingly using new technologies to improve their teaching practices. Among others, they are starting to use Game-Based Learning (GBL) activities to explore new ways to educate their students. Studies like [1–3] support that videogame features such as high interactivity, engagement and challenge can impact the learning processes positively, both by increasing students' motivation and academic performance. The latest Horizon Reports from 2012 to 2013 [4,5] on emerging educational technologies include GBL as a technology that is almost ready for massive adoption.

In turn, assessment processes are central in educational contexts: instructional designers must assure the validity of their methods, instructors need to track the progress of the students in order to provide support and to measure the acquisition of knowledge or skills for formal grading, usually involving the use of some sort of test [6].

All these forms of assessment play a major role in current research on game-based learning. As an innovative form of content, games should undergo validation processes. Furthermore, it would be necessary to develop new forms of assessment especially devised for videogames, given that they present content in new ways and that much more student interaction data is produced with video games than with less interactive contents. In addition, the engaging interactive nature of videogames leads naturally to authentic learning tasks [7], which suggests that videogames may even be used as an assessment tool that can be better than traditional exercises [8]. However, assessment is not thoroughly contemplated in many GBL initiatives, leading to errors and lack of results. It also difficulties adoption because assessment is key in formal education, and having gaps in this area creates distrust in teachers and policy makers alike [9].

In order to increase the adoption of GBL approaches, it is necessary to create reliable assessment systems for videogames that are easy to use, that facilitate the different forms of assessment (e.g. formative, summative, etc.) and that leverage the interactive features of videogames in a cost-effective way.

Among the different perspectives from which this task can be approached, in this work we focus on the potential (and challenges) of applying the techniques typically used in Learning Analytics [10] and Game Analytics [11,12] in GBL scenarios. Learning Analytics addresses the processing and visualization of data collected from interaction and navigation through educational contents. In some cases, Learning Analytics techniques are used to predict future students' outcomes in different educational goals.

☆ This paper has been recommended for acceptance by Matthias Rauterberg.

* Corresponding author. Tel.: +34 913947623.
E-mail addresses: angel.serrano@fdi.ucm.es (Á. Serrano-Laguna), jtorrente@fdi.ucm.es (J. Torrente), pablom@fdi.ucm.es (P. Moreno-Ger), balta@fdi.ucm.es (B. Fernández-Manjón).

http://dx.doi.org/10.1016/j.entcom.2014.02.003
1875-9521/

For that purpose, Learning Analytics is usually implemented as a combination of several technologies and techniques, like data mining, Web-Analytics [13] and Business Intelligence [14]. In turn, the discipline of Game Analytics spans a set of techniques widely used by the game development industry to better understand how users play their games, find errors and improve the game play experience. For the sake of simplicity, along this paper we will simply use the term "Learning Analytics", while we may be referring to either Learning Analytics, Game Analytics or a combination of both techniques.

However, transferring these techniques to the specific challenges of assessment in GBL scenarios poses unique challenges: the potentially available data for applying Learning Analytics techniques is much bigger because interaction is very intense during a gameplay session, and the existing constructs from game analytics techniques are often game-specific (which reduces scalability) and do not cover the special requirements of educational contexts.

In order to facilitate the convergence, in this work we present a two-step approach to define a scalable Learning Analytics System that can support different forms of assessment in GBL activities. The two steps are driven by the technical requirements of scalability, uniformity and reuse of efforts: we start by defining a small and easily treatable set of generic traces, and then build higher level assessment rules by combining those generic traces, resulting in game-specific traces that do not require a full redesign of the assessment system for each individual game.

The paper is structured as follows: first, we describe the general context of this work, the challenges and the requirements. Second, we define a set of basic universal game-agnostic traces that are easily applied, gathered and processed. Then, we describe the second step, in which game-specific assessment rules can be constructed by combining the basic traces. We then exemplify the use of these steps through two exploratory case studies, in which we studied how the system would support the different types of assessment required. Finally we summarize our conclusions about the advantages and disadvantages of our approach and outline some future lines of work.

2. Games, assessment and Learning Analytics

The term assessment is often used to describe different activities and therefore it is necessary to clarify its meaning along this paper. In the literature, it is common to make the distinction between assessment *of* learning (e.g. for certification) and assessment *for* learning (e.g. to provide support for students during the learning experience) [15]. A third form is typically contemplated, namely assessment *as* learning (e.g. self-evaluation and peer-evaluation), although in this work we have focused on the first two concepts (assessment of/for learning).

In addition, when using innovative forms of content, an additional previous step is required: the assessment of the content artifact itself, in order to find design and implementation issues that may hinder the learning experience. The next subsections briefly discuss the forms of assessment considered in this work, along with their relationship with Learning Analytics.

2.1. Assessment of the game artifact

When new instructional materials are created, the assessment of their appropriateness is important, given that the extra effort required for their adoption must be justified. Indeed, most instructional design approaches are based on the generic ADDIE model [16], where the E stands for Evaluation.

Such evaluations may vary, ranging from basic "student acceptance" evaluations to complex multi-stage evaluation procedures, but as a rule of thumb the more expensive the content is, the more rigorous the evaluation should be. One of the first steps is evaluating the overall suitability of the game for the target audience and context. In this sense, de Freitas and Oliver proposed four distinct dimensions to evaluate the suitability of a specific game design, focusing respectively on *pedagogic considerations*, the *learner specification*, the *context* and the *mode of representation* [17]. However, when the proposed game is a new development, basic software validation techniques are typically used, usually in the form of formative evaluations. In some scenarios with strict formal requirements, further evaluations are typically conducted to validate the appropriateness of the game formally (e.g. by comparing their effect in a randomized trial).

When the instructional material is as complex as an educational game, testing and validating the games before their application is a significant challenge. First, making sure that a game is engaging and fun is an elusive process, since these abstract constructs are difficult to measure, and usually require applying invasive techniques. In addition, the usability of games in general and serious games in particular is complex and time-consuming due to the specific traits of this family of applications [18]. Sim, MacFarlane and Read also explored the challenges associated with understanding when an educational game is successfully performing its function [19], and Eladhari and Olilla have explored the complexity of the problem of evaluating interactive games in general [20].

From this situation, it seems that there is a lack of universal evaluation/validation procedures that can be applied to any game to detect issues and identify solutions. In this context, we believe that it may be possible to leverage the use of Learning Analytics to support the first part of the process (detecting spots where the game misbehaves or where the users get lost) so that game designers can try new solutions to avoid those issues.

2.2. Assessment of learning

Assessment with games has been traditionally dealt with by conducting debriefing sessions or regular written tests after the activity. In this sense, the information generated within the game activity is not always considered for assessment. Sometimes the reason is that the videogame lacks the necessary tools to facilitate collecting and extracting game play data. However, some games include features that can be useful for assessment, like Questions and Answers (Q&A), where the game calculates a final score based on students' answers. More rarely the games implement an ad-hoc assessment system and the results are shown as feedback (during game play or at the end) to the student.

Q&A is the natural translation of traditional exams into a game, yet this mechanism is not fully compatible with the nature of the games. Prompting the student to answer questions very frequently can break the game flow, putting engagement and motivation at risk [21]. This has resulted in a growing search for stealth assessment approaches [22,23] that evaluate the performance of the student transparently.

We expect that the Learning Analytics approach, when applied to serious games, can represent a powerful mechanism to conduct stealth assessment procedures.

2.3. Assessment for learning

Assessment for learning typically focuses on conducting early assessment not with the purpose of grading, but with the purpose of providing adequate instructional scaffolding to the students [24]. Indeed, early formative assessment is a keystone in any educational process that aims to support the students as they learn, and some authors argue that scaffolding and formative assessment are essentially the same thing [25].

In the specific context of serious games, providing this type of scaffolding requires the instructor to understand how students are interacting with the game, in order to identify those students that are either struggling or working incorrect assumptions about the content. When the number of students and the format of the gameplay sessions is adequate, merely observing the students play may be enough for the instructor, who may detect these situations and provide support when required. However, such methods are difficult to apply, and it is more common to resort to posterior debriefing sessions [26] or to using games that offer reports about how each student played the game [27].

The typical use of Web Analytics in general and Learning Analytics in particular is often aligned with the goal of identifying users that are encountering obstacles in their navigation, sometimes even in real time. When we apply the Learning Analytics approach to the field of serious games, we expect to be able to identify gameplay patterns that pinpoint students who are getting stuck within the game, or displaying low performance records. Depending on the context, this identification and the provision of additional support for those students may be performed by inspecting the gameplay traces after the session. And, if the context allows it, real-time identification of students finding difficulties would provide instructors with a powerful tool to improve those gameplay sessions by helping struggling students right when they need it.

3. A "universal" set of traces

The application of Learning Analytics techniques to serious games poses significant technical challenges. There is a great variety of educational videogames including a wide range of game genres and types;in factsome videogames can even be considered as unique pieces that blend art and software, requiring individual treatment. This variety hinders the creation of reusable assessment systems that can be applied to different videogames. However, in all videogames the player interaction and the internal game logic produce simple events that are meaningful in the development of the game flow. These events represent everything that happens in the game and can be potentially used for assessment purposes. In the following subsections we present a set of meaningful interactions, called traces, which can capture common events present in all games (see Fig. 1). These traces will be the basis to build up the assessment system

3.1. Game traces

A first step to assess the educational videogame activity in general and each student in particular, is to identify which students actually played the game and if they finished it. This is especially important if the game is designed for unsupervised use (e.g. to play after class). To cover this aspect we propose three types of traces:

- *Game Start*: This trace is generated when a student begins to play a game. We use this trace to identify the user through credentials given in the tracking system (e.g. a user and a password). If students can play the game several times, the tracking system should generate a session number for the player each time. This trace includes basic information (e.g. initial timestamp) and could also contain data about the execution context. For example, technical data as the operating system or the browser where the game runs, or any other valuable data as the system's localization and language settings.
- *Game End*: This trace is generated when the student finishes the game. This trace could contain additional information about the end reached (e.g., if the game had several endings).
- *Game Quit*: This trace is generated when the student quits the game. This trace can also contain some context that identifies the point where the student was when he quitted the game (e.g. the phase or level, a part of the game state, etc.) and about the completion status. Due to technical restrictions, sometimes games cannot generate this trace (e.g., if the game is running in a browser and the user closes the window, the game cannot send a "game quit" signal to the server). To address this problem (at least partially) another trace containing contextual information could be issued in a fixed timely ratio basis (e.g. every 60 s). The last trace received could be considered as the *game quit* trace.

3.2. Phase changes

Most educational games include some kind of internal narrative structure organized in chapters, missions or phases. These phases usually set out sub-goals that must be accomplished in the game (e.g. collect items in some phase, defeat a final boss). In some games players must fulfill all the secondary goals to accomplish the main goal. Sometimes players must complete these phases in a specific order (like levels), while in others, they can explore them

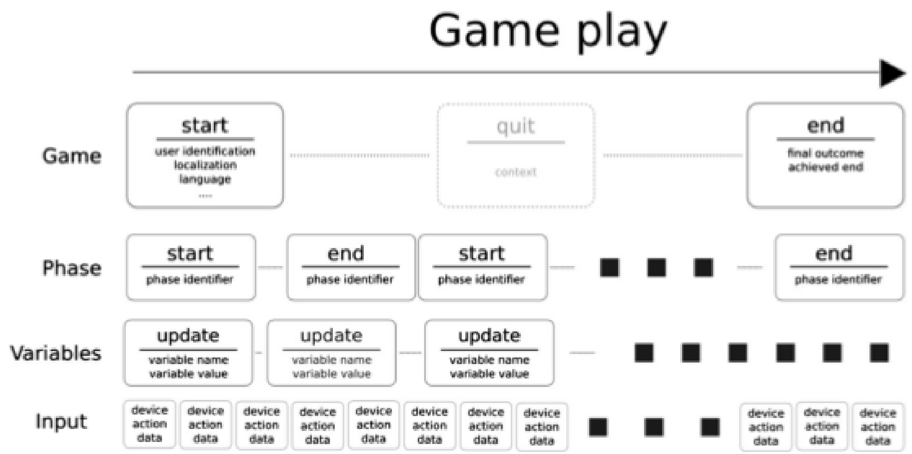


Fig. 1. Diagram with all the traces of our "universal" set, with the data associated to every of them, ordered by frequency during the gameplay.

freely. We propose logging how phases are explored with two types of traces:

- *Phase Start*: This trace is generated when the phase starts. This trace should contain an id identifying the phase.
- *Phase End*: This trace is generated when a player ends a phase. This trace could contain extra information about the completion status, like a "success" or "fail".

3.3. Meaningful variables

All games use variables to keep some sort of game state. Among these variables we can find scores or attempts used to complete the game. Posterior analysis can use final values for relevant variables to define a heuristic for player performance in the game. Variable updates over time represent the player's progress. By monitoring the evolution of a set of relevant variables over time, we can observe intermediate states the player goes through.

This type of trace can play as a wildcard to record various types of traces, as almost everything that happens in a game is registered somehow in a variable.

3.4. Input traces

The previous traces track in-game situations. They contain data related to the game logic and they are "unintentionally" generated by the game engine. But while students play, they produce a lot of data derived from their direct interaction with input devices used to control the game (e.g. mouse, keyboard, controller, joystick, etc.).

Ideally, trackers should be able to collect every single click, key or button pressed by the player. This information can be transmitted (assuming that the game engine is capable of reporting it) to a back-end and stored for latter processing. Grouping all those traces, it would even be possible to reproduce the entire play session afterwards.

However, storing traces for every single interaction can be resource consuming, demanding significant bandwidth and storage space. In most cases, the number of traces tracked can be reduced to a small subset of interactions and that is powerful enough to feed the Learning Analytics System. Therefore, while the proposed framework is designed with the capability of storing every single interaction, we explicitly contemplate the possibility of filtering either on the client side (culling irrelevant information so that it is never transmitted) or on the server side before committing the data to the database.

In any case, all input traces should contain information about the device that generated it (e.g. keyboard, mouse, controller, joystick, etc.), the action performed (e.g. button pressed, moved, clicked, etc.) and the input data associated (e.g. key press code, mouse position, controller button, etc.).

4. Analytics and reporting

The open issue is how to turn simple game and user interaction data into assessment information that could be relevant and useful for the teacher. Using general Web-Analytics techniques on the traces defined in Section 2, analysis tools can create generic reports containing game-agnostic information. For example, using *game traces* it is possible to know how many students played the game, how many finished it, and how many of them quit before finishing. As all traces can contain a timestamp, we can calculate best, average and worst completion time for a group of students. With the *quit game* trace context, we can identify in which points of the game the users stopped playing. This information can help to infer the reason why they stopped playing. When students recurrently quit the game at the same point, the authors may be pointed to a game design problem or a technical error. It could also denote a flaw in the educational content used or poorly balanced difficulty. By identifying these flaws the game can be refined iteratively.

Using *phase traces* it is possible to generate visual reports (e.g. using charts) representing how players access the game structure (e.g., represented in a directed graph), and using timestamps, how the overall time spent is distributed between all the phases. These reports also allow the identification of game design flaws by displaying excessively time-consuming phases, or which educational topics are more cumbersome for the students. If players can explore phases freely, it is also possible to show reports with the completion status of all the phases.

Using *meaningful variables* traces, we can display the final values for some interesting variables. For example, if the game has a score variable, or a variable tracking the final state of a secondary goal, their final values for every player can be shown.

Using *input traces*, we can partially (or completely, depending on the amount of input recorded) recreate the complete gameplay session. In turn, combining these traces with the information recorded by *phase traces*, it is possible to generate reports of how interaction is distributed in every phase. More specifically, we have explored the generation of heat-maps for each scene: this is achieved by aggregating the coordinates where each click of the mouse happens, and visualizing their locations on top of the scene's background. This allows the identification of those points in the screen that receive more attention from the users.

However, although this information provides insight into how students interact with the game and can provide some relevant feedback, it is not enough to support a full assessment process. More specific reporting is required to gain insight into how students are interacting with (and learning) the material.

5. Assessment conditions and goals

In the previous section we described the basic reporting approach based on simple analyses inspired in traditional web-analytics reporting. Those reports can be generated automatically for all games without a significant effort. In turn, educational assessment is typically based on goals achieved by students in certain areas. A goal is considered fulfilled when certain conditions are met, and these conditions are heavily dependent on each specific game. However, the design and collection of specific information traces for each individual game is not desirable, as the results will often be expensive.

In summary, tracing simple data is not enough by itself, and tracing advanced data is game-dependent and therefore more costly. Alternatively, we advocate for building game-specific assessment rules on top of the generic information traces described in the previous section. Our assessment process is based on this idea. As a starting point, using the "universal" set of traces, we define the following types of conditions:

- *Time from A to B is less/more than T*: Since all traces contain a timestamp, it is possible to define conditions were time passed between times A and B is less/more than a value set by the teacher. For example, time spent to complete the game must be less than 10 min; time spent to complete "phase two" must be less than 30 s.
- *Variable X has as final value equals to/less than/greater than Y*: The condition checks the value of some variables at the end of the game. X can be obtained from operating two or more variables, and Y could be a configurable value or other variable. For example, the "number of deaths" value must be less than five; the variable "score" must be greater than one thousand; variable "goal 1 completed" must be true.

- *Trace Z is present*: A specific trace was generated during game play. For example, "game finish".

More complex conditions could be defined by combining these ones with logical operators (or, and, not). Obviously these reporting rules cannot be made generic: the specific locations (*A*, *B*), or values (*T*, *Y*) are closely dependent on each specific game. Moreover, they must be carefully adjusted to match the educational objectives of the games. However, the information sources for these assessment rules are events broadly covered by the "universal" set of traces, and therefore it would be possible to create template rules that can be applied to different games, with only minor adjustments.

Combining these simple rules, it is possible to define complex goals that define the situations that are considered to reflect that the student has achieved the expected educational gain. The educational goals can be individual, or set for the group.

Including some Business Intelligence mechanisms, the teacher could set alarms if some goals are not accomplished after a certain date. Or, if some goals are generally poorly performed, teachers can make interventions to improve the results (e.g. debriefing the students about their insufficient performance, or trying to understand the reasons using the information collected).

6. Technical considerations

The generic traces have been designed with the intention of simplifying their collection from different games and platforms. However, it is obvious that the game platform used for running the games must allow collecting or generating traces as described in this paper.

In order to apply this approach in different games, it would be necessary to extend open-source game engines or to use game engines that are already prepared to track and report this information. In addition, since we expect to collect meaningful variables, the game platform must include an explicit model to represent the definition of a game. This model should be kept in separate files from the code that runs the games in a format that is easy to process by a machine.

Finally, the recorded data logged needs to be stored and collected separately. This will typically require a remote server that receives and collects the traces, storing them in a database, which increases the complexity of deploying the games in many settings (for example, primary school computers are usually behind a firewall/proxy that prevents direct access to external servers.

7. Case studies

In this section, we put into practice all the concepts presented in the previous sections using two different games developed by the e-UCM group, both implementing a tracking system similar to the one described in [28]. In each of the case studies we put into practice different aspects of this work: Case study 1 focused on assessment of the learning artifact and used reports based on both generic traces and very simple aggregated assessment rules. In turn, Case 2 focused primarily on assessment *of* learning (although we also explored assessment *for* learning), using assessment rules built by combining generic traces as described above (see Fig. 2).

7.1. The Big Party

"The Big Party" is an adventure game developed with the eAdventure platform [29]. The main goal of the game is to teach persons with psychical disabilities certain habits and common skills important for every person in daily life, like taking care of one's personal hygiene, clothing, social behavior in professional environments and how to address other people (Fig. 3). The game starts setting the players in their bathroom. They must shower and get dressed to go to a party with people from the office where they work. Once there, they must interact with their co-workers.

7.1.1. Study setup

"The Big Party" was the first game where we applied our Learning Analytics System. The game had been previously developed and tested with a group of users as part of a project for empowerment of disabled persons. However, its specific characteristics made it especially interesting for a deeper study focused on *assessment of the game artifact*. Its target audience represents a challenge for game designers, who may not anticipate which elements of the game may end up being a distraction, or causing users to get lost inside the game.

The tracking system was directly attached to the game, and we used exclusively the universal traces to gather information about how the users interacted with the product. We therefore had two objectives in this case study:

(1) Checking whether it was technically feasible to add the tracking mechanisms to a game that was developed separately.
(2) Testing whether a simple analysis of low-level interactions could be sufficient to identify game design issues and points in which the users were getting lost.

We invited 19 users from the target group, but who had not had any previous experience with the game, and allowed them to play without further directions, while the games generated the complete set of traces and sent it to the server. Then, we performed different simple analyses of the data, and tested possible visualizations that may help the designers learn about in-game issues, as well as to understand how much time users employed to complete the game. The analyses focused solely on assessing the design of the game rather than the actual acquisition of knowledge.

7.1.2. Results

Fig. 4 shows one of the reports generated for this game. To generate the *heatmap*, we collected all the *input traces* containing clicks in that concrete *phase* (i.e. those that happened before the *phase start* trace and the *phase end* trace). Each click is represented with a pale blue dot, and when an area receives more clicks the dots are clustered and we use warmer colors to identify the most clicked areas (with the color red representing the maximum number of clicks).

This report was relatively easy to create, and only used basic traces (input events and the markers for when the phase started and ended). However, it made it very easy to detect that users were mostly trying to interact with the actual active objects in the scene (toothbrush, toothpaste, comb, deodorant and shaving razor).

Interestingly, it also helped to identify minor pitfalls. For example, in the picture there is a mirror. The mirror is not an active part of the scene, and had never received any attention during the development process. However, there is a reflection in the mirror, and the analysis showed that some students actively tried to click on objects reflected on the mirror.

We also generated other reports using the available data. For example, we measured the time elapsed between *game start* and *game end* traces for every player. With this report, we could identify maximum and minimum completion times, as well as which players were faster and which ones were slower. This type of report is extremely simple, but it is still useful for detecting outlier cases that required abnormal amounts of time to complete the game.

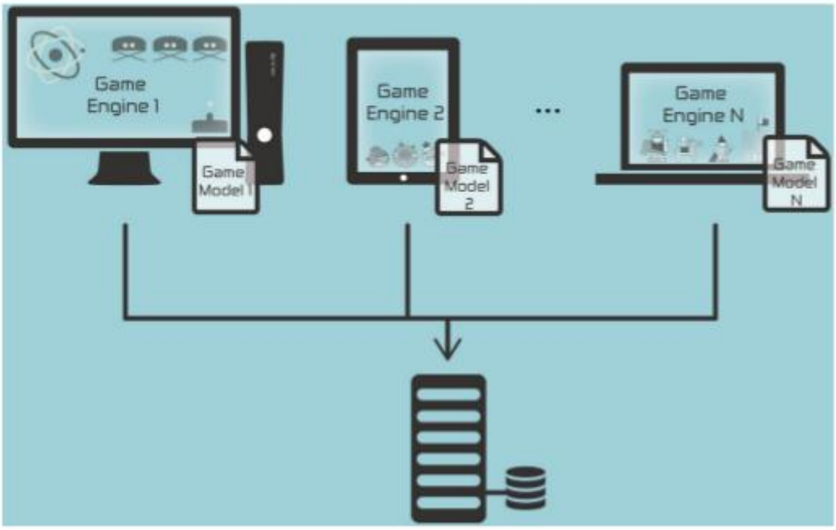


Fig. 2. General architecture.

Fig. 3. Screenshot from The Big Party game.

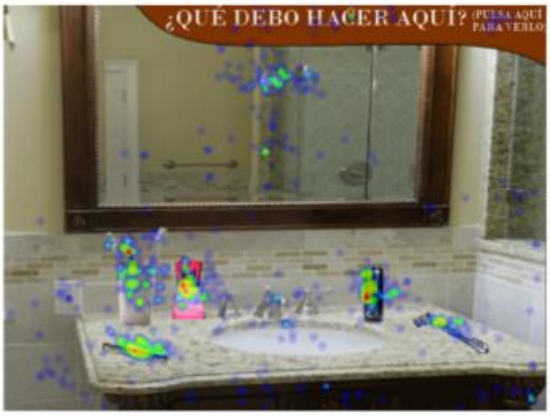


Fig. 4. A heatmap representing all the clicks performed by players in one scene.

In turn, we also used all the *phase start* and *phase end* traces to analyze how each student explored the game. Using the timestamps from these traces, it was possible to reproduce the path followed by the players. With this report, we could identify how many students were accessing "help phases" (special phases containing clues to complete the game) or getting caught in undesired loops.

Fig. 5 shows a screen capture of this animated report. It recreates how each player moved through all the phases of the game in real time (or as a later replay that can be accelerated). Each colored circle represents a student and his localization inside the game.

7.1.3. Discussion and lessons learned

The main objective of the case study was to test whether simple combinations of the universal traces would be able to provide valuable feedback about game design. The analysis based on the heatmap was especially useful as it provided an intuitive visualization. In fact, it allowed the designers to see how users were trying to interact with the mirror. The developers reacted by blurring a bit the image in the mirror so that it would receive less attention.

On the other hand, the time measurements were useful for identifying outlier figures. Two players took abnormally long to finish the game. While this does not provide directly useful information about the design of the game or the actual stumbling points, it allows the identification of important cases, prompting the designers to take a close look at the full pathway took by those specific students.

For this closer inspection, the directed graph diagram (Fig. 4) that shows the pathway of the students can be especially helpful for identification of undesired cycles or excessive amounts of time spent in a specific phase. As a last resort, having the detailed input traces it would be theoretically possible to replay the entire play session, although this would also require specific modifications of the game engine that we have not explored yet.

7.2. Lost in Space <XML>

Lost in Space <XML> is a puzzle game designed to learn basic XML and DTD concepts (Fig. 6). The game was divided in

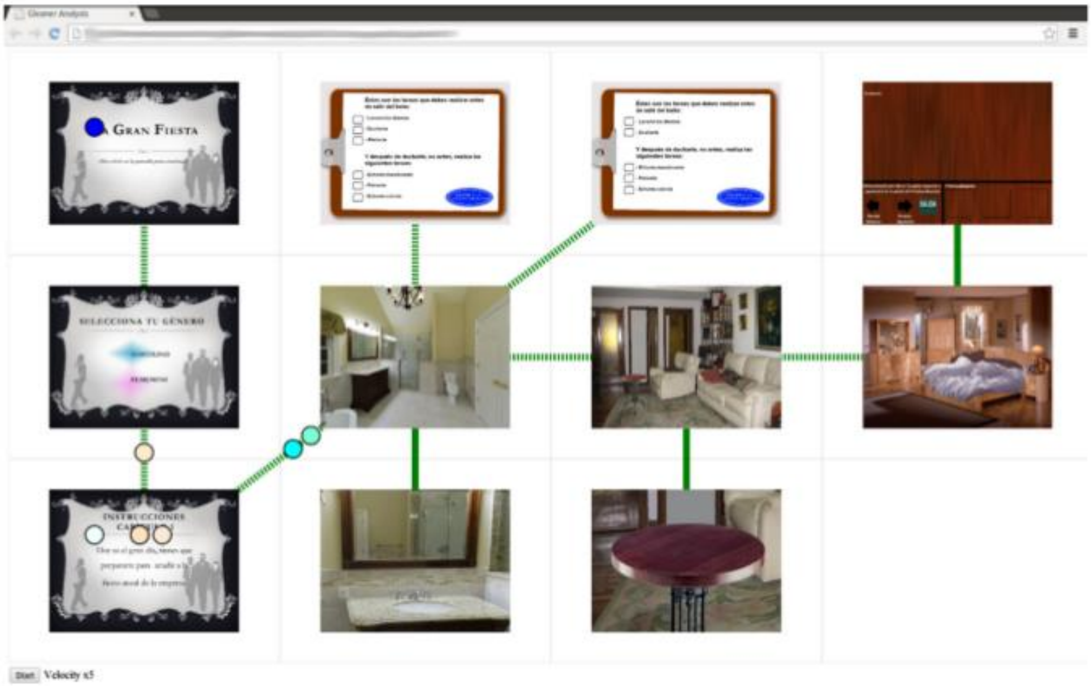


Fig. 5. Graph representing how players explore all the different phases contained by the game. Each circle represents a student and her localization in the game. Circles on edges represent students that are transitioning from one scene to another.

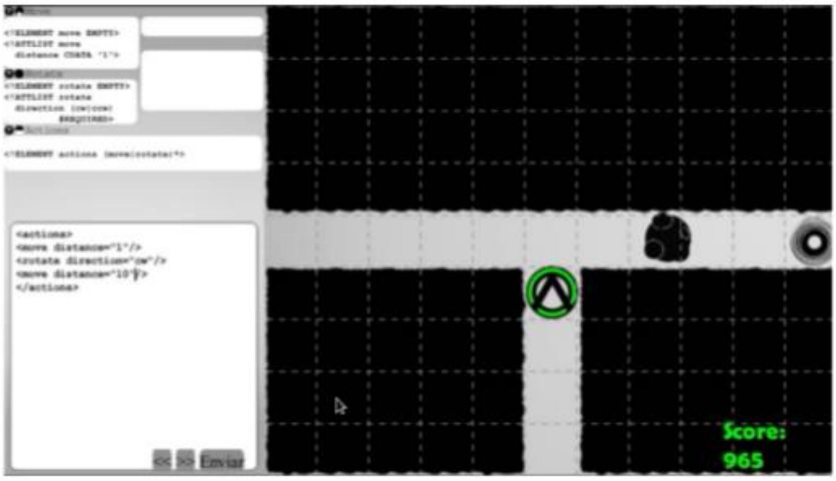
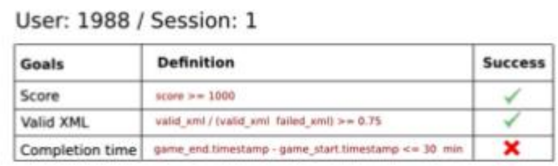


Fig. 6. Lost in Space XML screen capture. Students must enter XML documents in the interpreter (bottom left area) following a DTD (top left area) defining movements necessary to lead the spaceship to the wormhole.

phases of incremental difficulty. In each phase, students must lead a spaceship on a grid board from an origin point to a wormhole. They must type actual XML snippets (complying with a DTD that defines spaceship actions) that are then sent to an interpreter that translates the snippets into in-game actions that guide the spaceship. Some of these XML interactions include advancing one space, rotating, waiting for a few seconds or shooting.


User: 1988 / Session: 1

Goals	Definition	Success
Score	score >= 1000	✓
Valid XML	valid_xml / (valid_xml failed_xml) >= 0.75	✓
Completion time	game_end.timestamp - game_start.timestamp <= 30 min	✗

Fig. 7. Lost in Space <XML> assessment report, showing the goals out comes for one particular player's session.

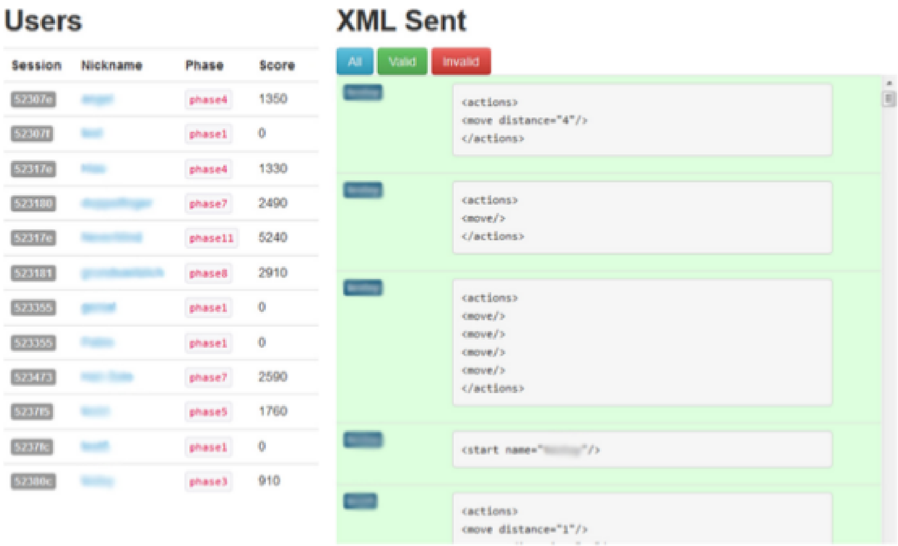


Fig. 8. Real-time dashboard for the Lost in Space <XML> assessment report.

7.2.1. Study setup

While the previous game focused on the assessment of the game artifact, this second case study focused on the other two forms of assessment (assessment of learning and assessment for learning). We wanted students to have to apply frequently the knowledge recently acquired as they played by writing actual XML documents that were the main source for assessment data.

We designed the case study to explore real-time data visualization and to generate complex scores, therefore covering the two forms of assessment mentioned above.

We set three assessment conditions, following the procedures presented in Section 5:

- *Score goal*: This focused on the universal trace based on meaningful variables. In particular, the game reported variations in the variable *score*, and the assessment condition was considered successful if it achieved a value greater than 1000 (maximum score 3000).
- *Valid XML documents*: For each session, the game also exposed a variable with the percentage of correct XML snippets. The assessment rule was activated if the percentage of correct XML documents ended up being higher than 75%.
- *Completion time*: This rule measured the time employed to complete the full game, as time from the first *phase start* to the *phase end* of the last level. The result should be less than 30 min.

Combining these traces, it is possible to create easy to read reports where the instructor can quickly verify individual progress in the game (see Fig. 7).

In addition, in order to test how these traces may be used in real time to assess the progress of each student (and perform corrections if needed), we tested a real-time visualization tool that also reported the current score and latest level successfully completed by each student.

7.2.2. Results

We tested the game with a group of 37 students from a web programming course, who played the game in a computer laboratory while the instructors watched, helped and kept track of the progress through the reporting screen.

The session was an open test of the feasibility of the approach and the usefulness of the reporting tools, and for this reason it was not structured as a formal evaluation. However 94% of the students completed the game successfully (all but two) and 81% of students achieved scores over 1000. In turn, only 24% of the students completed the game in less than 30 min.

However, the real goal of the session was not to obtain assessment of the students, but to try the approach in a real setting. The results in this sense were interesting, with the real-time visualization proving to be a very valuable tool for the instructors conducting the live session (Fig. 8).

7.2.3. Discussion and lessons learned

Through this case study, we tested the possibility of creating more complex assessment tools as combinations of the basic sets of traces. Regarding assessment *of* learning, we established different assessment conditions and goals that could use the universal traces as data sources. The result successfully allowed the instructors to get basic feedback about how well each student had performed during the play session. It should be noted, however, that these scores were exploratory, and they were not considered in computing the official grade of the students.

Regarding assessment *for* learning, the real-time tracking screen was useful for the instructors, displaying the progress and allowing the identification of those students that were having issues. The remediation (scaffolding) action was to approach the students and explain why they were failing to formulate the correct solution.

8. Conclusions and future work

In this paper we have proposed a two-step approach to apply Learning Analytics techniques to support the assessment of educational videogames. The main concern that we try to address is that, while LA techniques seem promising in general and in games in particular, their application to serious games is hard to generalize, expensive to scale and difficult to organize. While typical LA approaches focus on measuring the interactions with a web server (which are often carried out through well-established protocols), game analytics require more complex processes to gather and analyze information [12], potentially resulting in very good results, but also with a high complexity and cost. We are trying to lay the foundations for further work in applying LA in such a way that the technology would not be an issue, by offering a common ground for different vendors to provide and use games with LA subsystems.

The key idea is to propose a first layer consisting of a simple set of generic traces applicable to most (if not all) serious games. This set of traces is intentionally very simple, in order to maximize its potential spread. In turn, an additional assessment layer is used to generate more elaborated reports by tapping into those generic traces to gain insight into how the students are playing and learning. These reports serve three purposes: first, the study of the traces allows the developers to balance the game design, spotting weak points in the game deserving authors' close attention. Secondly, setting assessment conditions and rules allows the definition of success indicators that could eventually be used as a grading mechanism. Finally, the provision of real-time visualizations offers the teachers insight into how each student is progressing, therefore providing an opportunity to intervene.

We have tested this approach and its technical validity through two basic case studies created with different game engines. The first case study (*The Big Party*) focused on game validation and minor adjustments, using a game previously created with the eAdventure game platform. The second case study (*Lost in Space <XML>*) focused on testing the application in terms of assessment *of* learning and assessment *for* learning, using a browser game created with HTML5 tools. While the case studies were not formal validation experiments, they showed the feasibility of the approach. In particular, we aimed to use very simple and direct analyses, demonstrating that even small-scale tracing could yield important insights into the gameplay experience.

Regarding the assessment of the game artifact, we managed to identify design issues with these simple analyses. However, the approach is mostly valid for identifying points in which the players get confused, which does not necessarily give the designers insight on how to solve the issues. However, the identification of the stumbling point is an important and necessary step that facilitates the designers' work.

In turn, regarding the assessment of learning, the traces can be used to create reasonably complex assessment conditions that can be combined in an assessment report. We created such reports in our case study, but our main intention was to test the viability of the approach, not to actually grade the students. Further work on using such games as formal tests would require the validation of the consistency of these scores, for example by correlating them with a previously validated formal test.

Regarding its application as a tracking tool to provide scaffolding, the real-time visualizations offer a promising tool for an instructor overseeing a gameplay session. During the case study, the instructors were actually the authors of this work, and therefore this study is not a validation of whether instructors in general would find it useful for their specific teaching environments. Further work should focus on generalizing the real-time dashboard so that it can be used in different games and validated by external instructors.

Another relevant lesson learned is the difference between the kind of reporting required for fine-tuning a game and identifying pitfalls and the kind of reporting required for assessing the educational gain. Interestingly, fine-tuning can often be performed using the universal set of variables, traces and processing rules, which are independent of the specific game being studied. Therefore, the addition of these evaluation features has a reduced cost and can be easily added to different games. In contrast, assessing the educational game does require specific traces and processing rules, tailored for each specific game, given its specific design and specific learning objectives. Fortunately, we have observed that it is relatively straight-forward to define game-specific rules based on game-agnostic information traces.

We think that this is a first step towards a new model of student assessment based on educational games that can complement other methods. While these case studies are not a formal validation of the approach, our preliminary results suggest a pathway towards affordable assessment systems based on applying Learning Analytics techniques that rely on standard and easily generalizable information traces.

Acknowledgements

The following sponsors have partially supported this work: the Spanish Ministry of Science and Innovation (Grant No. TIN2010-21735-C02-02 and FPU12/04310); the European Commission, through the Lifelong Learning Programme (projects "SEGAN Network of Excellence in Serious Games" – 519332-LLP-1-2011-1-PT-KA3-KA3NW and "CHERMUG" – 519023-LLP-1-2011-1-UK-KA3-KA3MP) and the 7th Framework Programme (project "GALA – Network of Excellence in Serious Games" – FP7-ICT-2009-5-258169); and the Regional Government of Madrid (eMadrid Network – S2009/TIC-1650).

References

[1] H. Tuzun, M. Yilmazsoylu, T. Karakus, Y. Inal, G. Kizilkaya, The effects of computer games on primary school students' achievement and motivation in geography learning, Comput. Educ. 52 (1) (2009) 68–77.
[2] L.A. Annetta, J. Minogue, S.Y. Holmes, M.-T. Cheng, Investigating the impact of video games on high school students' engagement and learning about genetics, Comput. Educ. 53 (1) (2009) 74–85.
[3] H.A. Spires, J.P. Rowe, B.W. Mott, J.C. Lester, Problem solving and game-based learning: effects of middle grade students' hypothesis testing strategies on learning outcomes, J. Educ. Comput. Res. 44 (4) (2011) 453–472.
[4] L. Johnson, S. Adams, M. Cummins, NMC Horizon Report 2012 Higher Education Edition, 2012.
[5] L. Johnson, S. Adams Becker, M. Cummins, V. Estrada, A. Freeman, H. Ludgate, NMC Horizon Report: 2013 Higher Education Edition, Austin, Texas, USA, 2013.
[6] L. Communiqué, Towards the European Higher Education Area: Responding to Challenges in a Globalised World, 2007.
[7] H. Jenkins, E. Klopfer, K. Squire, P. Tan, Entering the education arcade, Comput. Entertain. 1 (1) (2003) 17.
[8] I. Garcia, A. Duran, M. Castro, Comparing the effectiveness of evaluating practical capabilities through hands-on on-line exercises versus conventional methods, in: 2008 38th Annual Frontiers in Education Conference, 2008, pp. F4H–18–F4H–22.
[9] R.T. Hays, The Effectiveness of Instructional Games: A Literature Review and Discussion, Naval Air Warfare Center, Orlando, FL, 2005.
[10] R. Ferguson, Learning analytics: drivers, developments and challenges, Int. J. Technol. Enhanc. Learn. 4 (5/6) (2012) 304.
[11] K. Hullett, N. Nagappan, E. Schuh, J. Hopson, Data analytics for game development, in: Proceeding of the 33rd International Conference on Software Engineering – ICSE '11, 2011, p. 940.
[12] M. Seif El-Nasr, A. Drachen, A. Canossa (Eds.), Game Analytics – Maximizing the Value of Player Data, Springer-Verlag, London, 2013.
[13] J. Burby, S. Atchison, Actionable Web Analytics: Using Data to Make Smart Business Decisions, Wiley, 2007.
[14] S. Williams, N. Williams, The Business Value of Business Intelligence, Intelligence 8 (310) (2003) 30–39.
[15] R. Berry, Assessment for Learning (Google eBook), Hong Kong University Press, 2008. 220.
[16] M. Molenda, In search of the elusive ADDIE model, Perform. Improv. 42 (5) (2003) 34–37.
[17] S. de Freitas, M. Oliver, How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?, Comput Educ. 46 (3) (2006) 249–264.
[18] P. Moreno-Ger, J. Torrente, Y.G. Hsieh, W.T. Lester, Usability testing for serious games: making informed design decisions with user data, Adv. Human Comput. Interact. 2012 (2012) 1–13 (article ID 369637).
[19] G. Sim, S. MacFarlane, J. Read, All work and no play: measuring fun, usability, and learning in software for children, Comput. Educ. 46 (3) (2006) 235–248.
[20] M.P. Eladhari, E.M.I. Ollila, Design for research results: experimental prototyping and play testing, Simul. Gaming 43 (3) (Apr. 2012) 391–412.
[21] J. Chen, Flow in games (and everything else), Commun. ACM 50 (4) (2007) 31.
[22] V.J. Shute, J.M. Spector, SCORM 2.0 White Paper: Stealth Assessment in Virtual Worlds, 2008.

[23] V.J. Shute, Stealth assessment in computer-based games to support learning, in: S. Tobias, J.D. Fletcher (Eds.), Computer Games and Instruction, Information Age Publishers, 2011, pp. 503–523.
[24] R. Sims, G. Dobbs, T. Hand, Enhancing quality in online learning: scaffolding planning and design through proactive evaluation, Distance Educ. 23 (2) (2002) 135–148.
[25] L.A. Shepard, Linking formative assessment to scaffolding, Educ. Leadersh. 63 (3) (2005) 66–70.
[26] V. Peters, G. Vissers, A simple classification model for debriefing simulation games, Simul. Gaming 35 (1) (2004) 70–84.
[27] P. Moreno-Ger, D. Burgos, J.L. Sierra, B. Fernández-Manjón, Educational game design for online education, Comput. Human Behav. 24 (6) (2008) 2530–2540.
[28] A. Serrano, E.J. Marchiori, A. del Blanco, J. Torrente, B. Fernandez-Manjon, A framework to improve evaluation in educational games, in: Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON), 2012, pp. 1–8.
[29] J. Torrente, Á. Del Blanco, E.J. Marchiori, P. Moreno-Ger, B. Fernández-Manjón, e-Adventure: introducing educational games in the learning process, in: IEEE Education Engineering (EDUCON) 2010 Conference, 2010, pp. 1121–1126.

6.4. Applying learning analytics to simplify serious games deployment in the classroom

6.4.1. Cita completa

Ángel Serrano-Laguna, Baltasar Fernández-Manjón (2014), ***Applying learning analytics to simplify serious games deployment in the classroom,*** Global Engineering Education Conference (EDUCON), 2014 IEEE, 3-5 abril 2014. DOI: 10.1109/EDUCON.2014.6826199

6.4.2. Resumen original de la contribución

In this paper we present our approach to introduce educational videogames as class exercises in face-to-face education. The main objective is to simplify teachers' task when using games by providing real-time information of the actual students' use of the games while in the classroom. The approach is based on defining the educational goals for the exercise/game precisely, designing a game that captures these goals, establishing relations between game interactions and educational goals and finally, create data capturing and visualizations of the relevant information to support the teacher. We applied this approach to a real case study, creating an educational videogame about the XML markup language that substituted the usual exercises in a Web Technologies class. This was tested with 34 computer science students with positive and promising results.

Applying learning analytics to simplify serious games deployment in the classroom

Ángel Serrano-Laguna
Department of Software Engineering and Artificial Intelligence
Complutense University
Madrid, Spain
angel.serrano@fdi.ucm.es

Baltasar Fernández-Manjón
Department of Software Engineering and Artificial
Complutense University
Madrid, Spain
balta@fdi.ucm.es

***Abstract*—In this paper we present our approach to introduce educational videogames as class exercises in face-to-face education. The main objective is to simplify teachers' task when using games by providing real-time information of the actual students' use of the games while in the classroom. The approach is based on defining the educational goals for the exercise/game precisely, designing a game that captures these goals, establishing relations between game interactions and educational goals and finally, create data capturing and visualizations of the relevant information to support the teacher. We applied this approach to a real case study, creating an educational videogame about the XML markup language that substituted the usual exercises in a Web Technologies class. This was tested with 34 computer science students with positive and promising results.**

Keywords—game based learning; serious games; learning analytics; classroom exercises; visualizations;

I. Introduction

Almost all teachers use practical exercises in their classes to help their students to consolidate knowledge and acquire new skills. While resolving exercises, students have to apply theoretical knowledge to solve specific problems demonstrating their understanding and identifying possible doubts or misunderstandings. This is especially important in scientific disciplines, where practice is fundamental in the learning process.

However, it is very complex to know what is actually happening in the classroom and it is even more complex when interactive content such as games or simulations is used. When teachers want to keep track of the progress of their class, among other things, they must monitor how students resolve classroom exercises. But the efforts to keep a complete view of students' progression in a normal class escalate exponentially as the number of students and exercises grow. In most cases, teachers only obtain a partial view that force them to rely on a large extent in objective evaluation (i.e. written exams) to assess their students [1]

Nowadays, new trends in education –combined with the emergence of new technologies– plead in favor of changing the number and type of interactions between teachers and students. One of these trends is Learning Analytics (LA), a discipline based on the analysis of student interactions with on-line educational resources to improve the educational process. LA results and metrics can benefit –with highly different purposes– teachers, organizations and students themselves [2].

Ideas behind LA can help teachers to keep track of the students –on a daily basis– through classroom exercises. We only need to satisfy two requirements: first, the students must resolve exercises using a connected device, so the device can communicate back the resolution process; and second, a LA system must listen to this user data, analyze it and present it to the teacher in a meaningful way.

We think that educational videogames can be a good tool to create "connected" classroom exercises. Many authors point out their qualities for education, and videogames are a perfect place to experiment and practice skills [3], which fulfills the needs for classroom exercises.

In this paper, we present our approach to successfully deploy videogames as classroom exercises whose results are automatically visualized by the teacher. This paper is structured as follows: first, we refer some related works that connects students in the classroom with the teacher using some device. Second, we define the steps of our approach and then we detail a study case in which we applied it. Finally, we discuss some results and conclusions.

II. Related work

The authors in [4] describe the process to prepare an observed pedagogical experiment defining two phases: the pre-experiment and the experiment. The first phase include two steps: the collection configuration and the structure configuration (which data to collect and how these data structures for subsequent analysis); and the second phase includes four steps: collect, structure, analyse and represent/visualize (the analysis process). In this work, authors

This work was supported in part by the Spanish Ministry of Science and Innovation (grant no. TIN2010-21735-C02-02); the Spanish Ministry of Education, Culture and Sport through the FPU programme (04310/2012); the European Commission, through the Lifelong Learning Programme (project "SEGAN Network of Excellence in Serious Games" - 519332-LLP-1-2011-1-PT-KA3-KA3NW) and the 7th Framework Programme (project "GALA - Network of Excellence in Serious Games" - FP7-ICT-2009-5-258169); the Complutense University of Madrid (research group number GR35/10-A-921340) and the Regional Government of Madrid (eMadrid Network - S2009/TIC-1650).

define “observable factors” as the low-level interactions that students perform with an e-learning system. A set of rules convert these low-level observable factors into more high-level events, with a particular meaning inside the educational context. This is approach is somehow general, and is intended to analyse abstract interactions of students. We want to extend this idea making educational goals the foundations of the process, and students assessment its result.

Authors in [5] refer one of the first examples of students using connected devices to participate in the classroom, back in 1996. Students answered direct questions using a palm-top computer (which had a full QWERTY keyboard and LCD screen) connected to a central computer. With this system, teachers presented test questions about the course content itself but also about students’ feelings and opinions of the class. This idea evolved parallel to the technology, and new research appeared with improved hardware, e.g. using RF clickers [6] or mobile phones [7].

One of the disadvantages of this type of systems is that the interaction students perform is limited, since in most cases answers consist of selecting an option or writing a word. Thus, both data for analysis and results’ visualizations are also limited. On the other hand, videogames are usually complex highly interactive applications that can present hundreds of situations and options to students therefore the data extracted can be much richer.

But simplify teacher task when using interactive content is essential for the acceptance of new technologies. Therefore, it is necessary to provide simple and useful LA data, such as visualizations. Some authors are trying to avoid the pure statistical graphs and reports, looking for more depurated views [8], and authors in [9] propose that LA visualizations should follow the idea of “goal oriented visualization”.

We take all the ideas presented in this section to define our approach.

III. Using Videogames as Classroom Exercises

We define an approach whose final purpose is to show teachers the results of their students in classroom exercises (presented in the form of small videogames). Figure 1 represents the steps in the process: 1) educational goals definition, 2) game design and implementation, 3) interaction analysis and 4) results visualization. In the next subsections, we detailed each of these steps.

A. Educational goals definition

Educational goals are the core and basis for the process. Teachers have to define what the students should learn in the exercise, and which concrete skills they want to test. Clear, concrete and precise goals will help in the subsequent steps.

The definition should follow a top-down approach, starting with a general goal (e.g. teach basics of structured programming) and then break it down into sub-goals (e.g. teach *for* loops, teach *while* loops, teach functions…). Teachers must narrow the goals domain, discard those too ambitious and keep them as simple and concrete as possible.

Each educational goal must have a binary result, (i.e. a student achieved the goal or not). Some goals can rely on a scale (e.g. student accomplished 40% percent of this goal), to give a more fine-grained result, but these results also should give a binary output (e.g. if a student accomplishes more than 60% of the goal then achieves the goal), to simplify later visualizations.

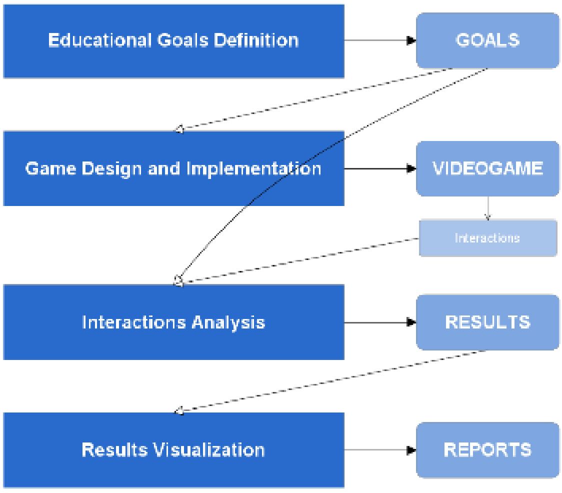


Fig. 1. Our approach follows four steps: 1. Definition of the educational goals for the exercise/videogame 2. Design and implementation of the game. 3. Interactions analysis to establish relations between educational goals and interactions 4. Visualization of the results.

Ideally, the game designer supports and gives guidance in the decisions of this step. The designer has more insight on which goals and ideas can be translated into the game and also keeps in mind requirements and constraints conditioned by the following steps of the process, like which game mechanics will be proper for the goals (if any) and which interactions will be necessary to capture to assess them.

B. Game design and implementation

In this phase, the game designer (or the game designers/programmers) takes control. His job consists in designing and implementing a game that covers all the educational goals defined in the previous step.

The design process will define the theme, the scope and the mechanics of the game. Settings like the target audience or the content subject will affect the theme. Variables like desirable time to complete the game, level of difficulty, in addition to the educational goals themselves, will define the scope. And finally, content of the goals will define for the most part the game mechanics.

Teachers will assist in the process, validating the educational and pedagogic approaches implemented by the game.

C. Translating game interactions into goals achievements

Although we treat it in a separate step, translating game interactions (the “observable factors” in [4]) into goal achievements is intimately bound up with the game design phase.

We face two aspects to connect interactions with educational goals: 1) game designer/programmer decides how the data are transmitted to the teacher (i.e. the communication-back process); and 2) teacher and designer have to define which concrete interactions *prove* that a student accomplished a goal (i.e. the analysis process).

Relation between interactions and goals can be as complex as desired. In many cases, however, it is feasible to establish direct relations between a concrete interaction and an educational goal. For example, if the game presents a puzzle that requires knowledge about two electronic components, solving it implies that the student knows those components (i.e. if the student solves the puzzle then the student achieves the goal), and we put aside complex data analysis to extract this result.

For goals that use a scale, it will be necessary to extract certain values to calculate its result. For example, in a quiz game, one goal could be the result of dividing the number of questions correctly answered between the number of questions asked.

Analysis of the results can take place in two modes:

- *In-game assessment*: the game assesses all the goals internally, and the teacher receives only the final results. In this mode, the teacher does not receive the interactions performed to achieve the goal. This mode can be appropriate for games in which relation between interactions and goals is very simple.
- *External analysis*: the game sends all the interactions to an external system, which collects, analyzes the data and finally shows the results to the teacher. This mode is more appropriate if relation between interactions and goals is more complex (e.g. several different interactions can lead to the same goal) so teachers can have more detailed information about the path followed by students for each goal completion.

D. Visualization

All the efforts from previous steps focus on giving to teachers a set of useful reports with feedback and information about students' performance.

Following the idea of goal oriented visualizations [9], reports for our approach primarily show goals achieved by each student in the concrete exercise/videogame. Combinations of students and goals can spot goals with best and worst success rates, and best and worst performers.

In addition, as secondary reports, student can also have access to some of the results from the data analysis, with auto-evaluation purposes (e.g. knowing which goals they accomplished and the knowledge associated with them).

E. Deployment in the classroom

Finally, teachers must decide the deployment of the videogame in the classroom: where students resolve the exercise (during laboratory practices or at home), if it is mandatory or optional, if the results have any impact on the students assessments, if there is a follow up session or other content related, etc.

IV. STUDY CASE

We tested our approach in a Computer Science Degree classroom, with students and content from a Web Technologies class in the Complutense University of Madrid.

In the usual mechanic for the class, the teacher first presents the lesson content through theory (with a slides presentation) and then proposes some related exercises to the students. Sometimes, they must develop a lab practice during several weeks, and others, they resolve basic exercises during classroom time, while the teacher clarifies doubts.

For this study case, we took one of the items of the course content –the XML markup language– and substituted its exercises for a puzzle videogame.

We now break all the steps taken, following the process described in section III:

A. Educational goals definition

We took as educational goals the same ones the teacher had defined for the substituted exercises, which were:

- *Create simple XML documents*: the student can create a document with a root and a few children nodes.
- *Create XML documents with attributes*: the student can use attributes in some of the nodes.
- *Create new documents interpreting a DTD*: the student can create documents based on a given DTD.
- *Create complex XML documents*: the student can create documents with several nested children and attributes.

B. Game design and implementation

The main educational goal for the exercises was to teach students to write XML documents. So we decided, inspired by some tools aimed to teach programming languages like Scratch [10], that students should introduce XML documents to control the game.

We choose puzzle game with several phases. In each phase, students must lead a spaceship to a wormhole (the exit), introducing XML documents (see figure 2). The XML documents represent actions that the spaceship can perform: move, rotate, shoot and disappear, each with several variants regulated by attributes.

C. Translating game interactions into goal achievements

The game mainly broadcasted two types of interactions: phase completions and XML documents introduced in the text area. Next, we detailed how these interactions were linked to the each educational goal:

- *Create a simple XML document*: this was the simplest goal, and it was achieved the first time a student sent a valid XML with a root and a child.
- *Create XML documents with attributes:* it was achieved after the student sent 5 valid XML documents with 5 different attributes.
- *Create new documents interpreting a DTD:* while players move through phases, they find spaceship parts that give them new abilities (rotate, shoot…). This new abilities are expressed through a DTD that is expanded every time a new part is found. In each phase the student find a new part, is mandatory to use the new ability, i.e. the student must interpret the updated DTD to finish the phase. This goal is achieved once the player passes the last phase in which a spaceship part is found.
- *Create complex XML documents:* players can complete many phases using several tiny XML documents, but they can also accomplish these phases grouping those tiny documents into a complex one (and obtaining a better score). However, there are two phases that require the use of a complex document to pass through, so this goal is achieved once the player beats these two phases.

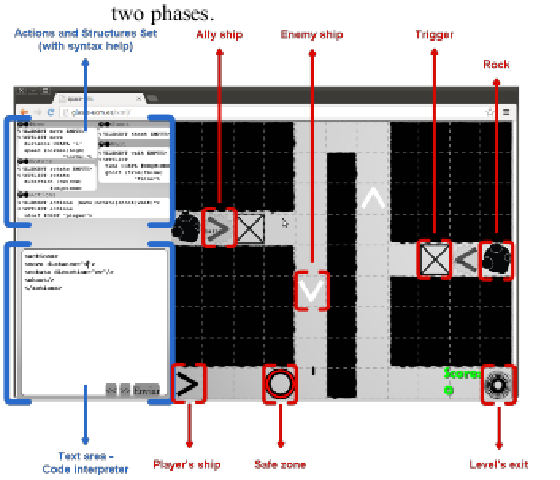


Fig. 2. Lost in Space <XML> screen capture. The goal in each phase is to lead the player's ship to the exit, introducing XML documents as orders in the text area.

As an extra (and although it has no real effect over the educational goals) the game also sent the score of each player in each of the phases.

D. Visualization

Figure 3 and figure 4 shows the teacher visualizations for the *Lost in Space <XML>* game. Figure 3 shows the results of each student individually. One row contains the data for one player in five columns: the session id, the name, the current phase, the score and the educational goals completed (in the form of badges).

Figure 4 box shows all XML documents sent by the students. This report can filter the results by student, and by valid or invalid documents. With this view, teacher could easily find those most common mistakes committed by the students, and intervene to fix them.

Fig. 3. Teacher visualization of students results. This visualization has 5 columns: a session identifier, user name, current phase, current score and educational goals achieved (in the form of badges that are "turned on" when they are achieved).

Fig. 4. Teacher visualization of XML sent by the students. This visualization contains 2 columns, one of the username and other with the XML document sent. Teacher can filter valid and invalid documents, and also filter the documents by user name.

Also, for this game, we developed a view for the students that showed them their individual results (figure 5). This view was mainly showed to students to encourage them to try again.

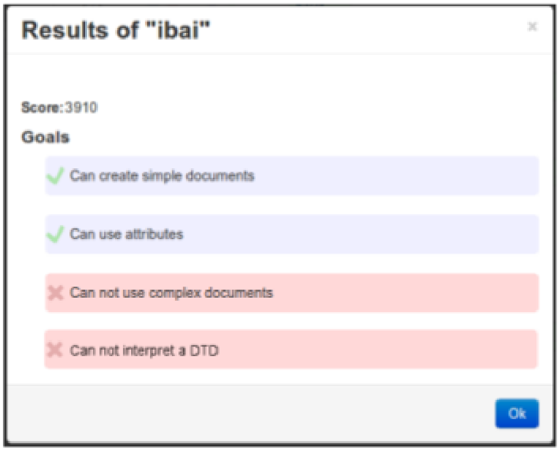


Fig. 5. Students visualization. In this report, they see their score and the educational goals achieved.

E. Deployment in the classroom

34 students played the game in a laboratory session (figure 6), two days after receiving a class about XML fundamentals. For the first half of the class, students were unaware that we were tracking their results, but we showed them the raking half way the class was through.

During the class, students enjoyed the game and remain active the whole class, exchanging comments and scores with their partners.

Fig. 6. Session with the students playing the game. Due to space restrictions, in this session students played by pairs.

V. CONCLUSIONS

In this paper, we present our approach to simplify teachers' task when using games by providing real-time information of the actual students' use of the games while in the classroom. This approach is specially focused on delivering assessment data to the teachers.

We emphasize three main aspects: define precisely the educational goal of the exercises, establish a reliable connection between game interactions and educational goals, and design clear visualizations that provide useful information for the teachers about the actual use of the games.

We tested this approach in a real case, and we obtain as results a game to teach XML basics and a tool for teachers to visualize the students' interactions and goals achievement.

Based on the experience with the study case, our approach fulfilled our needs. The goals for the *Lost in Space <XML>* game were simple but adequate four our needs, and the analysis require simple techniques, which leads us to think that for small exercises, it is not really necessary a complex data analysis.

Also, the goal oriented visualizations served the teacher to have a more complete view of the students' performance.

However, this study case was isolated, and some challenges remain opened, like how we can integrate results from several exercises/videogames and the long term effect on teachers' perception.

VI. REFERENCES

[1] "Towards the European Higher Education Area: Responding to Challenges in a Globalised World," in *Communiqué from the Ministerial meeting in London*, 2007.

[2] P. Long and G. Siemens, "Penetrating the Fog: Analytics in Learning and Education," *Educ. Rev.*, vol. 46, no. 5, pp. 31–40, 2011.

[3] K. Squire, "Video Games in Education. International Journal of Intelligent Simulations and Gaming," *Int. J. Intell. Simulations Gaming*, vol. 2, no. 1, pp. 49–62, 2003.

[4] J. M. Carron, T., Marty, J. C., France, L., & Heraud, "Preparing an Observed Pedagogical Experiment," in *CELDA*, 2005, pp. 526–531.

[5] R. J. Dufresne, W. J. Gerace, W. J. Leonard, J. P. Mestre, and L. Wenk, "Classtalk: A classroom communication system for active learning," *J. Comput. High. Educ.*, vol. 7, no. 2, pp. 3–47, Mar. 1996.

[6] L. Lugaric, M. Delimar, S. Krajcar, and I. Rajsl, "Real-time student assessment using a system of RF clickers," in *ITI 2008 - 30th International Conference on Information Technology Interfaces*, 2008, pp. 83–88.

[7] R. Jana, Y.-F. Chen, D. C. Gibbon, Y. Huang, S. Jora, J. Murray, and B. Wei, "Clicker - An IPTV Remote Control in Your Cell Phone," in *Multimedia and Expo,*

2007 IEEE International Conference on, 2007, pp. 1055–1058.

[8] L. France, J.-M. Heraud, J.-C. Marty, T. Carron, and J. Heili, "Monitoring Virtual Classroom: Visualization Techniques to Observe Student Activities in an e-Learning System," in *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pp. 716–720.

[9] E. Duval, "Attention Please! Learning Analytics for Visualization and Recommendation," in *Learning Analytics and Knowledge*, 2011, pp. 9–17.

[10] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, and Y. Kafai, "Scratch: Programming for all," *Commun. ACM*, vol. 52, no. 11, pp. 60–67, 2009.

6.5. A methodology for assessing the effectiveness of serious games and for inrring player learning outcomes

6.5.1. Cita completa

Ángel Serrano-Laguna, Borja Manero, Manuel Freire, Baltasar Fernández-Manjón (2017), ***A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes,*** Multimedia Tools and Applications [Aceptado y pendiente de publicación] [Factor de impacto JCR: 1.331; Segundo cuartil (Q2) 31/106 en "Computer Science, Software Engineering"]

6.5.2. Resumen original de la contribución

Although serious games are proven to serve as educational tools in many educational domains, there is a lack of reliable, automated and repeatable methodologies that measure their effectiveness: what do players know after playing serious games? Do they learn from them? Previous research shows that the vast majority of serious games are assessed by using questionnaires, which is in stark contrast to current trends in the video game industry. Commercial videogame developers have been learning from their players through Game Analytics for years via non-disruptive game tracking. In this paper, we propose a methodology for assessing serious game effectiveness based on non-disruptive in-game tracking. The methodology involves a design pattern that structures the delivery of educational goals through a game. This structure also allows one to infer learning outcomes for each individual player, which, when aggregated, determine the effectiveness of a serious game. We tested the methodology by having 320 students play a serious game. The proposed methodology allowed us to infer players' learning outcomes, to assess the game effectiveness levels and to identify issues in the game design.

Multimed Tools Appl
DOI 10.1007/s11042-017-4467-6

A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes

Ángel Serrano-Laguna [1] · Borja Manero [1] · Manuel Freire [1] · Baltasar Fernández-Manjón [1]

Received: 20 July 2016 / Revised: 13 December 2016 / Accepted: 31 January 2017

Abstract Although serious games are proven to serve as educational tools in many educational domains, there is a lack of reliable, automated and repeatable methodologies that measure their effectiveness: what do players know after playing serious games? Do they learn from them? Previous research shows that the vast majority of serious games are assessed by using questionnaires, which is in stark contrast to current trends in the video game industry. Commercial videogame developers have been learning from their players through Game Analytics for years via non-disruptive game tracking. In this paper, we propose a methodology for assessing serious game effectiveness based on non-disruptive in-game tracking. The methodology involves a design pattern that structures the delivery of educational goals through a game. This structure also allows one to infer learning outcomes for each individual player, which, when aggregated, determine the effectiveness of a serious game. We tested the methodology by having 320 students play a serious game. The proposed methodology allowed us to infer players' learning outcomes, to assess the game effectiveness levels and to identify issues in the game design.

Keywords Serious games · Learning analytics · Game design · Learning outcomes analysis · Educational games

1 Introduction

Serious games are video games designed for purposes beyond pure entertainment [25]. Serious games are multimedia tools by nature. As a subfamily of videogames, they combine different forms

✉ Ángel Serrano-Laguna
angel.serrano@fdi.ucm.es

1 Department of Software Engineering and Artificial Intelligence, Facultad de Informática, Complutense University of Madrid, C. Profesor José García Santesmases, s/n., 28040 Madrid, Spain

Published online: 15 February 2017

of media (animations, music, text, etc.) to create immersive experiences for players. Their versatility allows them to be used as tools with many applications in different domains. One of the main applications is education, whereby they have become proven learning tools: they are used across several domains with multiple goals and formats, and their acceptance and effectiveness has almost always proven positive [10, 39]. Traditionally, a large percentage of serious games has been both developed and deployed by educational researchers, limiting their scope and reach. This trend is beginning to change. Currently, the widespread use of Virtual Learning Environments (VLE) has allowed for the application of serious games at unprecedented scales. To reach their full potential, serious games should apply the latest advances in education and commercial videogames [26].

On-line education has increased exponentially in recent years, and many students now learn through Internet-connected devices. This has vastly increased the amount of educational data available for analysis. Disciplines such as Learning Analytics (LA) or Educational Data Mining (EDM) study patterns of student interactions to better understand underlying learning processes [7, 16]. This information can be used by different stakeholders for various purposes: from university administrators calculating dropout rates for each class to teachers identifying students at risk of course failure [18].

Serious games (and video games in general) are particularly well suited for data analysis. Their highly interactive nature based on a constant loop of user input followed by game feedback designates them as rich sources of interaction data. These interactions can be analysed to explore how users play and in the case of serious games to understand how users learn.

The video game industry has been performing these types of analyses on commercial games for years via Game Analytics (GA) [17]. One of the main uses of GA is to measure balance in gameplay: a balanced video game is one that keeps its players in the flow zone, a state wherein the player feels challenged by the game but is neither bored nor frustrated [8]. GA is used to locate points of gameplay at which players become stuck or quit as well as moments at which a game's mechanics or internal rules fall short. GA is also used to identify clues on ways to fix these problems.

Commercial video games typically non-disruptively collect data from their players through tracking systems that go unnoticed by players [36]. However, according to the literature [5], aspects of serious games are typically assessed through questionnaires completed by players. There is a clear need to combine the emerging disciplines of LA and EDM with the non-disruptive techniques of GA to generate reliable, automated and repeatable assessments of serious games.

Serious game assessments can focus on many outcomes, such as usability, engagement or motivation. However, the learning outcomes are the results most stakeholders wish to obtain from serious games [1]. Learning outcomes have also been the results most frequently assessed when examining recently developed serious games [5], and some authors even believe that such outcomes could be used to replace standardized tests [3]. However, multiple issues with serious games must first be addressed. One pertains to a lack of methods available to assess serious game effectiveness [40]: teachers, lecturers and policy-makers need to guarantee that serious games are effective enough to be used in the classroom. In this regard, the application of GA techniques to serious games can provide stakeholders with objective and reliable data.

In this paper, we propose a methodology for inferring learning outcomes and serious game effectiveness based on non-disruptive tracking. The methodology targets two different phases in the life of a serious game: 1) its design and implementation, for which we propose a game-design pattern to shape the delivery of educational content throughout a game, and 2) its validation and deployment, for which we propose an analysis based on the game-design pattern to infer learning outcomes and game effectiveness levels.

The paper is structured as follows. Section 2 presents a literature review of serious game assessment methods. Section 3 presents the methodology proposed, and section 4 describes an experimental case study in which the methodology was applied. Section 5 presents the results of the case study, which are then discussed in Section 6. Finally, Section 7 presents our conclusions, some limitations, and avenues for future work.

2 Serious game assessment

Although questionnaires are most commonly used to assess serious games [5], several authors have addressed the implications of using non-disruptive tracking methods for this task. Authors have proposed a set of minimum requirements to enable the automatic assessment of serious games [32] and have addressed the game design implications of combining learning analytics with serious games [20]. The ADAGE project [34] is a framework that defines several "assessment mechanics" that capture basic gameplay progression and critical achievements. Similarly, we have previously proposed a set of universal "traces" of particular interest in the case of serious games that can be emitted through any video game [37].

Other authors have implemented their own ad-hoc analytics, for instance, to analyse players' steps taken while completing a math puzzle to predict their movements based on current game states [24], to assess learning outcomes by analysing answers to quizzes integrated in a game [15], and to analyse how players progress through learning-language courses to create rich visualizations for teachers [41].

We note that serious game designers must take into account analytics and assessment constraints from a game's inception and throughout the design phase [32]. Many authors have defined methodologies and guides for designing serious games [3, 11, 12, 14, 30]. However, this body of research proposes methodologies that are applicable to any analytics-aware video game, serious or not. In particular, these works do not typically address key serious game features, such as ways to deliver knowledge and educational content through gameplay or ways to infer corresponding learning outcomes. Some work has started to explore these issues, proposing a taxonomy of possible elements that a serious game should include to be more effective [6].

To summarize, we found research that describes effective analytics-aware serious game design, but which lacks reference to concrete methodologies for inferring learning outcomes. On the other hand, some works have proposed ways to analyse serious game learning outcomes either via general frameworks or ad-hoc analysis, but without addressing the implications of such assessments for game design. We propose combining both approaches in defining a methodology that tackles all phases of serious game development: from game design and implementation to deployment and learning outcome analysis.

3 Proposed methodology

Our methodology pursues two goals: 1) to ease the measurement of serious game learning outcomes and 2) to provide a systematic way to assess the effectiveness of serious games as a whole. To achieve these goals, our approach covers the complete lifecycle of a serious game (Fig. 1). The process starts in the design phase, when the learning goals and target population forms the basis for creating a learning and game design. The combination of these designs is

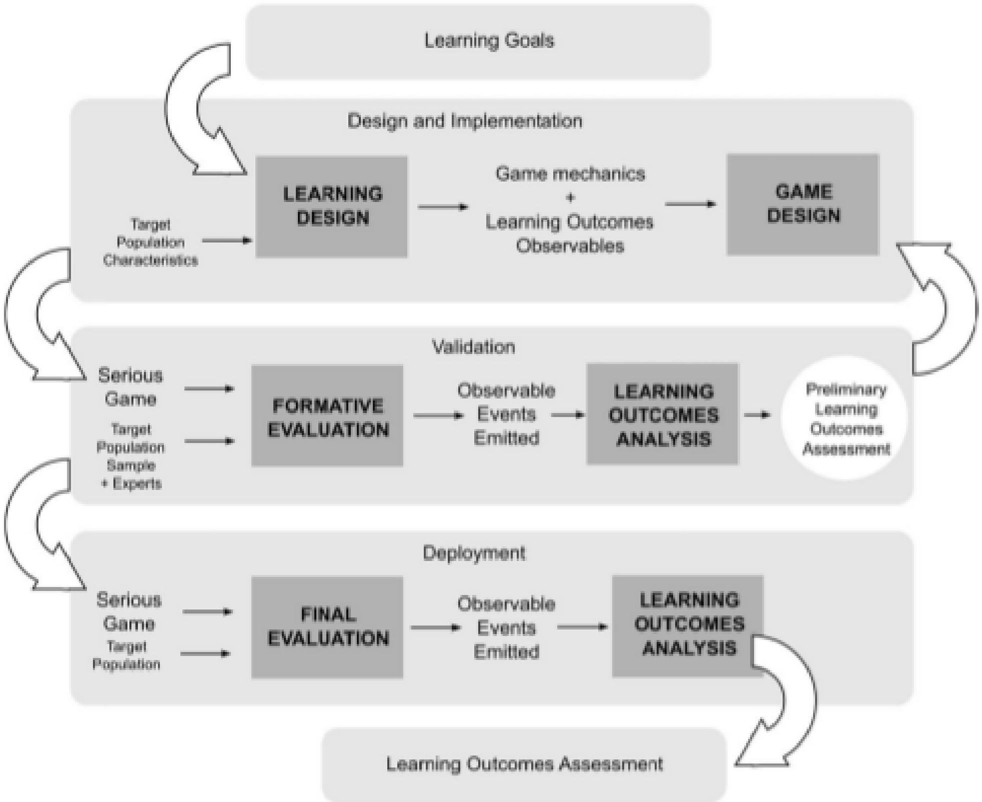


Fig. 1 Serious game design and deployment process with zlearning outcomes assessment

used to create the game, which is then validated through a formative evaluation with a sample of the target population. This process is repeated until the game is fully validated. The game can then be used by the target population (deployment). In the following subsections, we describe each step of the process in greater detail.

3.1 Design and implementation

Within the context of our methodology, we define "learning design" as the transformation of learning goals into game mechanics and learning outcomes observables based on characteristics of the target population.

The chosen game mechanics should fulfil two requirements: 1) they should be appropriate for the learning goal content based on models, such as that presented in [4], in which learning mechanics are mapped to game mechanics; 2) gameplay from players should produce learning outcome observables (also termed *events*) that attest to the players' knowledge or level of skill.

During game design, these constraints, along with many other considerations for a given game (such as art styles, storytelling or technologies) shape the creation of a serious game. Additionally, during this phase, designers must define how a serious game should scaffold the delivery of learning goals. Although there are few concrete methodologies that translate educational theories into game design aspects [23], some authors have proposed models that describe the learning processes of videogames. For instance, in the serious games domain, Kiili proposes the experiential gaming model [22] and problem-based learning [23], both of which are based on an iterative process through which players form a strategy, experiment in the game world, receive feedback, and reflect on the results. In the commercial videogames domain, similar proposals have been made to split experimentation in two sub-steps: experimentation in a “safe game environment”, whereby the level of the difficulty of the challenge to overcome is low and mistakes are not punished, and experimentation in an “unsafe game environment”, where the level of the difficulty is higher and mistakes are punished (e.g., losing game lives, coins, score levels, etc.) [33].

For the purposes of our methodology, we have combined and extended these principles into a game design pattern that also considers learning outcome observables. Each learning goal is presented to players throughout 3 phases based on 2 points of non-disruptive measurement (Fig. 2):

1) **Strategy**: Players are first introduced to the learning goal. This can involve knowledge they might need in subsequent steps as well as concrete instructions on how to interact with the game world (e.g., through non-interactive scenes or game tutorials). The player receives information on the challenge behind the learning goal and can start identifying ways to tackle the challenge. The strategy phase is similar to initial exposition plus exploration behaviours that are very common of games.
2) **Practice**: players start to apply the knowledge presented in the previous phase. This practice must occur in a game environment in which players” mistakes have either no consequence at all or only mildly adverse consequences (“safe game environment”). This experimentation must be designed in such a way that players can make deductions and test hypotheses on both the knowledge presented in the previous phase and on the game’s mechanics. In this phase, students test and practice their strategies. Strategies that work better will later be refined by the player during the mastery phase.

   During this phase, players apply knowledge associated with the current learning goal for the first time. This allows us to collect initial observables from which their initial knowledge can be estimated.
3) **Mastery**: players are required to prove that they have acquired the intended knowledge while facing challenges similar to those presented in the practice phase, but with increasing difficulty and higher in-game consequences (“unsafe game environment”).

   During this phase, players prove the degree to which they have acquired the targeted skill or knowledge. We can in turn collect final observables that allow us to measure their final progression towards a learning goal.

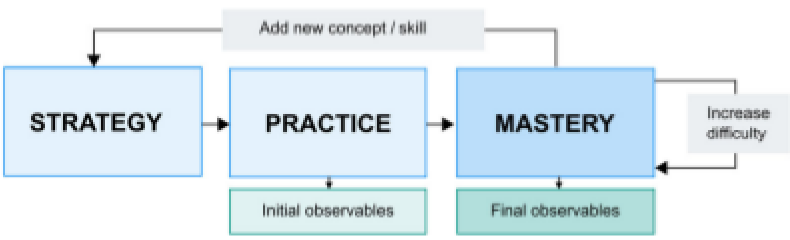


Fig. 2 Game design pattern phases. The phases would be repeated for each sub-goal. Observables are emitted during the practice and mastery phases

These three phases can be iterated to deliver multiple learning goals or to deliver a single goal with increasing difficulty while adding a new related concept or skill during each cycle. Additionally, this game pattern optimizes the period in which players occupy the flow zone [8], as it alternates phases in which players learn new things in a safe environment (practice) with phases in which they are challenged to prove their skills (mastery) and through an incremental approach to prevent frustration.

3.2 Collecting observables

Players perform different interactions to advance in a game: they make choices, resolve puzzles, beat bosses, and so on. These events are the core observables on which we perform our learning outcomes analysis. The following principles (many of them shared with general GA) facilitate this analysis:

1. Observables should result in time-stamped events that describe simple interactions between the player and the game [37]. These events should be sent to a central server where all player interactions are stored for later access and analysis.
2. Events sent to the server should be raw interactions rather than opaque scores [37, 38]. For instance, if the mastery phase involves two puzzles, the events to transmit would be the interactions performed to resolve the two puzzles rather than a combined score of the final result. This ensures flexibility, as scores can be later recalculated from interaction data if the subsequent analysis reveals a need to do so.
3. Data collection should be as non-disruptive as possible during gameplay. Ideally, game flow should never be interrupted to collect data – players should not be explicitly asked to stop their play to pass an exam or to answer questions not integrated in the gameplay.

Once all interaction events are stored in a central location, analysis can begin.

3.3 Learning outcome analysis

We store all gameplay interaction events on a single server. Following our design pattern, each interaction is associated with a learning phase (strategy, practice or mastery) of a specific learning goal. Interactions from the strategy phase are not used to infer learning outcomes (as this phase should only contextualize the learning goal). By analysing interactions from the other two phases, we can calculate two assessment scores:

1. **Initial assessment** (*IA*) using initial observables of the practice phase. IA estimates the learner's initial level of knowledge. A high value denotes that the player likely possessed the targeted knowledge before starting to play while a low value denotes the opposite.
2. **Final assessment** (*FA*) using final observables of the mastery phase. FA estimates the learning outcome. A high value denotes that the player achieved the learning goal while a low value denotes his or her failure to do so.

The specific steps used to transform events into *IA* and *FA* will be different for each serious game. However, they can generally be expressed through formulas that combine data from each interaction. In section 4, we provide details on this process through a real case study.

We define two assessment thresholds: an initial threshold (*IT*) associated with the *IA*, and a final threshold (*FT*) associated with the *FA*. These thresholds are used to determine whether a phase is successfully accomplished or not. For instance, when *FA*'s value ranges from 0 to 1, a possible value for *FT* could be 0.5, and so we assume that a player achieving an *FA* value of equal to or greater than 0.5 has successfully completed the mastery phase.

For serious games that include multiple learning goals, we can calculate their global *IA* and *FA* values using a weighted average while combining the results of each learning goal: for a game with *N* educational goals, each with two assessments (IA_i, FA_i), two thresholds (IT_i, FT_i) and a weight (W_i), we can calculate the global assessment value (*A*) of the initial and final assessments as:

$$\mathrm{A} = \frac{\sum_{i=1}^{N} \mathrm{A_i} \times \mathrm{W_i}}{\sum_{i=1}^{N} \mathrm{W_i}}$$

and the global threshold value (*T*) for the initial and final thresholds as:

$$\mathrm{T} = \frac{\sum_{i=1}^{N} \mathrm{T_i} \times \mathrm{W_i}}{\sum_{i=1}^{N} \mathrm{T_i}}$$

With these values, we can now estimate learning outcomes and assess serious game effectiveness levels.

3.3.1 Inferring players' learning outcomes

The analysis of observables or signals provides two measures for each learning goal: *FA* and *IA*. With these values, we can measure two concrete learning outcomes:

- ***FA* as the player's final score**: We can use *FA* as a score or mark for players when they are considered as students (essentially scoring what they know after playing the game). We must avoid using *IA* to calculate this marker. Although it represents a player"s level of knowledge, using it to calculate final marks would be unfair, as *IA* takes into account mistakes made during the practice phase while a fair grade should only consider what students know at the end of the game and not what they initially ignored.
- **The difference between accomplishments in the practice and mastery phases as game effectiveness**: If we compare *IA* and *FA* to their respective thresholds (*IT* and *FT*), we can determine whether a player has succeeded in the practice and mastery phase. A game is most effective when players who failed in the practice phase ended up succeeding in the mastery phase, as this denotes a knowledge gain. This difference forms the base from which we calculate serious game effectiveness.

3.3.2 Assessing serious game effectiveness

Within the context of our methodology, we assume that a serious game is effective when we find a positive change in the player's knowledge level. We can determine this change from *IA* and *FA* with respect to *IT* and *FT*. From these values, we can classify each player into a different learning category:

- When $FA \geq FT$, the players have successfully completed the mastery phase and have acquired the targeted skill. Depending on the *IA* value, we can classify players as either:
  - **Learners**, when $IA < IT$: players committed errors during the practice phase, indicating that they did not possess the targeted skill or knowledge before playing the game. However, they ended up being successful in the master phase, suggesting an educational gain during gameplay.
  - **Masters,** when $IA \geq IT$: the players did not commit errors during the practice phase, indicating that they likely already possessed the skill or knowledge before playing the game.
- When $FA < FT$, the players failed the mastery phase and do not possess the targeted skill. Depending on the *IA* value, we can classify players into two different categories:
  - **Non-Learners**, when $IA < IT$: the players also failed the practice phase, indicating that they struggled throughout the game with potentially little or no benefit.
  - **Outliers,** when $IA \geq IT$: the players succeeded during the practice phase but were unable to apply the acquired knowledge in the mastery phase.

We determine serious game effectiveness by classifying each gameplay session according to these criteria and by then comparing the total number of players in each category.

When the majority of players are learners, the game is considered highly effective: most players learned something while playing. When the majority are masters, the game is considered to have no learning effect, as most of the players had already possessed the targeted knowledge before playing the game. When the majority are non-learners, the game is considered not effective at all, as most of the players were unable to achieve success at any phase. Finally, a majority of outliers denotes that a game and/or the chosen *FA* and *IA* formulas likely present design flaws.

It is important to note that most serious games will output different results for different populations. A serious game could be highly effective for children of 10 to 12 years of age and not effective at all for children aged 7 to 9 years. The key is to have a well-defined target population during the design of a serious game and to follow a validation process to ensure that effectiveness goals are met.

3.4 Validation and deployment

After applying a serious game along with infrastructure to track its observables in relation to a learning outcomes analysis, we must to validate it.

During the validation phase, domain experts and ideally a sample of the target population play a serious game and engage in gameplay that is later assessed through a learning outcomes analysis, yielding preliminary results through a process typically referred to as formative evaluation [19]. This process is iterative and designed to detect ways to fix, polish, tweak or improve a serious game that can range from changing the game mechanics of a learning goal (e.g., when preliminary results suggest low performance) to altering how *FA* and *IA* are calculated (e.g., when experimental results contradict certain game design hypotheses).

Once a game is validated, it can be used in production for final deployment. In this final phase, the serious game and its learning outcomes analysis results are used to assess students that play it (final evaluation).

4 Case study

In the above sections, we presented our methodology for modelling and inferring learning outcomes and effectiveness in serious games. This section describes a case study that illustrates how this methodology works when applied. The case study is based on the following research questions:

RQ1. What are the implications of using our game-design pattern during the design and implementation of a serious game?
RQ2. What results in regards to learning outcomes and effectiveness levels can be obtained from a serious game developed and analysed through this methodology?

To answer these questions, we used the proposed methodology to implement and analyse "The Foolish Lady", a serious game[1] based on the homonymous theatre play by Spanish playwright Lope de Vega. In this game, players are presented with several language-related and literature challenges. Its main learning goal is to teach high school students about Spanish Golden Century poetry. In the following subsections, we describe the design and implementation process, the data collection and analysis process, and the results of an experiment on 320 high school students who played the game.

4.1 Design and implementation

"The Foolish Lady" serious game [27, 28] is an adventure game based on a classical Spanish play. In the game, players advance through scenes of the play by making decisions that affect the overall storyline and final scene. Along the way, they are presented with puzzles and mini-games in which they must apply their knowledge on language and literature. The game is designed to be completed within 30 to 40 min.

One of its main learning goals is to teach poetry structure and rhymes, in particular, "redondilla", a Spanish poetic composition form that uses a specific rhyming scheme and verse length. During the learning design phase, we chose to use point-and-click mini-games as our game mechanic, which relies on drag-and-drop puzzles and option selection in conversations with non-playable, in-game characters. This approach is typical of adventure games, a genre with a track record of proven educational benefits [13]. During the game's design, we subdivided our goal into the three phases according to our design pattern. Figures 3, 4, and 5 show in-game screen captures representing each of the three phases.

Players are first presented with a textual description of rhymes and of the "redondilla" structure (Fig. 3). These instructions appear in two non-interactive scenes that can be skipped (after reading the content or not) with a click. These scenes belong to the *strategy phase*.

[1] Available (in Spanish) at https://play.google.com/store/apps/details?id=es.eucm.androidgames.damaboba.

Fig. 3 The game presents basic features of the "redondilla" on two screens with textual explanations

Later on, players are presented with a mini-game in which they must complete a poem composed as a "redondilla" (Fig. 4). The poem is missing five words and the players can fill in the blanks by dragging words from a container on the right side of the screen. Once filled, they can check the correctness of the poem they have created by clicking a check button. They can try to do this as many times as they wish until they find the right combination of words: as the *practice phase* of the goal, the results of this mini-game are irrelevant to the final score.

Finally, players are presented with two mini-games (Fig. 5). In the first one, players must fight a knight by exchanging rhyming verses. A player can win this battle if he or she selects three correct rhyming replies in a row and loses it if he or she fails three times in a row. The player's score decreases with each error. In the second and final mini-game, the foolish lady"s father assesses the protagonist's suitability as a son-in-law by asking the player a series of questions on the "redondilla" poetic format. Players can answer these questions only once, and both the score and the game protagonist's marital prospects decrease when they fail. Both mini-games belong to the *mastery phase* of the goal, and therefore the results of these games affect the final score.

Fig. 4 In the first puzzle, players must apply their knowledge of the "redondilla" format. They can try to do this as many times as they wish

Fig. 5 In the last mini-games, players must prove their knowledge. In both cases, the player's score decreases with each error

4.2 Collecting observables

To record and analyse the gameplay sessions of all students, we developed a framework composed of a *tracker* bundled within the game itself for sending interaction events (observables) and a *collector server* for receiving and storing events. The types of events are fully detailed in [37, 38]; here, we only highlight those events relevant to the learning outcomes analysis:

- Events resulting from a new attempt to complete the "redondilla puzzle". Every time the player clicks the "check" button and the result is incorrect, a new attempt is made.
- Events resulting from a new attempt to beat the "fight mini-game". Every time the player loses the fight and restarts the mini-game, a new attempt is made.
- Answers chosen by the player during the final mini-game.

The game itself does not make any assessment calculation: only raw events are sent to the server.

4.3 Learning outcomes analysis

All players encounter the 3 mini-games during their playthroughs: the "redondilla" puzzle mini-game in the practice phase, and the fight and test mini-games in the mastery phase. For each mini-game, we calculate a score of between 0 and 1:

- **Redondilla Game score (*RG*)**: when A is the observable representing the number of attempts made to solve the "redondilla" puzzle mini-game, RG is computed using $RG = 1 - (\text{MIN}\ (A - 1, A_{MAX})\ /\ A_{MAX})$, where A_{MAX} is the reasonable number of attempts needed to solve the game. The initial assessment result takes a value of 1 when the player beats the puzzle on the first attempt, i.e., $A = 1$. The initial assessment result takes a value of 0 if the player does not complete the puzzle on any attempt or tries over A_{MAX} times.

- **Fight Game score (*FG*)**: if E is the observable representing the number of erroneous options chosen before completing the fight mini-game, FG is calculated from FG = MAX (0, 1 - (MIN (E, E_{MAX}) / E_{MAX})), where E_{MAX} is the maximum number of reasonable errors needed to beat the game.
- **Test Game score (*TG*):** In the test mini-game, each question has four potential answers, and only one of them is correct. Each answer is given an associated score. The correct answer always has a score of 0, and the remaining answers have scores that correspond to their distance relation the truth: 1 for answers that are almost right, 2 for answers that are wrong, and 3 for answers that, due to their content or formulation, are clearly listed as jokes. If I is the observable representing the accumulated score of incorrect answers after finishing the test mini-game, TG = MAX (0, 1 – I / 4), as 4 questions are posed.

We set A_{MAX} = 3 and E_{MAX} = 6. These values were agreed upon by game designers and educators in consideration of the educational and game challenges that each mini-game presents to players. However, as we wish to track raw A and E values, A_{MAX} and E_{MAX} values can always be changed a posteriori if, after running the validation process, the data suggest that more appropriate values should be used.

With these values, we can now calculate IA and FA:

- $IA = RG$, as the "redondilla" puzzle mini-game is the only one presented in the practice phase.
- $FA = FG \times 0.5 + TG \times 0.5$, as the fight and test mini-games are presented in the mastery phase, and we decided to give both equal weighting in the final score.

For all mini-games, we set the assessment threshold to 0.5, making the *IA* and *FA* thresholds 0.5 as well.

Table 1 shows possible values for *RG*, *FG*, *TG*, *IA* and *FA* used in the analysis of this experiment.

Table 1 Some illustrative values for IA and for components of FA

Initial Assessment / Redondilla game A_{MAX} = 3 $IA = RG = 1 – (MIN(A - 1, A_{MAX}) / A_{MAX})$		Final Assessment (FA) Fight game E_{MAX} = 6 $FG = 1 – (MIN(E, E_{MAX}) / E_{MAX})$		Test game TG = MAX (0, 1 - I/4)		Final assessment $0.5*FG + 0.5*TG$
Attempts (A)	RG = IA	Errors (E)	FG	Incorrect score (I)	TG	
1	**1**	*0*	*1*	*0*	*1*	**1**
2	**.66**	*1*	*.83*	*1*	*.75*	**0.8**
3	**.33**	*2*	*.66*	*2*	*.5*	**0.58**
4	**0**	*3*	*.5*	*3*	*.25*	**0.375**
5	**0**	*4*	*0.33*	*4*	*0*	**0.165**
6	**0**	*5*	*0.16*	*5*	*0*	**0.08**
7 or more	**0**	*6 or more*	*0*	*6 or more*	*0*	**0**

Initial and final assessments' formulas are presented in italic. Resuls for both assessments are presented in bold

4.4 Case study

To answer RQ2, we ran an experiment on high school students who played the serious game.

4.4.1 Experimental design

Before the sample of high school students (our target population) played the game, and as part of the validation process, we first ran a formative evaluation with graduate students [27] and with the teachers involved in the experiment. The results of this validation allowed us to address some implementation flaws and to improve the gameplay and overall learning design. For instance, two questions from the final mini-game were changed to improve their alignment with the learning goal.

After the validation was conducted, high school students played "The Foolish Lady" for 30 to 40 min on PCs under the supervision of a researcher who did not provide any assistance (only brief direction on how to start the game). We collected one gameplay per student (deployment phase). We consider a gameplay session as the set of traces (interactions with the game) generated from the first screen to the final screen of the game.

From each gameplay, we computed 3 values: *RG*, *FG* and *TG*. Students who did not complete a mini-game scored 0. From these variables, we calculated *IA* and *FA* from the formulas presented above. Using their results, we classified each student to a learning category (learner, master, non-learner or outlier) to draw conclusions on the game's effectiveness.

To gain insight into our methodology, we sought to determine whether we could answer the following case-study questions (CSQ) concerning "The Foolish Lady" serious game:

- CS1: Did the students acquire the targeted skill by the end of "The Foolish Lady" game? Given our demographic variables, were there differences between groups?
- CS2: Is "The Foolish Lady" game effective at teaching to targeted skill to our population? Given our demographic variables, were there differences between groups?

4.4.2 Participants

The experiment involved $N = 320$ high school students from 8 different schools in Madrid. Thirty-two of the gameplay sessions were corrupted or not completed due to various technical problems that arose during gameplay (power outages, Internet connection issues and computer malfunctions) and were therefore discarded.

The resulting population ($N = 288$) was 44.4% female and 55.6% male. The participants were between the ages of 12 and 16 (with a mean age of 13.70 ± 1.27) and were students at high schools in the Madrid area. Three of these schools were charter or private schools (accounting for 58% of the students), and 4 were public schools (accounting for 42% of the students). In regards to gender, age and school type characteristics, the participants are a representative sample of the student population of Madrid for this age [9, 31].

We also recorded the participants' game habits to classify each student into a player category by evaluating what types of games they played and how often. According to the instrument developed by [29], 14.9% were non-gamers (they never play any video games), 28.8% were casual gamers (they play video games casually for short periods of times), 31.6% were hard-core gamers (they frequently play games such as FPS or MMORPG) and 24.6% were well-rounded gamers (they play all types of games frequently). A more detailed explanation of each category is presented in [29].

5 Results

In this section, we present the results of the learning outcomes analysis of the deployment phase, i.e., the results of the experiment on the high school students.

5.1 Game completion

Figure 6 shows the number of players who completed each phase of "The Foolish Lady": all 288 players started the game and also completed the strategy phase; 281 completed the "redondilla" puzzle mini-game; 246 completed the fight mini-game; and 231 completed the test mini-game. The largest drop in player participation (35) occurs between the "redondilla" puzzle and fight mini-game stages.

In summary, 80.21% of the players finished the game at least once.

5.2 Learning outcomes

To determine whether the students acquired the targeted skill level by the end of the game, we calculated the values of *RG*, *FG* and *TG* and therefore *FA* and *IA*. In total, 196 players (68.05% of the total population and 84.84% of the players who completed the game) scored higher than 0.5 (adequacy threshold set for the game during the design phase) in both *FA* and *IA*.

The second part of Case Study Question 1 (CS1) led us to calculate *FA* and *IA* across the different demographic groups: gender (M/F), age (12 to 16) and gaming habits (4 clusters).

Fig. 6 Number of players who accomplished each phase of "The Foolish Lady" game

To explore whether there were statistically significant differences within each group, we first determined whether the different groups (e.g., males vs. females for division by gender) had a different starting point score. In other words, we needed to determine, for instance, whether males had a statistically significant different *IA* score than females. Regardless of whether such differences existed, we needed to adjust the scores of each group according to their *IA* values before carrying out the analysis.

We therefore performed a one-way analysis of variance (ANOVA) over *IA* to find initial differences across groups. As is shown in Table 2, *IA* showed statistically significant differences for each group, which were especially significant in the case of gender and game habits. Thus, for all of the groups studied, we needed to adjust the initial values through an analysis of covariance (ANCOVA) rather than using the analysis of variance (ANOVA) method, which is generally recommended for similar initial values.

The ANCOVA allowed us to evaluate differences in *FA* scores (dependent variable) across groups (independent variables) by overriding differences in *IA* (that is, by using *IA* as a covariate). Before conducting the ANCOVA analysis, we had to perform standard preliminary checks to confirm that there was no violation of assumptions of normality, linearity, variance homogeneity and regression slope homogeneity [35].

Table 3 shows the ANCOVA results for the 3 independent variables, which present statistically significant differences ($p < 0.05$) among groups by age and game habits but not by gender. This suggests that the dependent variable (*FA*) values differ statistically by player age and gaming habits.

The first ANCOVA [between-subjects factor: age (12 to 16); covariate: IA scores] reveals main effects for age F(4288) =7.28, $p < 0.01$ and a moderate $\eta_p^2 = .094$. According to Table 4, which shows the adjusted means once the effect of *IA* is omitted, the 16-year-old players scored moderately (according to η_p^2) higher (FA = .766) than their younger classmates: the 12- and 13-year-olds presented the lowest adjusted means (FA = .508).

The second ANCOVA [between-subjects factor: gender (male, female); covariate: IA scores] shows no main effects for gender F(1288) = .62, $p = .43$, $\eta_p^2 = .002$. Thus, we can argue that *FA* does not depend on player gender.

A third ANCOVA [between-subjects factor: game-habits (4 clusters); covariate: IA scores] reveals main effects for game habits F(3, 288) =2.880, $p = .036$. In this case, the effect of gaming habits ($\eta_p^2 = .030$) was lower than the effect of age. However, the game worked better for well-rounded players and worse for non-gamers.

5.3 Serious game effectiveness

Figure 7 shows the total number of players grouped by learning category. Most players are masters followed by learners. The number of outliers is higher than that of non-learners.

Table 2 ANOVA results on IA showing significant differences among the three groups

Independent variable	One-way ANOVAs on IA			
	N	*df*	*F*	*p*
Age	288	4	2.5	.031
Gender	288	1	18.41	<.005
Game Habits	288	3	12.10	<.005

Table 3 Test scores and ANCOVA results by age, gender and gaming profile

Independent variable	ANCOVAs on FA				
	N	*df*	*F*	*p*	*Partial* η^2
Age	288	4	7.28	.000*	.094
Gender	288	1	.62	.43	.002
Game Habits	288	3	2.88	.036*	.030

*$p < 0.05$

Figures 8 and 9 show the players grouped by learning category and segmented by age and gaming habits. In all groups, the number of masters exceeds that of the other categories, and especially for the 14-year-old group. For all of the groups, the number of outliers is greater than the number of non-learners, except for the group of students aged 16.

6 Discussion

In this section, we first present our answers to the case-study questions and then further elaborate on the methodological research questions.

CS1: Did the students acquire the targeted skill by the end of "The Foolish Lady" game? Given our demographic variables, were there differences between groups?

Yes; 80.2% of the students completed the game, which required, by design, a basic understanding of the principles of the learning goal. Regarding the students' final scores, the ANCOVA analysis (dependent variable: FA; covariate: IA) reveals significant differences when data are segmented by age and gaming habits.

Table 4 FA adjusted means by age, gender and gaming profile

Ind. Variable	Values	ANCOVA		
		N	*Adj. Mean**	*Std. Err.*
Age	12	69	.508	.038
	13	50	.508	.044
	14	96	.708	.032
	15	43	.631	.048
	16	30	.766	.057
Gender	Female	128	.603	.029
	Male	160	.633	.026
Game Habits	Casual	83	.559	.035
	Non-gamer	43	.554	.049
	Well-rounded gamer	71	.680	.038
	Hardcore	91	.659	.034

*Adjusted mean using practice phase scores as covariates (ia = .6146)

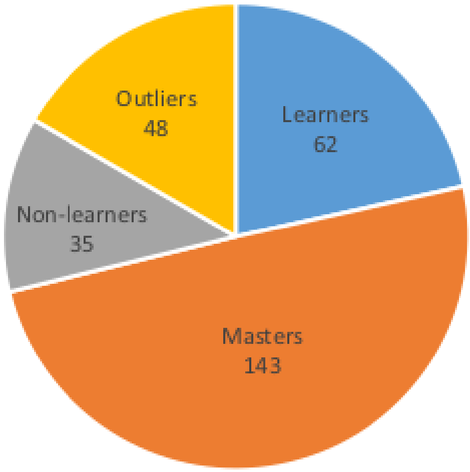


Fig. 7 Categorization of players by assessment category

As is shown in a previous section, by age, students aged 12 and 13 obtained the lowest values (Adj. Mean = .508), and students aged 16 obtained the highest values (Adj. Mean = .766). This seems natural: older students found the game easier.

In terms of gaming habits, the well-rounded gamers generated the best results (Adj. Mean = .680) followed closely by hard-core gamers (Adj. Mean = .659). These two types of players are used to playing games with complex mechanics. "The Foolish Lady" is an adventure game with fairly simple mechanics, and so these players' expertise likely helped them complete the game more effectively. At the other end of the scale, we have non-gamers (Adj. Mean = .554), which supports our hypothesis.

CS2: Is "The Foolish Lady" game effective at teaching to targeted skill to our population? Given our demographic variables, were there differences between groups?

No; this is not because the players did not learn, but rather because according to the results, most of them were categorized as "Masters", i.e., many of them already knew most of the educational content. This could mean that the game was too easy for most of the players. However, we think that an additional problem in the game's design prevented us from capturing a more accurate IA (and, consequently, a more accurate learning profile): as we wanted to keep the game short—it had to be completed over a standard 40-min session—the practice phase was made deliberately shorter than the mastery phase. This forced us to limit the

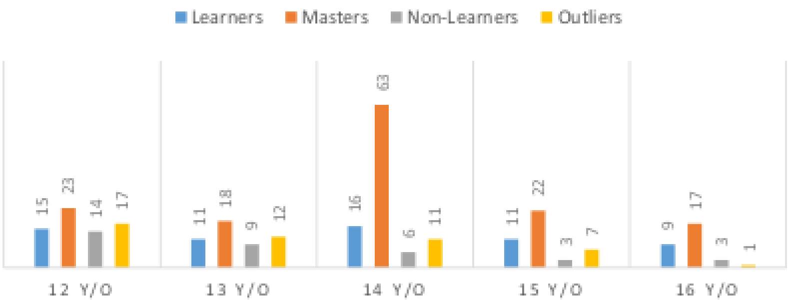


Fig. 8 Distribution of players across assessment categories segmented by age

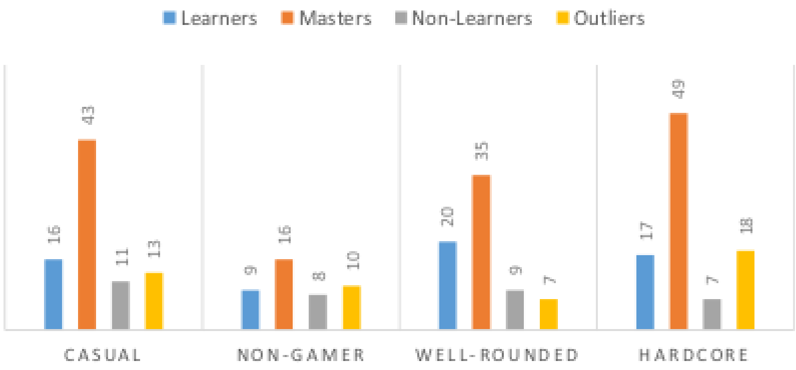


Fig. 9 Distribution of players across assessment categories segmented by game-habits

practice phase to a single comparatively easy mini-game, which proved insufficient as a full measure of initial knowledge. This flaw went unnoticed during the validation process, as the players initiated tested were domain experts, and it seemed natural to classify them as masters. This oversight shows why serious games must be validated with a sample from the target population and not only by domain experts.

When segmenting groups by age and game habits, there is no particular group for which the game was more effective.

These results do not imply that the game has no value as an educational tool. Students playing this game enjoyed other benefits, such as a measurable increase their motivation to attend the theatre, as demonstrated in [28].

RQ1. What are the implications of using our game-design pattern during the design and implementation of a serious game?

Our use of this methodology forced us to define a clear learning goal from the start and to continue to use it throughout the game development process.

In cooperation with education experts, we clearly defined which role each of the mini-games played in meeting the educational goal. The proposed game-design pattern provided guidance when defining mini-game difficulty, weight and placement measures. The mini-games were implemented in such a way that interactions and events involved in their resolution were clearly identified; and mapping from these events to assessments was also determined early on.

We also integrated a tracker into the game engine to capture all relevant interactions. This approach is commonly used in the gaming industry for analytics-related tasks, although its difficulty varies depending on the chosen game engine. In our case, we used an open source engine where all required events and interactions were generated in a handful of locations within the code. This made integrating the tracker a relatively simple task.

We also needed a service to collect all data sent by the tracker. Ours consisted of a REST back-end system that processed HTTP requests describing events, a database for storing these traces, and some Python scripts to query the database. Although we used a customized solution, the serious game could be integrated with any other VLE. This raises new interesting questions regarding the sharing of data between such systems and a serious game, which however fall outside of the scope of this paper.

RQ2. What results in regards to learning outcomes and effectiveness levels can be obtained from a serious game developed and analysed through this methodology?

The identification of relevant educational observables during game design significantly simplified the task of measuring learning outcomes and game effectiveness. These results helped us answer several interesting questions from the case study.

By default, our methodology converts serious games into assessment tools: it relies on clear assessment locations that are associated with both learning design and goals and which are then combined to infer learning outcomes. However, using our design pattern, we can also determine whether students actually learn from playing a game, which is key to assessing a game's effectiveness for a particular population group.

From our case study, we conclude that the initial assessment results are higher than what was expected. The number of outliers generated denotes a design flaw in the practice phase. Such findings, which are based on actual data, are very useful: after altering the game design, the result would be an improved game for the next round of players.

Finally, when student demographic data are available, we can use statistical analyses to identify and characterize those groups enjoying better learning outcomes or game effects. This can help narrow down and characterize the ideal target population of a game. It can also guide changes to adapt the game to other audiences.

7 Conclusions

In this paper, we present a means of structuring the design and assessment of serious games at two levels: inferring learning outcomes and assessing the effectiveness of serious games as educational tools. We think that this will help systematize serious game development and improve several of the methodologies identified in our literature review: our methodology is fully integrated with the production cycle of a serious game (from design to deployment) and involves a non-disruptive assessment as an alternative to questionnaires, the most common assessment method used for serious games. The method poses extra requirements during game development (a tracker in the game engine and a server to collect data), but with the help of today's big data technologies, this is now an affordable task.

We tested our methodology by developing a serious game entitled "The Foolish Lady", which was played by 320 students. The methodology clearly guided the design process and later our analysis of the game's effectiveness. While "The Foolish Lady" proved to be an effective assessment tool (i.e., we were able to assign a mark to each student), it was unable to fully capture the initial knowledge level of the students studied.

One of the conclusions of the experiment is that the design of the practice phase is key to the implementation of an effective serious game. However, balance in the practice phase can be difficult to maintain: the designer wants the player to advance flawlessly while identifying their mistakes to obtain an accurate assessment of their initial knowledge level. In addition, the implementation result derived from the case study (i.e., a tracker and a basic server infrastructure for receiving and analysing traces) is currently being used in the RAGE European Project [21] as an important part of the infrastructure for assessing games.

Although our case study is focused on a serious game designed to deliver knowledge and to teach several skills, we think that the methodology could be applied to any serious game whose goal can be measured in a quantitative way. For instance, a serious game designed to help diabetics control their blood glucose levels could use players' real blood glucose levels (e.g., reading them through a device connected to the game) to determine whether the goal was achieved rather than relying on the use of puzzles or mini-games.

In summary, we conclude that the methodology presented in this paper provides a richer and easily understandable assessment analysis method for serious games. We make one major point that once a game starts sending observable events, everything is automated and all of the assessments are based on how learners interact with the game rather than through the use of traditional out-of-game questionnaires. Additionally, the assessment model is adaptable to researchers' needs, as it is not hardwired to game signals: the way each dependent variable (*FA* and *IA*) is calculated can be changed "a posteriori", allowing constants used in assessment model to be updated when required (for instance, A_{MAX} and E_{MAX} in our case study). Additionally, results obtained via this methodology could complement formal experiments in measuring serious game effectiveness, which remains as an issue to address [2].

We believe that this methodology opens up avenues for future research. In this paper, we limited the students' assessments to 3 particular points in the interest of clarity (the 3 mini-games). In the future, we plan to enrich our game design pattern with more observables for both phases. Additional data will provide us with more information on student progression patterns while affording researchers greater insight into the evolution of the learning process. We plan to go one step further by analysing other gameplay data (such as the time spent on each phase) that may shed light on why some players struggle in certain parts of the game. We also wish to further explore the transformation from game observables into assessment scores by identifying and addressing common patterns of different game mechanics.

Finally, the integration of serious games that follow our proposed methodology within VLEs also raises interesting questions. What standards should be used for such communications? What visualizations should be provided to different stakeholders? Addressing such integration methods will constitute an important step towards realizing the full potential of combining serious games with learning analytics.

Acknowledgements This research study was partially financed by the Ministry of Education, Culture and Sport of Spain through its FPU Programme (grant FPU12/04310), by the Regional Government of Madrid (eMadrid S2013/ICE-2715), by the Complutense University of Madrid (GR3/14-921340), by the Ministry of Education (TIN2013-46149-C2-1-R), by the RIURE Network (CYTED 513RT0471) and by the European Commission (RAGE H2020-ICT-2014-1-644187, BEACONING H2020-ICT-2015-687676).

References

1. All A, Nuñez Castellar EP, Van Looy J (2015) Towards a conceptual framework for assessing the effectiveness of digital game-based learning. Comput Educ 88:29–37. doi:10.1016/j.compedu.2015.04.012
2. All A, Nuñez Castellar EP, Van Looy J (2016) Assessing the effectiveness of digital game-based learning: best practices. Comput Educ 92–93:90–103. doi:10.1016/j.compedu.2015.10.007
3. Annetta LA (2010) The "I's" have it: a framework for serious educational game design. Rev Gen Psychol 14:105–112. doi:10.1037/a0018985
4. Arnab S, Lim T, Carvalho MB et al (2015) Mapping learning and game mechanics for serious games analysis. Br J Educ Technol 46:391–411. doi:10.1111/bjet.12113
5. Calderón A, Ruiz M (2015) A systematic literature review on serious games evaluation: an application to software project management. Comput Educ 87:396–422. doi:10.1016/j.compedu.2015.07.011
6. Carvalho MB, Bellotti F, Berta R et al (2015) An activity theory-based model for serious games analysis and conceptual design. Comput Educ 87:166–181. doi:10.1016/j.compedu.2015.03.023
7. Chatti MA, Dyckhoff AL, Schroeder U, Thüs H (2012) A reference model for learning analytics. Int J Technol Enhanc Learn 4:318–331. doi:10.1504/IJTEL.2012.051815

8. Chen J (2007) Flow in games (and everything else). Commun ACM 50:31. doi:10.1145/1232743.1232769
9. Comunidad de Madrid (2011) Datos y Cifras de la Educación http://www.madrid.org/cs/Satellite?blobcol=urldata&blobheader=application/pdf&blobheadername1=Content-Disposition&blobheadervalue1=filename=DATOS+Y+CIFRAS+2010_2011.pdf&blobkey=id&blobtable=MungoBlobs&blobwhere=1271936872331&ssbinary=true. Accessed 12 Dec 2016
10. Connolly TM, Boyle EA, MacArthur E et al (2012) A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 59:661–686. doi:10.1016/j.compedu.2012.03.004
11. Denis G, Jouvelot P (2005) Motivation-driven educational game design. In: Proc. 2005 ACM SIGCHI Int. Conf. Adv. Comput. Entertain. Technol. - ACE '05. ACM Press, New York, pp 462–465
12. Dickey MD (2006a) Game design and learning: a conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. Educ Technol Res Dev 55:253–273. doi:10.1007/s11423-006-9004-7
13. Dickey MD (2006b) Game design narrative for learning: appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. Educ Technol Res Dev 54:245–263. doi:10.1007/s11423-006-8806-y
14. Dondlinger M (2007) Educational video game design: a review of the literature. J Appl Educ Technol 4:21–31. doi:10.1108/10748120410540463
15. Dudzinski M, Greenhill D, Kayyali R et al. (2013) The Design and evaluation of a multiplayer serious game for pharmacy students. In: Proc. 7th Eur. Conf. Games Based Learn, vols 1 2. pp 140–148
16. Elias T (2011) Learning analytics : definitions, processes and potential. Learning 23. http://learninganalytics.net/LearningAnalyticsDefinitionsProcess
17. El-Nasr MS, Drachen A, Canossa A (2013) Game analytics: maximizing the value of player data. doi:10.1007/978-1-4471-4769-5
18. Ferguson R (2012) The state of learning analytics in 2012: a review and future challenges. Tech Rep KMI-12-01. doi:10.1504/IJTEL.2012.051816
19. Fuchs LS, Fuchs D (1986) Effects of systematic formative evaluation: a meta-analysis. Except Child 53:199–208. doi:10.1177/001440298605300301
20. Hauge JB, Berta R, Fiucci G et al. (2014) Implications of learning analytics for serious game design. In: Proc. 14th Int. Conf. Adv. Learn. Technol. IEEE, pp 230–232
21. Hollins P, Westera W, Manero B (2015) Amplifying applied game development and uptake.
22. Kiili K (2005) Digital game-based learning: towards an experiential gaming model. Internet High Educ 8:13–24. doi:10.1016/j.iheduc.2004.12.001
23. Kiili K, Ketamo H (2007) Exploring the learning mechanism in educational games. In: proc. Int. Conf. Inf. Technol. Interfaces, ITI, pp 357–362
24. Lee SJ, Liu Y, Popovic Z (2014) Learning individual behavior in an educational game : a data-driven approach. In: Proc. 7th Int. Conf. Educ. Data Min. pp 114–121
25. Liu G, Fu L, Rode AV, Craig VSJ (2011) Water droplet motion control on Superhydrophobic surfaces: exploiting the Wenzel-to-Cassie transition. Langmuir. doi:10.1021/la104669k
26. Loh CS, Sheng Y, Ifenthaler D (2015) Serious games analytics: theoretical framework. In: Serious games anal. Springer International Publishing, Cham, pp 3–29
27. Manero B, Fernández-Vara C, Fernández-Manjón B (2013) E-learning a escena: De La Dama Boba a Juego Serio. Vaep Rita 1:51–58
28. Manero B, Torrente J, Serrano Á et al (2015a) Can educational video games increase high school students' interest in theatre? Comput Educ 87:182–191. doi:10.1016/j.compedu.2015.06.006
29. Manero B, Torrente J, Fernández-Vara C, Fernández-Manjón B (2015b) Gaming preferences and habits, gender and age on educational videogames effectiveness: an exploratory study. IEEE Trans Learn Technol (in press)
30. Marne B, Wisdom J, Huynh-Kim-Bang B, Labat J-M (2012) The six facets of serious game design: a methodology enhanced by our design pattern library. In: twenty-first Century Learn. Twenty-First Century Ski. pp 208–221
31. Ministerio de Educación (2008) Escolarización y población http://www.mecd.gob.es/dctm/ievaluacion/indicadores/2011-e1.2.pdf?documentId=0901e72b810b4d41. Accessed 12 Dec 2016
32. Moreno-Ger P, Burgos D, Martínez-Ortiz I et al (2008) Educational game design for online education. Comput Hum Behav 24:2530–2540. doi:10.1016/j.chb.2008.03.012
33. Nutt C, Hayashida K (2012) The structure of fun: learning from Super Mario 3D Land's director. In: Gamasutra. http://www.gamasutra.com/view/feature/168460/the_structure_of_fun_learning_.php?page=4. Accessed 13 Jan 2016
34. Owen VE, Ramirez D, Salmon A, Halverson R (2014) Capturing learner trajectories in educational games through ADAGE (Assessment Data Aggregator for Game Environments): a click-stream data framework for assessment of learning in play. Am Educ Res Assoc Annu Meet 1–7
35. Pallant J (2013) SPSS survival manual: a step by step guide to data analysis using IBM SPSS. Open Univ Pr

36. Santhosh S, Vaden M (2013) Telemetry and analytics best practices and lessons learned. In: Game Anal. Maximizing Value Play, Data, pp 85–109
37. Serrano Á, Marchiori EJ, del Blanco Á et al. (2012) A framework to improve evaluation in educational games. In: IEEE Glob. Eng. Educ. Conf. IEEE, pp 1–8
38. Serrano-Laguna Á, Torrente J, Moreno-Ger P, Manjón BF (2012) Tracing a little for big improvements: application of learning analytics and videogames for student assessment. In: Procedia Comput. Sci. pp 203–209
39. Squire K (2003) Video games in education. Int J Intell Simul Gaming 2:49–62
40. Vargas JA, García-Mundo L, Genero M, Piattini M (2014) A systematic mapping study on serious game quality. In: Proc. 18th Int. Conf. Eval. Assess. Softw. Eng. EASE 2014. pp 1–10
41. Ye F (2014) Validity, reliability, and concordance of the Duolingo English Test.

Ángel Serrano-Laguna finished his Master Degree in Computer Science at the UCM in 2012. Currently he is working as a full-time researcher in the e-UCM research group, while working on his PhD. His research focuses in the design and implementation of Serious Games and the creation of tools to ease its introduction in the classroom, as well as the study of Learning Analytics techniques to assess those games.

Borja Manero has a PhD in Computer Science by University Complutense of Madrid. He received a Bachelor in Physics (major in Computer Science) from the UCM (1999) and in Dramatic Arts (2007). Currently he is a Lecturer in the Department of Software Engineering and Artificial Intelligence of the Complutense University in Madrid, Spain. His research interests focus on educational games and their use on arts (particularly in theater, cinema, or dance).

Manuel Freire has a PhD in Computer Science from the Universidad Autónoma de Madrid (UAM). His research interests are centered on Information Visualization, Human-Computer Interaction, Online Learning and Serious Games. In 2008, he enjoyed a a postdoctoral Fulbright grant at the University of Maryland's HCIL. In 2010, he became a member of the e-UCM research group at the Universidad Complutense de Madrid (UCM), where he is an Associate Professor since 2013.

Baltasar Fernández-Manjón is a Professor in the Department of Software Engineering and Artificial Intelligence (DISIA) at the Complutense University of Madrid (UCM). He leads the Complutense e-learning research group eUCM (www.e-ucm.es) and is the holder of an honorary Telefonica-Complutense Chair in Serious Games. In 2010–2011 he has been Visiting Associate Professor at Harvard University and Visiting Scientist at LCS-MGH. He is IEEE Senior Member and associate editor of IEEE Transaction on Learning Technologies. He is involved in different EU projects (e.g. H2020-RAGE, FP7 NoE GALA) and other projects about technology transfer with industry. His main research interests are educational uses of (serious) games, learning analytics, e-learning technologies, and application of educational standards topics. He has published more than 150 papers on that topics.

6.6. Learning Analytics and Educational Games: Lessons Learned from Practical Experience

6.6.1. Cita completa

Ángel Serrano-Laguna, Javier Torrente, Borja Manero, Ángel del Blanco, Blanca Borro-Escribano, Iván Martínez, Manuel Freire, Baltasar Fernández-Manjón (2013), ***Learning Analytics and Educational Games: Lessons Learned from Practical Experience,*** International Conference on Games and Learning Alliance: GALA 2013, pp 16-28. DOI: 10.1007/978-3-319-12157-4_2

6.6.2. Resumen original de la contribución

Learning Analytics (LA) is an emerging discipline focused on obtaining information by analyzing students' interactions with on-line educational contents. Data is usually collected from online activities such as forums or virtualized courses hosted on Learning Management Systems (e.g. Moodle). Educational games are emerging as a popular type of e-learning content and their high interactivity makes them potential sources for relevant educational user data. However, it is still uncertain how to deploy and combine these two incipient technologies, as multiple challenges remain unresolved. This paper reports on our practical experience using LA to improve assessment of experimental research on educational game-based instruction. In the last year, we conducted four experiments evaluating game-based instruction under different conditions (using three adventure games and a puzzle game respectively) in 13 educational institutions including schools, universities and vocational training organizations. A LA system was used to track interaction around six hundred students. In these experiences, we encountered several problems in each of the steps of the process, from issues related to the design of the experiment and the game to different technical and practical problems, derived from the very diverse conditions of the facilities and policies of each institution (computer laboratories, computers hardware, software installed, irregular Internet access), which hinders the data collecting process (e.g. the system had to deal with high latency Internet connections and backup plans had to be devised for collecting data when no Internet access was available). We present the lessons learned and propose guidelines from a technical and practical perspective for the design of experimental research using both LA systems and educational games.

Learning Analytics and Educational Games: Lessons Learned from Practical Experience

Ángel Serrano-Lagunaa(✉), Javier Torrentea, Borja Maneroa, Ángel del Blancoa, Blanca Borro-Escribanoa, Iván Martínez-Ortiza, Manuel Freirea, and Baltasar Fernández-Manjón

Complutense University of Madrid, C/Profesor José García Santesmases, 9, 28040 Madrid, Spain
aserrano@e-ucm.es

Abstract. Learning Analytics (LA) is an emerging discipline focused on obtaining information by analyzing students' interactions with on-line educational contents. Data is usually collected from online activities such as forums or virtualized courses hosted on Learning Management Systems (e.g. Moodle). Educational games are emerging as a popular type of e-learning content and their high interactivity makes them potential sources for relevant educational user data. However, it is still uncertain how to deploy and combine these two incipient technologies, as multiple challenges remain unresolved. This paper reports on our practical experience using LA to improve assessment of experimental research on educational game-based instruction. In the last year, we conducted four experiments evaluating game-based instruction under different conditions (using three adventure games and a puzzle game respectively) in 13 educational institutions including schools, universities and vocational training organizations. A LA system was used to track interaction around six hundred students. In these experiences, we encountered several problems in each of the steps of the process, from issues related to the design of the experiment and the game to different technical and practical problems, derived from the very diverse conditions of the facilities and policies of each institution (computer laboratories, computers hardware, software installed, irregular Internet access), which hinders the data collecting process (e.g. the system had to deal with high latency Internet connections and backup plans had to be devised for collecting data when no Internet access was available). We present the lessons learned and propose guidelines from a technical and practical perspective for the design of experimental research using both LA systems and educational games.

1 Introduction

This paper aims to be a practical guide for researchers interested in conducting experimental research on educational gaming who might also want to explore the opportunities that embedding a learning analytics system provides to improve the research. Building on our experience, we identify potential issues that can arise while conducting this kind of research, and also provide recommendations to tackle them. Some of the issues discussed are technical, related to the implementation and

A. De Gloria (Ed.): GALA 2013, LNCS 8605, pp. 16–28, 2014.
DOI: 10.1007/978-3-319-12157-4_2

deployment of the technology, while others are rather logistical, operational and related to the design of the experiments.

The field of learning analytics (LA) refers to the collection, analysis and visualization of large amounts of data related to educational processes. In its heart, LA aims to harness the power of big data and data-mining techniques to improve the assessment of the learning processes. LA also can create new opportunities for adaptive and personalized learning, which has remained an unfulfilled promise of the e-learning field for years. As an incipient new era of online learning, promoted by Massive Online Open Courses (MOOCs) comes upon our shoulders, LA could find the perfect conditions to make an impact in education in the next few years [1].

As any big data system, LA requires gathering a wealth of data from different sources. It usually uses data generated from tracking students' interactions with the online platform used to support learning (e.g. a Learning Management System like Moodle or Sakai). For example created posts, pages and resources accessed or time spent on each piece of content. Research in the LA field is currently exploring how to leverage other data sources to improve the effectiveness of the paradigm. One of the proposals is to use educational games, which are another pushing educational technology. Although educational games are still far from reaching massive adoption due to unresolved limitations [2–4], their effectiveness to improve learning is accepted among the academic community [5], as attributed benefits to educational games like increased student motivation, improved engagement and better knowledge acquisition have been recently backed up with experimental research [6–9]. Serious games can pose an advantage to feed LA systems as they are inherently highly interactive pieces of software which can produce massive user data.

Achieving a successful synergy between serious games and learning analytics poses significant challenges that are added to the difficulties of designing and conducting experimental research in education. In the next sections we will present an overview of our recent research and discuss the issues found and the solutions we came up with.

2 Overview of Games Used in the Experiences

During the last year, we have conducted 4 experimental research studies using educational games. In round numbers, 4 games have been deployed at 13 educational institutions including universities, high schools and vocational training institutions, involving more than 600 students.

The four games were independently developed and cover different knowledge areas. Three of them were conversational point-and-click games developed with the eAdventure platform [10, 11], while the fourth was programmed from scratch.

Experiments carried out with these games followed a similar process. In a starting phase, the games were designed. The games were implemented and then internally tested. Once the games were polished and ready for distribution, they were deployed to conduct experiments in several educational institutions. During the experiments, students completed a pre-test to measure initial levels of motivation and knowledge on the subject. Then they played the game while the LA system collected data gathered from their interaction with the game. At the end of the session, students completed a post-test. Differences between pre and post tests were measured to estimate effect of instruction.

Next subsections briefly introduce each experiment and the games used.

2.1 The Big Party

The Big Party [12] is an eAdventure game (Fig. 1) that aims to teach persons with different levels of cognitive disabilities life management skills related to personal hygiene, safety, social interaction and transportation. The goal of the game is to reach a party organized by the player's company for all the staff. To complete the game, players must find their way through different common situations of daily life until they reach the venue of the party. This game was played by 19 players of different ages and with different levels of cognitive disabilities.

Fig. 1. Screenshots of the game "The Big Party".

2.2 Lost in Space <XML>

This puzzle, level-based game (Fig. 2) was designed to teach XML syntax to students with different programming backgrounds. In the game, players control a spaceship they must lead to a target point by writing little XML documents with instructions (e.g. move spaceship two units forward, rotate 90º, shoot, etc.). 89 students from computer science and social science studies played this game distributed in two different settings.

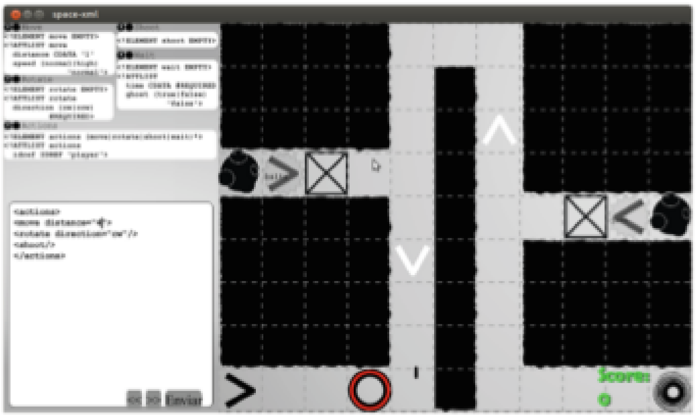


Fig. 2. Screenshot from the game "Lost in Space <XML>".

2.3 La Dama Boba

The game La Dama Boba [13] is an eAdventure game (Fig. 3) that was designed to motivate youngsters on classic theater plays. This goal is achieved making avatars interesting to players, by incorporating different theater techniques within their personalities and including the elements that can be used as audiovisual contents, such as music, scenery and dressing.

The game is based on the equally named comedy by the Spanish playwright Lope de Vega, wrote in 1613. The player becomes Laurencio, the main male character of the comedy, and has to live the story through his eyes. As the plot unveils, several minigames about grammar and literature are also introduced.

This game has been deployed in 9 different high schools in Madrid where 370 students aged from 11 to 15 played it.

2.4 Donations

Donations is an eAdventure game-like simulation (Fig. 4) developed in collaboration with the Spanish National Transplant Organization (ONT). This game simulates the process that the ONT staff follow for deceased donation management. These steps are (1) organ and donor evaluation, (2) the organ distribution and (3) the organ transportation. The game was developed to facilitate instruction of new personnel within the organization as well as to support transferring this successful workflow to other transplant coordinators. The game has been evaluated by 150 students in three training courses.

Fig. 3. Screenshots of the game La Dama Boba. Character design inspired in customs by Agatha Ruiz de la Prada for an adapted version of the play.

3 Designing and Planning the Experiment

Our experiments are oriented to evaluate the impact that using videogames has on the learning process. Mainly, we try to measure the educational gain (knowledge acquired by the students as a consequence of the intervention) and changes in the motivation towards the subject of interest (e.g. literature or computer programming).

In this section we provide a set of considerations we take when designing the experiments.

3.1 Instruments for Measuring Variation in Knowledge and Motivation

We use a twofold approach to measure variations of motivation and educational gain. First we use questionnaires to perform an external analysis of the learning process. This type of analysis does not consider how or what makes learning happen; instead, it just focuses on the outputs. Second, GLEANER, a LA system [14], is used to perform an internal evaluation of the process, gathering insight on what aspects of the game promote learning.

Several instances of the questionnaire are used along the experiment to have multiple values to compare with. At least the questionnaire is used twice: a pre-test that

Fig. 4. Screenshots of the game Donations. Main central office of the ONT and action of organ evaluation.

is administered just before students start playing and a post-test administered after the intervention, but it can be used more times for longer or longitudinal experiments. The results of different measures of the questionnaires are compared (e.g. pre-test Vs post-test) to determine the effect of the game.

A thorough design of the questionnaire is strongly recommended. The questions included in the test must tackle exactly what it is meant to be measured in the study. We combine several strategies to iteratively design and refine our questionnaires. First, we identify the aspects that we want to capture, usually distributed in three or more sections: demographics (e.g. age, gender, video game preferences, etc.), knowledge on the subject, and motivation towards the subject and/or evaluation of the experience (for post-tests only). Second, we write a set of candidate questions exploring different wordings and evaluation scales for each of the aspects. We also design several questions oriented to measure aspects that are difficult to capture because they cannot be observed, like the knowledge or motivation. This allows us to build subscales that aggregate all the questions related to the same aspect to generate a variable that is more powerful for statistical analysis. Third, the candidate questions are distributed for iterative review among an internal team of experts in educational games and also experts in the knowledge area. Finally, one or more testing sessions are organized where the instruments and questionnaires are piloted with a group of selected students.

As a preferred format, we use 7-point Likert questions for subjective aspects (e.g. motivation) as this allows sometimes applying parametric statistical methods for analysis. For objective aspects (e.g. knowledge) we tend to use multiple-option questions to simplify evaluation.

3.2 Group Comparison

It is essential to compare the effect of game-based instruction to traditional instruction to allow drawing valid conclusions at the end of the study [2], especially if the same questionnaire is used at different points of the experiment, which introduces bias in the variation measured.

For that purpose, we distribute students in at least two groups: experimental and control. Students in the experimental group attend the session where the game is played. Students in the control group attend a traditional instruction session, which is usually a lecture driven by an experienced teacher. In some cases we use a third group where the best possible instruction is delivered to the student, which is usually driven by experts in the field using (perhaps) special equipment. All groups must use the same pre and post tests to measure variation. This allows discussing the effectiveness of the game comparing to bottom and top lines.

3.3 Duration of the Experiment and Logistics

Our experiments are designed to require an intervention of maximum 50 min, which is the typical schedule slot in education. All instruments and also the game must be designed to respect this constrain.

Longer exposure to instruction is desirable, but it is hard to achieve in many educational organizations as logistics and organizational costs for an experiment like these, which are already high, increase. It must be noted that in many cases the experiment will require taking up most of the computer resources of the institution for a while. Computer clusters in schools usually have less than one computer per student (around 20 computers per cluster) to minimize costs, which implies that students usually work in pairs. If the game under evaluation is designed for individual use, it will be needed to use two computer clusters if possible, or to split students in the experimental group in two turns, requiring to have access to the cluster for a longer period.

Some institutions will not meet these requirements. In that case, a different kind of experiment must be planned, following a qualitative approach. Students can play the game in groups, and the sessions would be recorded for later analysis. These data can be combined with interviews or debriefing sessions.

3.4 Getting Support from the Educational Institution

It is essential to get support from the educational organization where the experiment will take place to ensure that researchers are given access to facilities, students and personnel. Also researchers must make sure that the institution understands all the requirements for the experiment. In our experience, we have observed that sometimes researchers' and institutions' interests conflict. On the one hand, researchers want to have the facilities and support required to conduct an experiment that is as controlled and well designed as possible to ensure that the conclusions obtained are valid. On the other hand, the institution has usually tight schedules and organizing an experiment like these has a considerable impact for them. Having strong institutional support will help researchers finding a balance between both perspectives that benefits the goals of the research.

A possible strategy to get better institutional support is to make the institution and its staff understand that getting involved in the experiment adds value to all the stakeholders involved. Principals and administrative staff can be enticed by the idea of participating in a pilot study that improves innovation in the institution. Motivated teachers or early adopters may be interested in having new software available to improve their instruction for free.

3.5 Maximize Control to Reduce Error Rates

Researchers should supervise all the important parts of the experience and ensure they are in control. People out of the research can unintentionally bias the results of the experiment. In this regard, understanding teachers' perspectives before the experiment can help to foresee potential problems and develop mitigation plans. In our experience we discovered that some teachers are worried about how their students will perform compared to other institutions. In this sense, they felt their work as teachers being under examination. These teachers unintentionally tended to give students more clues to solve the tests, introducing bias.

If the experience requires randomized allocation of students in groups that will receive different instruction, we suggest not leaving the entire responsibility to the teacher, or at least researchers should supervise the process. Teacher used to consider playing the game as a reward for the best students, and this could bias data.

3.6 Ethical and Legal Issues

Ethical and legal issues must always be considered, especially when the target audience include minors. Privacy of data collected must be ensured. For that reason, we use anonymization techniques and anonymous questionnaires. To pair questionnaires and results within the game, each student is given a unique code that the write on paper tests and also introduce in the games to identify data sent to the LA system.

All students should receive the same treatment and have access to the same content. For that reason, we usually let students in control group play the games once the experiment is complete and all data have been collected. Also students in the experimental group are given traditional instruction after the experience.

3.7 Workflow

For the reader's convenience, in this section we provide a summary of the workflow with all the activities conducted in these experiments.

- Make an initial design of experiment, materials and game. This will help know the requirements of your experiment.
- Start recruitment. Approach institutions that may be interested in participating. This process may be long and tedious, until the desired number of participants is ensured for the project. Contact institutions, get their attention and explain your needs. Make sure to explain that they have the opportunity of participating in a research project for free (teachers are used to companies offering different services for the schools, and tend to refuse them immediately).
- Meeting the teachers that will participate. Normally, in a school, there are several teachers for a single subject. The meeting must be planned to involve the maximum number of teachers. Explain the requirements and schedule activities according to the needs of the institution. Also get information about the facilities of the institution, special requirements, etc.
- Refine or adapt the design of the experiment and materials to fit the needs or limitations of each institution, if necessary.
- Install the software and revise the computers in advance. We strongly recommend testing the computers and to install the game in advance as not all the institutions have maintenance plans for the computers. All software and computers must be installed and ready before the experiment takes place.
- The day of the experiment. We suggest a minimum of two researchers to carry out the experience. They should arrive ten minutes in advance. One researcher will lead the experimental group and check that the computer facilities are ready. The other researcher will lead the control group and supervise all activities conducted by teachers.
- Post-mortem activities. After the experiment, link data collected from questionnaires and data obtained from the LA system and analyze the results. Then, conduct debriefing sessions and interviews with teachers involved to get insight on their impressions about the experience and present the results obtained.

4 Designing and Implementing the Game

In this section we provide an overview of technical issues that must be considered when designing and implementing the game to ensure that (1) data collected from LA will be of interest and (2) the game is as easy to deploy as possible to meet the varying conditions and requirements of the institutions.

4.1 Designing the Game for Learning Analytics

Educational game design is a complex activity and it is not our intention to delve into all the considerations that make a game both educational and entertaining. For further reading on this matter we would recommend some of the articles that can be found in the literature [15, 16]. Instead, in this section, we elaborate on the requirements that a game design must meet to allow effective use of LA.

In short, the game must interlace mechanics oriented to facilitate building new knowledge with mechanics oriented to assess the new knowledge acquired. This behavior is frequently present in video games where players acquire new skills throughout a game level and at the end they have to apply these skills in a new way to defeat a final "boss".

We did not consider this issue when we designed The Big Party. The game was linear and students did not have to apply any new piece of knowledge acquired. We learned from this experience, and the following games were designed having two phases: one to let the student build new knowledge, one to assess. The Table 1 shows a summary with each game and how the two phases were distributed.

Table 1. Description of the learning and assessment phases

Game	Educating phase	Assessment phase
Lost in Space <XML>	Phase presenting a new power-up (representing a new syntax structure to be learned)	Phases not presenting any new power-up (players must use structures already learned)
La Dama Boba	Several text screens explaining literature concepts, mini-games about grammar and spelling	Direct questions at the end of the game, mini-games repetition. Final assessment is shown to the player
Donations	First phase, where players are guided through all the donation process, allowing them to commit mistakes and correcting them	Second phase, where players go through all the donation process, with no help and no mistakes allowed

4.2 Game Technical Requirements

The game must be implemented using a technology that can be easily deployed using the computers of the institution(s) where the experiment will be held. Having a detailed description of the settings can be used to make an informed decision on the appropriate technology. However, it is not always possible to get this information. Our recommendation is to choose a technology that is lightweight, flexible and easy to deploy across platforms (Windows, Linux and Mac). In our games, we use two technologies that meet this requirement: Java, and the Web Browser (HTML plus JavaScript).

Educational institutions may also have different hardware configurations. Especial attention must be placed on the dimension of the displays and audio support. While some institutions may have modern high definition displays, others may support low resolution applications. Similarly, some institutions may provide headsets while in others sound may be disabled. For these reasons, we use an 800 × 600 resolution which works well in low resolution screens and which also looks nice in larger screens. We also design the games so that they can be played with or without sound.

4.3 Tracking System Requirements

This section presents some of the technical considerations to bear in mind when deploying a videogame that communicates with a LA system.

The game must establish communication with a remote server and also univocally identify the player in a way that allows pairing data collected through all the instruments of the experiment. If a good Internet connection is available identification can be done on the server side. However, in many educational institutions Internet access is not reliable or bandwidth is insufficient. To deal with these situations, an offline alternative must be designed. For example, the game can store the information locally and the identification can be done using a code provided by the researchers. After the experiment researchers can collect the files storing tracking information and upload it manually to the LA system.

The server is in charge of receiving and storing the traces. To facilitate the analysis of the data, the server must store not only the traces but also the user id, game id, session id (continuous period of time where the user plays the game), user's group and learning activity/experience. These metadata allows the researcher to contextualize the statistical analysis of the data giving other variables to the analysis.

Depending on the size of the experiment and the data collected, the data traffic generated can be significant. Hence, it is needed to prepare the server to manage quite high workloads efficiently. The server must minimize the response time, and must use a storage system specifically optimized for writing loads of data. For that purpose, it is handy to use a NoSQL database (e.g. Apache Cassandra, Apache HBase, MongoDB, etc.) because they are particularly optimized for writing throughput. However it is possible to use a traditional relational database system, usually by mixing clustering and sharding techniques. Note that the storage system used to store the traces received by the tracking system may not be the same used to analyze the data, that is, the received data can transformed to another representation (e.g. graph, relational, etc.) and stored into another database that can facilitate the analysis of the data.

Finally, in order to minimize the response time and to avoid the collapse of the tracking system, it is usually needed to throttle the requests to the tracking system (limit the number of request per second) and limit the amount of data to be sent per request. This request throttling requires a close collaboration between the client and the server, on one hand the server can control (even dynamically) the number of allowed request per second and notifies with an especial error that the game has reached this limit or that the request cannot be currently processed, and on the other hand the client must be aware of this error conditions and resend the request after a certain amount of time.

It is desirable to have a tool to monitor the client-server communication during the experiment. This allows adapting the settings to deal with any technical issue that may prevent correct data collection. Ideally, this tool gives real time feedback on the communication and summary statistics. In our case, we had a service that showed how many students were playing and its identifiers, allowing us to visually confirm that the number of students playing were the same as the number of users that server was receiving traces for.

To help to interpret part of the collected data, it can be helpful to write notes about observations made during the experience, associated to concrete student identifiers.

For example, we observed that most high school girls carefully read screen texts, while most boys don't. Something that we confirmed after analyzing traces with the time-stamps to know the time expend in those screens that include instructions for the game.

5 Final Remarks

Probably the most important recommendation, according to our experience, is to follow a careful process to design and execute the experiment. All the aspects are important - from the design of the game to the wording of the questionnaires. To achieve success attention must be placed on all the details. Also it is important to be prepared for unexpected situations and to have a backup plan for carrying out the experiment.

References

1. Johnson, L., Adams, S., Cummins, M.: NMC Horizon Report: 2012 K-12 Edition, p. 44. The New Media Consortium, Austin (2012)
2. Hays, R.T.: The effectiveness of instructional games: a literature review and discussion. Naval Air Warfare Center, Orlando, FL (2005)
3. Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., Ludgate, H.: NMC Horizon Report: 2013 Higher Education Edition, Austin, Texas, USA (2013)
4. Pivec, P.: Game-based Learning or Game-based Teaching? Becta (2009)
5. Hwang, G.-J., Wu, P.-H.: Advancements and trends in digital game-based learning research: a review of publications in selected journals from 2001 to 2010. Br. J. Educ. Technol. **43**(1), E6–E10 (2012)
6. Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., Boyle, J.M.: A systematic literature review of empirical evidence on computer games and serious games. Comput. Educ. **59**(2), 661–686 (2012)
7. Sadler, T.D., Romine, W.L., Stuart, P.E., Merle-Johnson, D.: Game-based curricula in biology classes: differential effects among varying academic levels. J. Res. Sci. Teach. **50**, 479–499 (2013)
8. Tuzun, H., Yilmazsoylu, M., Karakus, T., Inal, Y., Kizilkaya, G.: The effects of computer games on primary school students' achievement and motivation in geography learning. Comput. Educ. **52**(1), 68–77 (2009)
9. Warren, S.J., Dondlinger, M.J., McLeod, J., Bigenho, C.: Opening the door: an evaluation of the efficacy of a problem-based learning game. Comput. Educ. **58**(1), 397–412 (2012)
10. Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J.L., Fernández-Manjón, B.: Educational game design for online education. Comput. Hum. Behav. **24**(6), 2530–2540 (2008)
11. Torrente, J., Del Blanco, Á., Marchiori, E.J., Moreno-Ger, P., Fernández-Manjón, B.: <e-Adventure>: introducing educational games in the learning process. In: 2010 Conference on IEEE Education Engineering (EDUCON), pp. 1121–1126 (2010)
12. Torrente, J., del Blanco, Á., Moreno-Ger, P., Fernández-Manjón, B.: Designing serious games for adult students with cognitive disabilities. In: Huang, T., Zeng, Z., Li, C., Leung, C.S. (eds.) ICONIP 2012, Part IV. LNCS, vol. 7666, pp. 603–610. Springer, Heidelberg (2012)

13. Manero, B., Fernández-Vara, C., Fernández-Manjón, B.: E-learning a Escena: De La Dama Boba a Juego Serio. IEEE-RITA **1**(1), 51–58 (2013)
14. Serrano, A., Marchiori, E.J., del Blanco, A., Torrente, J., Fernandez-Manjon, B.: A framework to improve evaluation in educational games. In: Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON), pp. 1–8 (2012)
15. Salen, K., Zimmerman, E.: Rules of Play: Game Design Fundamentals. MIT Press, Cambridge (2003)
16. Ak, O.: A Game Scale to Evaluate Educational Computer Games. Procedia - Social Behav. Sci. **46**, 2477–2481 (2012)

6.7. Applying standards to systematize learning analytics in serious games

6.7.1. Cita completa

Ángel Serrano-Laguna, Iván Martínez-Ortiz, Jason Haag, Damon Regan, Andy Johnson, Baltasar Fernández-Manjón (2017), ***Applying standards to systematize learning analytics in serious games***, Computer Standards and Interfaces, Vol. 50, pp. 116-123. DOI: 10.1016/j.csi.2016.09.014 [Factor de impacto JCR 2015: 1.268; Segundo cuartil (Q2) 35/106 en “Computer Science, Software Engineering”]

6.7.2. Resumen original de la contribución

Learning Analytics is an emerging field focused on analyzing learners’ interactions with educational content. One of the key open issues in learning analytics is the standardization of the data collected. This is a particularly challenging issue in serious games, which generate a diverse range of data. This paper reviews the current state of learning analytics, data standards and serious games, studying how serious games are tracking the interactions from their players and the metrics that can be distilled from them. Based on this review, we propose an interaction model that establishes a basis for applying Learning Analytics into serious games. This paper then analyzes the current standards and specifications used in the field. Finally, it presents an implementation of the model with one of the most promising specifications: Experience API (xAPI). The Experience API relies on Communities of Practice developing profiles that cover different use cases in specific domains. This paper presents the Serious Games xAPI Profile: a profile developed to align with the most common use cases in the serious games domain. The profile is applied to a case study (a demo game), which explores the technical practicalities of standardizing data acquisition in serious games. In summary, the paper presents a new interaction model to track serious games and their implementation with the xAPI specification.

Computer Standards & Interfaces 50 (2017) 116–123

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Applying standards to systematize learning analytics in serious games

Ángel Serrano-Laguna[a,*], Iván Martínez-Ortiz[a], Jason Haag[b], Damon Regan[b], Andy Johnson[b], Baltasar Fernández-Manjón[a]

[a] Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, Madrid, Spain
[b] Advanced Distributed Learning Initiative, United States

ARTICLE INFO

Keywords:
Serious games
learning analytics
standards
xAPI

ABSTRACT

Learning Analytics is an emerging field focused on analyzing learners' interactions with educational content. One of the key open issues in learning analytics is the standardization of the data collected. This is a particularly challenging issue in serious games, which generate a diverse range of data. This paper reviews the current state of learning analytics, data standards and serious games, studying how serious games are tracking the interactions from their players and the metrics that can be distilled from them. Based on this review, we propose an interaction model that establishes a basis for applying Learning Analytics into serious games. This paper then analyzes the current standards and specifications used in the field. Finally, it presents an implementation of the model with one of the most promising specifications: Experience API (xAPI). The Experience API relies on Communities of Practice developing profiles that cover different use cases in specific domains. This paper presents the Serious Games xAPI Profile: a profile developed to align with the most common use cases in the serious games domain. The profile is applied to a case study (a demo game), which explores the technical practicalities of standardizing data acquisition in serious games. In summary, the paper presents a new interaction model to track serious games and their implementation with the xAPI specification.

1. Introduction

A serious game is a video game designed with a purpose other than pure entertainment [1]. Serious games have been proven to be effective educational tools in many domains, such as mathematics, physics, engineering, medicine, economics, history and literature [2–5]. The methods used to measure their effectiveness vary from study to study (some standard guidelines are starting to arise [6]). However, a large number of serious games research studies primarily depend on data from surveys and questionnaires [7].

Meanwhile, data-driven approaches that rely on collecting and analyzing data from learners' on-line activity are a current trend in the e-learning community. Disciplines such as Learning Analytics (LA) [8] and Educational Data Mining (EDM) [9] are studying the way learners perform online activities within Virtual Learning Environments (VLE). Their main goal is to better understand educational processes to find ways to improve them and assure an accurate assessment of the student. LA applications vary from identifying students at risk of failing a course [10] to recommending additional educational materials for those students who might need them [11]. As the number of LA applications increases, there is a growing interest in which educational standards can be used to share and exploit data, easing the collaboration between LA tools and VLEs.

The interactive nature of serious games makes them a good source of LA data. Tracking the learner's interaction within a serious game and storing it provides multiple benefits for all the stakeholders involved in the learning process. For instance, teachers could follow a student's progression while the student plays, and could take action on any identified learning problem in real-time. In addition, researchers can now harvest and store activity data in a centralized location, and can more easily conduct a deeper analysis for understanding consistent student behaviors and performance in serious games.

Current research shows that questionnaires, which are outside the game context, are the most common method to collect data in formal experiments with serious games [7]. This heavily contrasts with the practices in non-formal environments. For instance, commercial videogames have been relying on Game Analytics (GA) to learn from their users for years [12]. GA researchers use questionnaires to assess game mechanics or gameplay [13], but their main source of data is the non-disruptive tracking embedded in their games (usually called telemetry [14]). They can track all types of interactions with different purposes: from predicting revenue to measuring engagement.

Serious games can greatly benefit from GA techniques (and non-disruptive tracking) to improve their analysis and, through the use of standards, to ease their integration and increase their usefulness inside VLEs. Using GA techniques poses some challenges, though. The video

* Corresponding author.

http://dx.doi.org/10.1016/j.csi.2016.09.014
Received 20 June 2016; Received in revised form 29 September 2016; Accepted 29 September 2016
Available online 30 September 2016
0920-5489/

game industry is very protective with their GA practices and usually relies on proprietary systems. Consequently, there are no standardized formats to represent players' interactions. Another limitation is that GA and LA goals differ: while GA aims to increase engagement (trying to maximize the time a player stays in the "flow" state [15]) and monetization, LA seeks to analyze and measure players' learning outcomes [16].

Some serious games track their learners' interactions, but use custom formats (Section 2 below shows a detailed review). These custom formats hinder support for serious games in educational tools (particularly VLEs) and the use of general tools to process and analyze the data. This is natural due to the immaturity of the field, however, there are enough case studies that identify common interactions tracked by serious games to start defining tracking models, which eventually can be standardized. This paper aggregates these case studies to infer an interaction model candidate for standardization, joining other proposals to track LA observables in a standard way [17]. The standardization will also help the creation of supporting infrastructure and decrease the cost of applying LA (e.g., the cost of integrating the game in a VLE) [18].

In this paper, we perform a review to detect the common interactions tracked in serious games found in the literature. We use that review to infer an aggregated interaction model. Then, we analyze the current LA standards and their suitability to represent the model. Finally, we present a reference implementation using the xAPI specification [21], along with a case study.

The rest of the paper is structured as follows: Section 2 presents a review of the interactions tracked in formal experiments with serious games; Section 3 shows the interaction model inferred from findings in Section 2; Section 4 analyzes current LA standards; Section 5 presents the xAPI implementation of the interaction model along with a case study; and Section 6 discusses the results, limitations, and future work.

2. Interactions tracked in serious games

In this paper we build upon a previous review by Calderon and Ruiz on serious games and evaluation [7]. They identified a total of 120 papers using key search terms such as "evaluation / validation / assessment" combined with "serious games / simulation games". We were looking for serious games that contained any type of tracking for assessment, and this review was additionally valuable for finding relevant serious games, and it was recent enough to provide a quite complete list. We reviewed the 120 papers and found 14 serious games tracking the players' in-game interactions. We analyzed them to infer the types of interactions most commonly tracked, and present the results below.

In general, video games use two interaction strategies: a) event-based, where the game logs pre-specified events when they occur; and b) state-based, where the game repeatedly sends game state at a specific frequency [22]. Each of the 14 serious games in our review opted for the events-based strategy. Most of the events include at least two attributes: a timestamp, representing the moment the event is generated, and a user id, identifying the player that originates the event. This basic data enable researchers to obtain metrics such as number of players, the number of times players accessed the game, or time played [23–26].

The serious games reviewed track the completion (binary value, yes or no) or the level of completion achieved (percentage) by the players. Some serious games just track whether the game was fully completed, and others have a more fine-grained level of detail, tracking completion in each of the levels within the game [23,27,28]. This type of interaction relies on the notion that a player completing the game or parts of the game is also progressing towards a learning goal. Common metrics extracted from these events are the quantity of completed levels/scenarios/scenes and the time it took to complete each one.

Many serious games also track the in-game choices performed by players in a given context. These choices most commonly involve questions with multiple answers [23,24,29]. These questions can be either presented directly to the player during the game or integrated in a dialog with a non-playable character (NPC). Some serious games also track general choices where the player must select an action among several options [25,26]. In many cases, when a player makes a choice in a serious game, he would need to apply specific knowledge to make the right decision. This feature makes the choice interesting for future assessment. Common metrics obtained from these interactions are: the time spent to make decisions and the rightness/wrongness ratio when selections can be scored positively or negatively (questions with a correct answer).

The serious games that were reviewed often rely on meaningful measurable variables to calculate players' learning outcomes [25,27,28]. The most common variables are scores, number of in-game deaths and kills or coins collected. All these variables are linked to the player's performance in the game, and can reveal the level of success in the learning goals involved. For instance, a final high score in a serious game can indicate good performance towards the game's learning goals. The most common metrics extracted from this interaction are the game's final high score value and the ratios that can be obtained combining several of them collected from different game attempts.

Finally, we need to highlight that several serious games tracked events that were specific to those games. For instance, Hauge [29] collected chat logs from their multiplayer game, Qudrat-Ullah [26] tracked the number of times a player asked for in-game help, and Buttussi [30] and Cowley [31] collected biometric information using several external devices. This leads us to conclude that, although we can identify common events and interactions, the use of game-specific interactions is also necessary to assess user performance in serious games.

3. Interaction model

The previous section covers how learner's interactions are tracked in serious games. This section presents an interaction model, derived from the previous analysis, to represent the most common interactions found in the analyzed serious games.

The proposed model uses events to represent the players' individual interactions within the game, since event-based tracking is the most widespread strategy among serious games. The vast majority of interactions are characterized by two attributes: an action and a target that receives the action. Sometimes, an additional value is necessary to quantify the result of the interaction (e.g., in the interaction "achieved 50% of Level 1", the action is "achieved", the target is "Level 1" and the value is "50%"). Finally, all interactions identify who generated them and when.

Thus, we define an interaction event with the following attributes: 1) a timestamp, representing the instant the event was generated in the game; 2) a user id, identifying the player that generated the event; 3) an action, representing the type of interaction performed by the player; 4) a target, representing a game element that is the objective of the player's action; and 5) an optional value, representing the parameters of the action. All attributes are required except value, which will only appear for those actions that need parameters to quantify their results.

Below, we present all the types of targets (i.e., the objective of the player's action) identified in Section 2 along with their related actions.

3.1. Completables

The target type *completable* deals with the player's level of progress in a serious game. A completable is something a player can start, progress on and complete within a serious game (even several times). It is a unit of progress inside the game, and can have different scopes. Some examples of completables are: game, game session, level, quest, world, stage, and race.

The action *start* marks when a player begins the fulfillment of a completable. No value is associated to this action. The action *progress* updates the total advance of a player in a completable. The associated value must be a decimal number, between 0.0 and 1.0, where 0.0 means the player has not made any progress towards the completable and 1.0 means the player fully satisfied the completable. The action *complete* marks when a player finishes a completable. No value is associated with this action. Considering these actions, we could calculate the following metrics per player and per completable: level of completeness, times completed, and time to complete.

3.2. Alternatives

The target type *alternative* deals with each of the in-game decisions a player performs during a gameplay. An *alternative* is a set of options among which the player has to choose at a given point in the game. The player can only choose one option, and some of the options can be unavailable (locked). Some examples of alternatives are: questions, menus, paths, and dialog tree (in a conversation with a non-playable character).

The action *select* marks when a player selects an option in an alternative. The associated value must be a string identifying the selected option. The action *unlock* marks when a player unlocks an option in an alternative. The associated value must be a string identifying the unlocked option. Considering these actions, we could calculate the following metrics per player and alternative: options selected and options unlocked.

3.3. Meaningful variables

The target type *meaningful variable* deals with each of the values that represent something meaningful in gameplay (e.g., a score). A *meaningful variable* is a value inside the game world with a special significance. Meaningful variables are usually numeric, but they could also be other data types, for instance, text, binary values (true or false), or simple structures (e.g., positions). The player can set a variable's value. Some examples of meaningful variables are: score, currency (e.g., coins, rings, and money), health, and the player's position.

The action *set* marks when the game sets a particular value in a meaningful variable after some player interaction. The associated value must be the new value. Considering these actions, we could calculate the following metrics per player and variable: final value and value's evolution.

3.4. Custom interactions

There are serious games and educational scenarios that will benefit from tracking very specific player interactions with great detail, for example, to facilitate a manual subjective analysis of the interaction.

If the serious game needs to track some event that is not able to be represented by the events defined above, the model can always be extended with new types of targets and a set of associated actions. For instance, when a serious game tracks chat logs [29], we could create a type of target called "chat message" (at the same level as completables, alternatives, and meaningful variables) with the action *send* to represent when a player sends a text message in a chat, where the value would be the message content.

However, if too much use is made of these extensions, then the result may be a loss of the automatic or semi-automatic processes that the standard is intended to support, hence requiring the extension of analysis tools or manually processing the interactions. The model presented in this section is a first step to cover the interactions most commonly found in serious games, settling down a common semantic for both serious games developers and analytics tool providers and allowing them to innovate and work without restrictive dependencies. The model can be extended with custom interactions that can be generalized and shared across many serious games (e.g., events for concrete game mechanics), along with the automatic processes to analyze them.

4. Learning Analytics standards

In the previous section, we have defined the targets and actions presented in our interaction model for serious games. Now we need a real notation to represent the model. As we discussed in the introductory section, we can take advantage of standardization efforts currently underway in the field of LA in order to represent serious games analytics. Therefore, we need to find an appropriate standard to represent our interaction model.

LA specifications and standards can deal with many types of data in educational contexts. Several authors [32] classify data handled by learning analytics into two categories: static (data that barely changes over time) and dynamic (data that is updated more often). These data can proceed from several sources: people, resources, services, learning activities, objectives, and assessments [33]. In our particular situation, we are looking for a standard that deals with dynamic data (interactions) inside a learning activity (serious game).

Some standards in the dynamic category focus on capturing achievements derived from users' interactions. For instance, Mozilla Open Badge Initiative (OBI) [34] is oriented towards the issuing of badges that represents the knowledge or skills gained by individual learners that can be checked by third parties, while IMS Basic Outcomes [35] offers the possibility to set a grade for the activity.

Within this category, SCORM deserves a special mention, since it is one of the most widespread standards used in the deployment and communication of education resources [36], and has been used in the past to track results of serious games [37]. SCORM supports communication of completion status, success status, score, and progress. It also provides some extra fields to report events from the educational activity: comments, in free text forms; interactions, to track questionnaires and other educational items, and objectives, to track sub-goals with score, status, and completion variables.

The main limitation of SCORM and the other standards we have mentioned, is that their principal design goal is to capture results. Some of them allow for interactions during the activity to be captured, but many tools and VLEs do not make these data available for analysis, hence the educational content behaves as a "black box" [37]: we know the results of the activity but we have no information about how the activity was actually carried out. New approaches in LA are advocating for the "white box" model [37], in which educational content provides more granularity for user interaction data, allowing more insights into educational results.

Efforts being made towards this model include standards dedicated to tracking activities and interactions inside educational resources (dynamic data about learning activities), usually in the form of a log. For instance, PSLC DataShop Tutor Message Format (DataShop) [38] is used to log activity in tutoring applications; Contextualized Attention Metadata (CAM) [39] is used to log user interactions in learning environments.

Another relevant specification for this task is Activity Streams [40]. This specification is able to represent sequences of actions performed by users in a specific context. Each action is represented by an "activity," whose main attributes are an actor (who performs the action), a verb (what action is performed), and an object (the target of the action). A sequence of activities is a stream, and represents the actions from a set of users. This is a general purpose specification, although it was initially based on the interactions given inside social networks. Its structure is a good fit to represent our model interactions.

ADL Experience API (xAPI) and IMS Caliper are two specifications greatly inspired by Activity Streams. Both are mainly used in educational contexts, so they are an even better fit to represent our interaction model. Below, we analyze each of them separately.

4.1. IMS Caliper

Caliper Analytics is a framework developed by the IMS Global Consortium [41], whose main goal is to establish a way to capture and obtain measures from a set of learning activities.

Each learning activity has one or several associated metric profiles. A metric profile defines the information model that shapes the types of events emitted by the learning activity. It also provides a semantic for later analysis. Some metric profiles are designed to track raw user activity (e.g., page views in an eBook) and others to track user learning outcomes (e.g., the score in an assessment) [42].

The Sensor API defines the standard representation of the events in each metric profile. The event structure derives from Activity Streams and contains, among other attributes, an actor, an action, and an object. The first version of IMS Caliper [43] does not provide a way to extend the vocabulary used by events to fit other learning activities (such as serious games). Some sort of extension can be accomplished through the extensions attribute present in events, although this would not be enough as our interaction model needs to define a whole set of new vocabulary.

A new metric profile would be necessary to represent serious games as learning activities in IMS Caliper. The profile should define as actions all interactions described in Section 3, and as objects all the targets. The task is technically accomplishable but it would require direct collaboration with IMS, since, to date, they fully control the development of new metric profiles.

4.2. Experience API

Experience API (xAPI) is a specification developed by an open community lead by the Advanced Distributed Learning Initiative (ADL) [21]. The specification's objective is to define a data and communication model to track user activities within learning environments.

xAPI defines each event tracked in a learning activity as a Statement. The format also derives from Activity Streams, and the main attributes in a Statement are actor, verb (action), and object. Fig. 1 illustrates its structure.

The statements can contain additional attributes with more information about the experience: *result*, containing the outcomes of the statement; *context*, representing the learning environment; or *authority*, specifying who assures the truthfulness of the statement.

All xAPI statements are sent to a Learning Record Store (LRS), a database that holds all the statements in sequential order. The LRS can be later queried to perform statements analysis.

Unlike IMS Caliper, xAPI does not set any constraints on the vocabulary that can be used in the statements. Practitioners can create their own verbs and activity types to define domain specific vocabularies. Additionally, xAPI allows extensions to expand the specification and fulfill new or unique requirements. The definition of new vocabularies along with these extensions is denominated a "xAPI profile". These profiles are usually developed through an xAPI Community of Practice (e.g., some groups have created vocabulary for eBooks or videos) [44], but other third parties are also developing vocabularies to fit their own needs [45].

In summary, the xAPI specification is designed to represent sequences of interactions, is widely adopted in the educational community, and it allows for the creation of domain specific vocabularies to fit new types of learning activities. These features pose it as one of the best candidates to use in the representation of our interaction model for serious games. In the next section, we present how our model can be implemented using xAPI.

5. Experience API implementation

A full version of a Serious Games xAPI Profile has been developed by the RAGE project in collaboration with ADL [46]. In this section, we detail the process followed to convert the simple events structure of the proposed model presented in Section 3 to full xAPI statements.

First, we create a mapping between event fields and Statement properties as an initial approach to the conversion and present the initial vocabulary of the profile. Then, we explore the use and semantics of xAPI attributes to create richer statements, specifically, the use of *result* to log game state variables and the use of *extensions* to add additional semantics useful for analysis. Finally, we present a case study that uses this implementation and present some statement examples.

5.1. Event fields to Statement properties

Table 1 shows the mapping of schema fields to xAPI Statement properties. Below, we summarize the considerations and adjustments made to each field in the proposed implementation.

The *userId* maps to *actor*. The VLE hosting the learning activity is responsible for communicating the value of *actor* to the client (the serious game). This communication is done during the activity launch, after an authentication process.

The *action* maps to *verb*. xAPI identifies a verb with a unique internationalized resource identifier (IRI). IRIs are usually URIs that, if resolved, are intended to return the verb definition in a machine-readable format. xAPI does not define any particular verb by default (with the exception of a reserved verb, *voided*) and suggests that verbs defined by the community should be reused. If the community does not provide a verb that represents what is needed, a new verb with a new IRI should be created.

The *target* maps to *object*. xAPI also identifies objects with IRIs, but in this case the IRIs identify specific objects of the activity, and should not be reused outside its context. However, the *object* can specify a *type* that defines its class. The IRIs for this attribute follow the same guidelines as verbs' IRIs.

The *value* maps to *result*. The *result* attribute holds all the outcomes associated with the statement, and it contains a set of predefined properties: *score*, *success*, *completion*, *response*, *duration*, and *extensions*. The proposed implementation will try to map the *value* field to these attributes. In cases where the semantics of the value disallow it, it will use the *extensions* attribute, which allow the definition of custom outcomes (see Section 5.2. below).

Finally, the *timestamp* maps directly to the xAPI statement *timestamp*.

5.2. xAPI profile vocabulary

This section presents the xAPI vocabulary needed to represent the types of events presented in Section 3. Table 2 shows the mapping of actions to xAPI verbs, Table 3 shows the activity types for common targets in serious games. Note that this new vocabulary is also based on our previous work and experience applying e-learning standards to serious games [19,37].


```
{
  "actor": {
    "name": "John Doe",
    "mbox": "mailto:johndoe@example.com"
  },
  "verb": {
    "id": "http://adlnet.gov/expapi/verbs/completed",
    "display": { "en-US": "completed" }
  },
  "object": {
    "id": "http://example.com/activities/programming-course",
    "definition": {
      "name": { "en-US": "Programing course" }
    }
  }
}
```


Fig. 1. An xAPI Statement representing a learning activity. Specifically, that "John Doe completed a Programming Course".

Table 1
Mapping of interactions event fields to xAPI statement attributes.

Interaction event field	xAPI statement attribute
UserId	Actor
Action	Verb
Target	Object
Value (Optional)	Result (Optional)
Timestamp	Timestamp

This vocabulary conceptualizes the most used concepts found in serious games. Note that the goal of this xAPI vocabulary (and profile) is to establish a basis for the application of LA standards to serious games. Once that is done, the Serious Games Community of Practice can extend this vocabulary to capture interactions from more specific game mechanics. For instance, it could add the verbs *jumped, killed* or *picked* to represent typical interactions in a platform game.

5.3. Game variables in result

The interaction model presented in Section 3 (also building on our earlier work [19,20]) proposes to generate a new event with the action *set* each time a game variable is updated (e.g., the score increased, the health bar decreased, the coins was set to a value…). This design decision presents a drawback: the primary action that generated the update gets lost. That is, if the player gets a score of 500 points because he completed a level, two events are generated: one with the level completion and one with the score update. Thus, it is necessary to analyze previous events to obtain the triggering action of the score update.

xAPI enables the problem to be solved through the use of the attribute *result*. This attribute allows the ability to specify any outcome related to the statement. Thus, the variable updates can be contained by the interaction that generated them instead of being in a separate statement. So, instead of generating two statements when a player completes a level and obtains 500 points, the xAPI tracker only generates one, containing the level completion as the main interaction and the score update in the result.

5.4. Analysis' semantics with extensions

A common problem in any type of analysis based on user-generated logs is to establish the semantics of each action in the log. xAPI solves this problem for statements by giving strong semantics to the vocabulary it defines. For instance, if the semantics of the verb *completed* and the activity type *exam* are fully defined, whenever a statement declaring that "a student completed an exam" is found, its interpretation for analysis will be unequivocal.

The verbs and activities definitions are fundamental to the analysis of statements. Additionally, xAPI defines *extensions* properties to establish further semantics whenever verbs and activity types are not enough. For instance, the attribute *result* contains an *extensions* property to add additional outcomes of the statement.

Table 4 shows the result extensions defined for the Serious Games xAPI Profile. For instance, the extension *progress* measures the level of achievement during an activity, *health* determines the current level of health of the player and *position* localizes the player inside the game world (with coordinates, *x, y,* and *z).*

These extensions represent common concepts found in serious games. Along with their definitions, they establish clear semantics for their analysis (e.g., a level of health near to 0 would indicate that the player is in danger of losing a life or attempt). As with verbs and activity types, the Serious Games Community of Practice can leverage extensions to capture more specific game mechanics.

5.5. Case study: the Countrix game

This section presents *Countrix*, a serious game that implements the Serious Games xAPI profile to represent and communicate learners' interactions. The game is a case study developed to demonstrate the use of the Serious Games xAPI profile. It is fully implemented and available for download.[1] The goals of this case study are two: 1) illustrate the use of the profile with a real serious game and 2) analyze the technicalities involved in the xAPI communication.

Countrix is a timed Q&A about Geography (countries, capitals, flags and continents). The player has a fixed amount of time to answer the greatest number of questions. When the player chooses a correct answer the score increases by one. If the player fails, the time to answer questions decreases (Fig. 2). The game uses a simple game mechanic to illustrate more clearly the statements generated.

Fig. 3 shows a statement generated right after starting the game. The statement uses the verb *initialized* to specify that the activity (of the serious game type) has just started. Fig. 4 shows a statement generated after the player selects an incorrect answer. The statement uses the verb *selected*, and an activity of type *question*. The result contains the variable that changes with the interaction. In this case, the value *health* (remaining time to answer questions) decreases with the error and its value is transmitted. The game also contains an xAPI viewer that allows the player to watch the statement being generated in real time.

The *Countrix* game is connected to a Learning Analytics framework that contains an LRS [18–20]. When the player starts the game, the game requests an authorization to the LA framework to start sending xAPI statements. If the framework grants it, it responds, in JSON format, with the LRS endpoint, an actor identifying the user and an activity id. The client tracker uses the returned data to start sending xAPI statements.

This initialization and authorization process is out of the xAPI specification scope, so it was necessary to design and develop it from scratch. The rest of the communication with the LRS is fully described by the xAPI specification. In this case, the framework uses an open source LRS to store the statements. Finally, the game contains an xAPI tracker, developed and integrated in such a way that it emitted the statements at the appropriate moments.

6. Discussion, conclusions and future work

In this paper, we reviewed 14 serious games that track learners' gameplay. This review allowed us to build an interaction model that can lay the foundation for a more systematic application of learning analytics in serious games. The model includes several types of events common in serious games and proposes some associated metrics. The model can be implemented with standards or specifications designed to track users' activity in the form of sequential streams (i.e., Activity Stream and its derivatives) and can decrease the cost of implementing LA for serious games. It sets a basis to start performing analysis in serious games methodologically, using the semantics concepts it contains: completable, alternatives, and variables.

However, this is just a first step in systematizing LA in serious games. The proposed interaction model and its associated metrics cover only the general aspects of learning in a serious game. The model can be extended by adding elements that represent the nuances of concrete game mechanics, which are usually correlative to the way players learn. It can also take elements from Game Analytics, since many of their insights can help to make serious games into better (e.g., more engaging) video games. These extensions will enable new types of analysis that will help to better understand the learning process inside

[1] Download from https://play.google.com/store/apps/details?id=com.anserran.countrix (Last accessed September, 2016).

Table 2
Mapping of event actions to xAPI verbs.

Action	Verb	Definition
Start	http://adlnet.gov/expapi/verbs/initialized	Indicates the activity provider has determined that the actor successfully started an activity.
Progress	http://adlnet.gov/expapi/verbs/progressed	Indicates a value of how much an actor has advanced or moved through an activity.
Complete	http://adlnet.gov/expapi/verbs/completed	Indicates the actor finished or concluded the activity successfully.
Unlock	https://w3id.org/xapi/seriousgames/verbs/unlocked	Indicates the actor unlocked an option previously unavailable.
Select	https://w3id.org/xapi/adb/verbs/selected	Indicates the selected choices, favored options or settings of an actor in relation to an object or activity.

Table 3
Mapping of target types to xAPI activities' types.

Target	Activity Type	Definition
Serious Game	https://w3id.org/xapi/seriousgames/activities/serious-game	A game designed for a primary purpose other than pure entertainment. For instance, an educational game or a game-like simulation
Level	https://w3id.org/xapi/seriousgames/activities/level	A level of a game or of a gamified learning platform.
Mission	https://w3id.org/xapi/seriousgames/activities/mission	An accomplishable mission or challenge presented inside a gamified activity.
Question	http://adlnet.gov/expapi/activities/question	A question is typically part of an assessment and requires a response from the learner, a response that is then evaluated for correctness.
Menu	https://w3id.org/xapi/seriousgames/activities/menu	A menu with several options whose selection produces different effects.
Dialog tree	https://w3id.org/xapi/seriousgames/activities/dialog-tree	An alternative presented during a conversation with a non-playable character.

Table 4
Extensions defined for the serious game profile.

Extension Name	Extension IRI	Definition
Progress	https://w3id.org/xapi/seriousgames/extensions/serious-game	A decimal number (3 significant figures with 2 figures following the decimal point) between 0 and 1 (inclusive) to indicate the value of progress in an activity.
Currency	https://w3id.org/xapi/seriousgames/extensions/currency	A variable indicating the count of and spendable currency within a game: coins, rings, diamonds, dollars...
Health	https://w3id.org/xapi/seriousgames/extensions/health	A decimal number (3 significant figures with 2 figures following the decimal point) between 0 and 1 (inclusive) to indicate the remaining health in an activity.
Position	https://w3id.org/xapi/seriousgames/extensions/position	An object, with attributes x, y and z, indicating the position of the player in the game world.

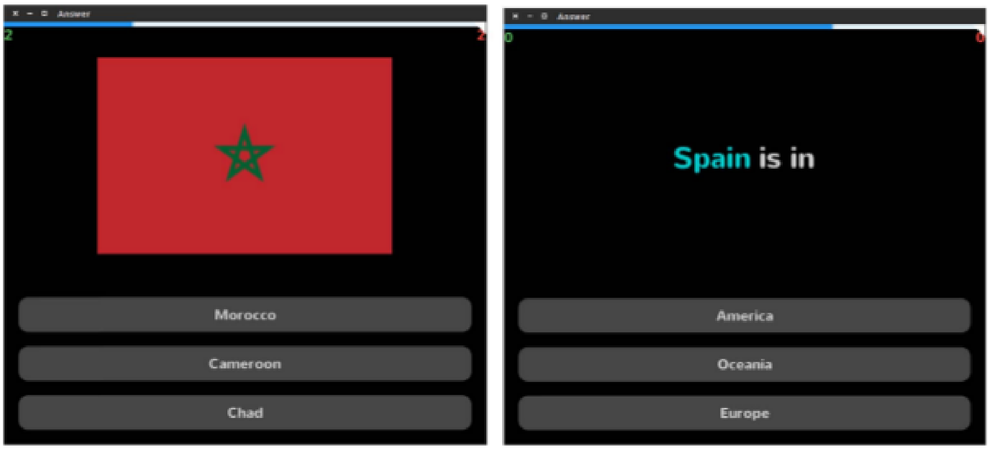


Fig. 2. Screen captures of *Countrix*. The game presents consecutive questions with 3 possible answers. The remaining time is represented by a blue bar at the top. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

```
{
  "actor": {
    "name": "John Doe",
    "mbox": "mailto:johndoe@example.com"
  },
  "verb": {
    "id": "http://adlnet.gov/expapi/verbs/initialized",
    "display": { "en-US": "initialized" }
  },
  "object": {
    "id": "http://rage.e-ucm.com/activities/Countrix",
    "definition": {
      "name": { "en-US": "Countrix Serious Game" },
      "type:" "https://w3id.org/xapi/seriousgames/activities/serious-game"
    }
  }
}
```


Fig. 3. xAPI statement generated every time the player launches the game.


```
{
  "actor": {
    "name": "John Doe",
    "mbox": "mailto:johndoe@example.com"
  },
  "verb": {
    "id": "https://w3id.org/xapi/adb/verbs/selected",
    "display": { "en-US": "selected" }
  },
  "object": {
    "id": "http://rage.e-ucm.com/activities/Countrix/questions/Capital_of_Spain",
    "definition": {
      "type:" "http://adlnet.gov/expapi/activities/question"
    }
  },
  "result": {
    "response": "Lisbon",
    "success": false,
    "extensions": {
      "https://w3id.org/xapi/seriousgames/extensions/health": 0.34
    }
  }
}
```


Fig. 4. xAPI statement generated after the player selected an incorrect answer.

serious games.

We wanted to explore the technical practicalities of using the proposed model, so this paper also presents a full implementation using the xAPI specification. The specification's flexibility enabled the full interaction model to be described, creating a new set of vocabulary along with an xAPI serious games profile. Using xAPI enables the integration with compatible VLEs, as well as the use of some basic report tools developed by the xAPI community. The *Countrix* serious game was presented as an example of use of the Serious Games xAPI profile to improve the understanding of the approach.

We did encounter some limitations. The xAPI specification is fully based on the notion of a self-contained statement: each statement emitted from an educational activity should have enough data to make sense on its own. These extra data include attributes such as *context* (the activity where the statement was generated) or *authority* (who assures the statement is valid and true), which are redundant across a sequence of statements belonging to the same gameplay session. Although these properties are optional, in some cases highly dense statements can create bandwidth problems for highly interactive activities that produce a substantial number of statements, as in the case of video games. This can be tackled by simplifying the content of the statement, for instance, sending only the minimum required attributes of the statement (namely, identifiers for actors, verbs and objects and values for results), and leaving to the receiving LRS (database) the task of filling the missing attributes. Another approach would be to send and store events in a compressed non-xAPI format, and enable the compressed data store to serve xAPI statements through a converter in response to queries. In addition to verbosity, the JSON data format might not be the most efficient representation regarding both bandwidth and CPU cycles metrics, an issue that is particularly relevant for mobile devices which are commonly used gaming devices. We are exploring the use of an xAPI gateway that can receive a more efficient representation (binary) and then transform it into the standard (and full) JSON statement representation.

The great flexibility of xAPI also presents some risks. Although the specification sets a common ground for the data format, it leaves to the communities of practice the development and agreement on common vocabularies for specific domains. The basic analysis tools available for xAPI are too general to analyze the semantics behind these specific vocabularies, so the development of analysis tools targeting their specific semantics is also necessary to extract the full potential of the specification. Thus, the success of an xAPI profile as an open standard requires a minimum number of adopters from different areas (business, education, research...) working together on tools they can develop and share. This can be a complex task in an area that is evolving so fast.

This paper is an exploration of the current issues in the application of LA standards to serious games.

We consider that the interaction model that we have presented, together with the Serious Games xAPI Profile, can establish some basic principles and open new research paths for serious game analysis.

Acknowledgments

We would like to thank Ana Rus Cano Moreno for her detailed review and suggestions on the paper. This research has been partially financed by the Ministry of Education, Culture and Sport of Spain, through their FPU Programme [Grant FPU12/04310], by the Regional Government of Madrid [eMadrid S2013/ICE-2715], by the Ministry of Education [TIN2013-46149-C2-1-R], by the RIURE Network [CYTED 513RT0471] and by the European Commission [RAGE H2020-ICT-2014-1-644187, BEACONING H2020-ICT-2015–687676].

References

[1] D. Djaouti, J. Alvarez, J.-P. Jessel, Classifying serious games: The G/P/S model, Handb. Res. Improv. Learn. Motiv. through Educ. games Multidiscip. approaches, 2011, no. 2005, pp. 118–136.
[2] M.J. Mayo, Video games: a route to large-scale stem education? Science (80-.)., vol. 323, no. 5910, pp. 79–82, Jan. 2009.
[3] B. Manero, C. Fernández-Vara, B. Fernández-Manjón, E-learning takes the stage: from La Dama Boba to a serious game, Rev. Iberoam. Tecnol. Aprendiz 8 (4) (2013) 197–204.
[4] K.H. Evans, W. Daines, J. Tsui, M. Strehlow, P. Maggio, L. Shieh, Septris: a novel, mobile, online, simulation game that improves sepsis recognition and management, Acad. Med. 90 (2) (2015) 180–184.
[5] K. Squire, M. Barnett, J.M. Grant, T. Higginbotham, Electromagnetism supercharged!: learning physics with digital simulation games, in: Proceedings of International Conference on Learning Sciences, 2004, pp. 513–520.
[6] A. All, E.P. Nuñez Castellar, J. Van Looy, Assessing the effectiveness of digital game-based learning: Best practices, Comput. Educ. 92–93 (2016) 90–103.
[7] A. Calderón, M. Ruiz, A systematic literature review on serious games evaluation: an application to software project management, Comput. Educ. 87 (2015) 396–422.
[8] T. Elias, Learning analytics: definitions, processes and potential, Learning (2011) 23.
[9] C. Romero, S. Ventura, Educational data mining: a review of the state of the art, IEEE 40 (6) (2010) 601–618.
[10] K. E. Arnold, M.D. Pistilli, Course signals at Purdue, in: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge – LAK '12, 2012, p. 267.
[11] V. Caputi, A. Garrido, Student-oriented planning of e-learning contents for Moodle, J. Netw. Comput. Appl. 53 (2015) 115–127.
[12] M.S. El-Nasr, A. Drachen, A. Canossa, Game Analytics: Maximizing the Value of Player Data, Springer, 2013 http://www.springer.com/us/book/9781447147688.
[13] B. Weedon, Game Metrics Through Questionnaires, in: Game Analytics, London: Springer London, 2013, pp. 515–537.
[14] S. Santhosh, M. Vaden, Telemetry and analytics best practices and lessons learned, Game Anal.: Maximizing Value Play. Data (2013) 85–109.
[15] J. Chen, Flow in games (and everything else), Commun. ACM 50 (4) (2007) 31.
[16] C.S. Loh, Y. Sheng, D. Ifenthaler, Serious Games Analytics: Theoretical Framework, in: Serious Games Analytics, Cham, Springer International Publishing, 2015, pp. 3–29.
[17] A. Cooper, Learning Analytics Interoperability – a survey of current literature and candidate standards, Cetis, [Online]. Available: (http://blogs.cetis.org.uk/adam/wp-content/uploads/sites/23/2013/05/learning-analytics-interoperability-v1p1.pdf), 2013.
[18] P. Hollins, W. Westera, B. Manero, Amplifying Applied Game Development and Uptake, Oct. 2015.
[19] Á. Serrano, E.J. Marchiori, Á. del Blanco, J. Torrente, B. Fernández-Manjón, A framework to improve evaluation in educational games, in: Proceedings of the IEEE Global Engineering Education Conference (EDUCON'12), 2012, pp. 1–8
[20] Á. Serrano-Laguna, J. Torrente, P. Moreno-Ger, B.F. Manjón, Tracing a little for big improvements: application of learning analytics and videogames for student assessment, Procedia Comput. Sci. 15 (2012) 203–209.

[21] ADL Initiative, xAPI Specification, [Online]. Available: ⟨https://github.com/adlnet/xAPI-Spec/blob/master/xAPI.md⟩, 2016 (accessed: 01.04.16).
[22] A. Drachen, M. Seif El-Nasr, A. Canossa, Game analytics – the basics, in: Game Analytics, London: Springer London, 2013, pp. 13–40.
[23] N. Adamo-Villani, T. Haley-Hermiz, R. Cutler, Using a serious game approach to teach 'operator precedence' to introductory programming students, in: Proceedings of the International Conference on Information Visualisation, 2013, pp. 523–526.
[24] M. Dudzinski, D. Greenhill, R. Kayyali, S. Nabhani, N. Philip, H. Caton, S. Ishtiaq, F. Gatsinzi, The design and evaluation of a multiplayer serious game for pharmacy students, in: Proceedings of the 7th European Conference on Games Based Learning, Vols 1 and 2, 2013, vol. 1, pp. 140–148.
[25] Á. López-Martínez, S. Santiago-Ramajo, A. Caracuel, C. Valls-Serrano, M. J. Hornos, M.J. Rodríguez-Fortiz, Game of gifts purchase: computer-based training of executive functions for the elderly, in: Proceedings of the 2011 IEEE 1st International Conference on Serious Games and Applications for Health, SeGAH 2011, 2011.
[26] H. Qudrat-Ullah, Perceptions of the Effectiveness of System Dynamics-Based Interactive Learning Environments: an Empirical Study, Comput. Educ. 55 (2010) 1277–1286.
[27] S. Berkovsky, M. Coombe, J. Freyne, D. Bhandari, N. Baghaei, Physical Activity Motivating Games : Virtual Rewards for Real Activity, CHI 2010 Games Play, pp. 243–252, 2010.
[28] N.A. Zin, W.S. Yue, A. Jaafar, Digital game-based learning (DGBL) model and development methodology for teaching history, WSEAS Trans. Comput vol. 8 (2) (2009) 322–333.
[29] J.B. Hauge, J.C.K.H. Riedel, Evaluation of Simulation Games for Teaching Engineering and Manufacturing, Procedia Comput. Sci 15 (2012) 210–220.
[30] F. Buttussi, L. Chittaro, Smarter phones for healthier lifestyles: an adaptive fitness game, IEEE Pervasive Comput. 9 (4) (2010) 51–57.
[31] B. Cowley, M. Fantato, C. Jennett, M. Ruskov, N. Ravaja, Learning when serious: psychophysiological evaluation of a technology- enhanced learning game, Educ. Technol. Soc. 17 (1) (2014) 3–16.
[32] D. Griffiths, T. Hoel, A. Cooper, Learning Analytics Interoperability: Requirements, Specifications and Adoption, 2016
[33] A. Cooper, Learning Analytics Interoperability – The Big Picture In Brief, Learn. Anal. Community Exch., 2014.
[34] Mozilla, Open Badges for Lifelong Learning, White Pap., 2011, pp. 1–14, .
[35] J. Fontenla, R. Pérez, M. Caeiro, Using IMS basic LTI to integrate games in LMSs Lessons from GaméTel, in: Proceedings of the 2011 IEEE Global Engineering Education Conference, EDUCON 2011, 2011, pp. 299–306.
[36] W. Capone, SCORM Sharable Content Object Reference Model, Info, 2004.
[37] Á. del Blanco, E.J. Marchiori, J. Torrente, I. Martínez-Ortiz, B. Fernández-Manjón, Using e-learning standards in educational video games, Comput. Stand. Interfaces 36 (1) (2013) 178–187.
[38] K.R. Koedinger, R.S.J. Baker, K. Cunningham, A. Skogsholm, A Data Repository for the EDM community: The PSLC DataShop, Handb. Educ. Data Min. (2010) 43–55.
[39] H.-C. Schmitz, M. Wolpers, U. Kirschenmann, K. Niemann, Contextualized attention metadata, Hum. Atten. Digit. Environ. 13 (2007) 1082–9873.
[40] J. Snell, M. Atkins, W. Norris, C. Messina, M. Wilkinson, R. Dolin, JSON Activity Streams 1.0, Act. Streams Work. Gr., 2011, vol. 22, no. 8, p. 2013.
[41] IMS Global Consortium, Caliper Analytics, [Online]. Available: ⟨http://www.imsglobal.org/activity/caliperram⟩, 2016 (accessed 01.04.16).
[42] IMS Global Consortium, IMS Caliper Best Practices Guide, [Online]. Available: ⟨http://www.imsglobal.org/caliper/caliperv1p0/ims-caliper-analytics-best-practice-guide⟩, 2016 (accessed 01.04.16).
[43] IMS Global Consortium, IMS Caliper Analytics Implementation Guide, 2015. .
[44] ADL Initiative, xAPI Community of Practice, [Online]. Available: ⟨https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/xapi-community-of-practice-cop/⟩, 2016 (accessed 01.04.16).
[45] M. Manso Vázquez, M. Caeiro Rodríguez, and M. Llamas Nistal, Development of a xAPI application profile for self-regulated learning requirements for capturing SRL related data, in Proceedings of IEEE Global Engineering Education Conference, EDUCON, 2015, vol. 2015–April, pp. 358–365.
[46] eUCM Research Group, xAPI Serious Games Profile, [Online]. Available: ⟨http://w3id.org/xapi/seriousgames⟩ 2016 (accessed: 01.06.16].

Bibliografía

ADL Initiative. (2016). xAPI Specification. Retrieved April 1, 2016, from https://github.com/adlnet/xAPI-Spec/blob/master/xAPI.md

Advanced Distributed Learning, U. S. D. O. D. (2011). *SCORM Users Guide for Programmers*. Retrieved from http://www.adlnet.gov/wp-content/uploads/2011/12/SCORM_Users_Guide_for_Programmers.pdf

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2015). Towards a conceptual framework for assessing the effectiveness of digital game-based learning. *Computers & Education*, *88*, 29–37. http://doi.org/10.1016/j.compedu.2015.04.012

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, *92–93*, 90–103. http://doi.org/10.1016/j.compedu.2015.10.007

Amory, A., Naicker, K., Vincent, J., & Adams, C. (1999). The use of computer games as an educational tool: identification of appropriate game types and game elements. *British Journal of Educational Technology*, *30*(4), 311–321. http://doi.org/10.1111/1467-8535.00121

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science (New York, N.Y.)*, *228*(4698), 456–462. http://doi.org/10.1126/science.228.4698.456

Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology*. http://doi.org/10.1037/a0018985

Arafeh, J. M., Hansen, S. S., & Nichols, A. (2010). Debriefing in simulated-based learning: facilitating a reflective discussion. *J Perinat Neonatal Nurs*, *24*(4), 301–302. http://doi.org/10.1097/JPN.0b013e3181f6b5ec\r00005237-201010000-00006 [pii]

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., de Freitas, S., Louchart, S., … De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, *46*(2), 391–411. http://doi.org/10.1111/bjet.12113

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning Analytics: From Research to Practice* (pp. 61–75). http://doi.org/10.1007/978-1-4614-3305-7_4

Baker, R. S. J. d., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM - Journal of Educational Data Mining*, *1*(1), 3–17.

Baranowski, T., Buday, R., Thompson, D. I., & Baranowski, J. (2008). Playing for Real. Video Games and Stories for Health-Related Behavior Change. *American Journal of Preventive Medicine*. http://doi.org/10.1016/j.amepre.2007.09.027

Baruque, C. B., Amaral, M. A., Barcellos, A., Freitas, J. C. D. S., & Longo, C. J. (2007). Analysing users' access logs in Moodle to improve e Learning. *Proceedings of the 2007 Euro American Conference on Telematics and Information Systems*, 1–4. http://doi.org/10.1145/1352694.1352767

Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., & Baghaei, N. (2010). Physical Activity Motivating Games : Virtual Rewards for Real Activity. *CHI 2010: Games and Players*, 243–252. http://doi.org/10.1145/1753326.1753362

Bernardini, S., Porayska-Pomsta, K., & Smith, T. J. (2014). ECHOES: An intelligent serious game for fostering social communication in children with autism. *Information Sciences*, *264*, 41–60. http://doi.org/10.1016/j.ins.2013.10.027

Bondzulic, N. (2015). Google Analytics in games. Retrieved from http://www.gamasutra.com/blogs/NemanjaBondzulic/20150825/252136/Google_Anal ytics_in_games.php

Borro Escribano, B., Del Blanco, A., Torrente, J., Borro Mate, J. M., & Fernandez Manjon, B. (2015). Educational game development approach to a particular case: The donor's evaluation. In *Transplantation Proceedings* (Vol. 47, pp. 13–18). http://doi.org/10.1016/j.transproceed.2014.11.006

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … 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 and Education*, *94*, 178–192. http://doi.org/10.1016/j.compedu.2015.11.003

Bravo, J., & Ortigosa, A. (2009). Detecting Symptoms of Low Performance Using Production Rules. *Proceedings of the 2nd International Conference on Educational Data Mining*, 31–40. Retrieved from http://www.educationaldatamining.org/EDM2009/uploads/proceedings/bravo.pdf

Burns, M., & Colbert, M. (2013). Web analytics: The new purpose towards predictive mobile games. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8253 LNCS, pp. 1–13). http://doi.org/10.1007/978-3-319-03161-3_1

Burr, L., & Spennemann, D. H. R. (2004). Patterns of user behavior in university online forums. *International Journal of Instructional Technology and Distance Learning*, *1*(10), 11–18. Retrieved from http://itdl.org/Journal/Oct_04/article01.htm

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education*, *87*, 396–422. http://doi.org/10.1016/j.compedu.2015.07.011

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic Analytics. *Educause Review*, *42*, 40–57.

Canossa, A., & Drachen, A. (2009). Patterns of Play : Play-Personas in User-Centred Game Development. *Spore*. Retrieved from http://www.digra.org:8080/Plone/dl/db/09287.49165.pdf

Capone, W. (2004). SCORM Sharable Content Object Reference Model. *Info*.

Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C. I., Hauge, J. B., … Rauterberg, M. (2015). An activity theory-based model for serious games analysis and conceptual design. *Computers & Education*, *87*, 166–181. http://doi.org/10.1016/j.compedu.2015.03.023

Chanchary, F. H., Haque, I., & Khalid, M. S. (2008). Web Usage Mining to Evaluate the Transfer of Learning in a Web-Based Learning Environment. In *First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008)* (pp. 249–253). IEEE. http://doi.org/10.1109/WKDD.2008.139

Chang, K., Beck, J., Mostow, J., & Corbett, A. (2006). A Bayes Net Toolkit for Student Modeling in Intelligent Tutoring Systems (pp. 104–113). Springer Berlin Heidelberg. http://doi.org/10.1007/11774303_11

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, *50*(4), 31. http://doi.org/10.1145/1232743.1232769

Chen, Y.-L., & Weng, C.-H. (2009). Mining fuzzy association rules from questionnaire data. *Knowledge-Based Systems*, *22*(1), 46–56. http://doi.org/10.1016/j.knosys.2008.06.003

Clow, D., & Doug. (2013). MOOCs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13* (p. 185). New York, New York, USA: ACM Press. http://doi.org/10.1145/2460296.2460332

Connolly, T. M., Boyle, E. a., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686. http://doi.org/10.1016/j.compedu.2012.03.004

Cooper, A. (2014). Learning Analytics Interoperability - The Big Picture In Brief. *Learning Analytics Community Exchange*.

Daniel, J. (2012). Making Sense of MOOCs. *Commonwealth of Learning Blog*. http://doi.org/10.1145/1787234.1787242

Daradoumis, T., Bassi, R., Xhafa, F., & Caballe, S. (2013). A Review on Massive E-Learning (MOOC) Design, Delivery and Assessment. In *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (pp. 208–213). IEEE. http://doi.org/10.1109/3PGCIC.2013.37

Dawson, S., Bakharia, A., & Heathcote, E. (2010). SNAPP : Realising the affordances of real-time SNA within networked learning environments. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. De Laat, D. McConnell, & T. Ryberg (Eds.), *Learning* (pp. 125–133). Retrieved from http://www.lancs.ac.uk/fss/organisations/netlc/past/nlc2010/abstracts/PDFs/Dawson.pdf

DeKanter, N. (2004). Gaming redefines interactivity for learning. *TechTrends*, *49*(3), 26–31. http://doi.org/10.1007/BF02763644

Dekker, G. W. ., Pechenizkiy, M. ., & Vleeshouwers, J. M. . (2009). Predicting students drop out: A case study. *EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining*, 41–50. http://doi.org/doi:10.1037/0893-3200.21.3.344

del Blanco, Á., Marchiori, E. J., Torrente, J., Martínez-Ortiz, I., & Fernández-Manjón, B. (2013). Using e-learning standards in educational video games. *Computer Standards & Interfaces*, *36*(1), 178–187. http://doi.org/10.1016/j.csi.2013.06.002

Del Blanco Aguado, A., Torrente, J., Martínez-Ortiz, I., & Fernández-Manjón, B. (2011). Analysis of SCORM standard for the integration of educational games | Análisis del uso del estándar SCORM para la integración de juegos educativos. *Revista Iberoamericana de Tecnologias Del Aprendizaje*, *6*(3).

Denis, G., & Jouvelot, P. (2005). Motivation-driven educational game design. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology - ACE '05* (pp. 462–465). New York, New York, USA: ACM Press. http://doi.org/10.1145/1178477.1178581

Dickey, M. D. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development*, *53*(2), 67–83.

Dickey, M. D. (2006a). Game design and learning: a conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research and Development*, *55*(3), 253–273. http://doi.org/10.1007/s11423-006-9004-7

Dickey, M. D. (2006b). Game design narrative for learning: Appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research and Development*, *54*(3), 245–263. http://doi.org/10.1007/s11423-006-8806-y

Diego, J. P. S., Ballard, J., Hatzipanagos, S., Webb, M., Khan, E., Blake, P., … Barrett, I. (2012). Do Moodle analytics have a role to play in learning design, feedback and assessment? *1st Moodle Research Conference*, 61–68. Retrieved from http://research.moodle.net/mod/data/view.php?d=3&rid=8

Djaouti, D., Alvarez, J., & Jessel, J.-P. (2011). Classifying serious games: The G/P/S model. *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches*, (2005), 118–136. http://doi.org/10.4018/978-1-60960-495-0.ch006

Dondlinger, M. (2007). Educational video game design: A review of the literature. *Journal of Applied Educational Technology*, *4*(1), 21–31. http://doi.org/10.1108/10748120410540463

Dougiamas, M., & Taylor, P. (2003). Moodle: Using Learning Communities to Create an Open Source Course Management System. *Research.Moodle.Net*, 1–16. Retrieved from http://research.moodle.net/pluginfile.php/15/mod_data/content/1121/Moodle - Dougiamas - 2003.pdf

Drachen, A., Seif El-Nasr, M., & Canossa, A. (2015). Game Analytics - The Basics. In *Game Analytics: Maximizing the Value of Player Data* (pp. 13–40). Springer.

Drachen, A., Thurau, C., Togelius, J., & Bauckhage, C. (2013). Game data mining. In *Game analytics* (pp. 205–253). http://doi.org/10.1007/978-1-4471-4769-5_12

Dreyfuss, S. E., & Dreyfus, H. L. (1980). A five-stage model of the mental activities involved in directed skill acquisition. *Operations Research Center*, (February), 1–18.

Dringus, L. P., & Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, *45*(1), 141–160. http://doi.org/10.1016/j.compedu.2004.05.003

Dudzinski, M., Greenhill, D., Kayyali, R., Nabhani, S., Philip, N., Caton, H., … Gatsinzi, F. (2013). The Design and Evaluation of a Multiplayer Serious Game for Pharmacy Students. In *Proceedings of the 7th European Conference on Games Based Learning, Vols 1 and 2* (Vol. 1, pp. 140–148).

Egenfeldt-Nielsen, S. (2004). Practical barriers in using educational computer games. *On The Horizon - The Strategic Planning Resource for Education Professionals*, 18–21. http://doi.org/doi: 10.1108/10748120410540454

El-Nasr, M. S., Drachen, A., & Canossa, A. (2013). *Game Analytics: Maximizing the Value of Player Data*. Springer. http://doi.org/10.1007/978-1-4471-4769-5

Entertainment Software Assotiation. (2015). 2015 Essential Facts About the Computer and Video Game Industry. *Social Science Computer Review*, *4*(1), 1–20. Retrieved from http://www.theesa.com/facts/pdfs/ESA_EF_2008.pdf

Evans, K. H., Daines, W., Tsui, J., Strehlow, M., Maggio, P., & Shieh, L. (2015). Septris: a novel, mobile, online, simulation game that improves sepsis recognition and management. *Academic Medicine : Journal of the Association of American Medical Colleges*, *90*(2), 180–4. http://doi.org/10.1097/ACM.0000000000000611

Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 146–150. http://doi.org/10.1145/2723576.2723589

Farmer, J., & Dolphin, I. (2005). Sakai : eLearning and More. *Strategy*, 22–27. Retrieved from http://www.immagic.com/eLibrary/ARCHIVES/GENERAL/IMM/S050622F.pdf

Fausett, L. V., & Elwasif, W. (1994). Predicting performance from test scores using backpropagation and counterpropagation. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)* (Vol. 5, pp. 3398–3402). IEEE. http://doi.org/10.1109/ICNN.1994.374782

Fitts, P. M. (1964). Perceptual-Motor Skill Learning. *Categories of Human Learning*. http://doi.org/10.1080/00140139208967796

Fontenla, J., Pérez, R., & Caeiro, M. (2011). Using IMS basic LTI to integrate games in LMSs Lessons from GamėTel. In *2011 IEEE Global Engineering Education Conference, EDUCON 2011* (pp. 299–306). http://doi.org/10.1109/EDUCON.2011.5773152

Freire, M., del Blanco, A., & Fernandez-Manjon, B. (2014). Serious games as edX MOOC activities. In *2014 IEEE Global Engineering Education Conference (EDUCON)* (pp. 867–871). IEEE. http://doi.org/10.1109/EDUCON.2014.6826198

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation and Learning: A Research and Practice Model. *Simulation & Gaming*, *33*(4), 441–467.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*. New York ; Basingstoke: Palgrave Macmillan. http://doi.org/10.1145/950566.950595

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, *2*(1), 5–16. http://doi.org/10.2304/elea.2005.2.1.5

Golding, P., & Donaldson, O. (2006). Predicting Academic Performance. In *Frontiers in Education Conference 36th Annual* (pp. 21–26). http://doi.org/10.1109/FIE.2006.322661

Goldstein, P. J. (2005). Academic Analytics : The Uses of Management Information and Technology in Higher Education. *Phoenix Usa*, *8*(December), 1–12. http://doi.org/10.1080/17439880802097659

Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *Br J Surg*, *99*(10), 1322–1330. http://doi.org/10.1002/bjs.8819

Griffiths, D., Hoel, T., & Cooper, A. (2016). *Learning Analytics Interoperability: Requirements, Specifications and Adoption*. Retrieved from http://www.laceproject.eu/wp-content/uploads/2016/01/LACE_D7-4.pdf

Hämäläinen, W., & Vinni, M. (2006). Comparison of Machine Learning Methods for Intelligent Tutoring Systems (pp. 525–534). Springer Berlin Heidelberg. http://doi.org/10.1007/11774303_52

Hasan, L., Morris, A., & Probets, S. (2009). Using google analytics to evaluate the usability of e-commerce sites. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 5619 LNCS, pp. 697–706). http://doi.org/10.1007/978-3-642-02806-9_81

Hassler, M. (2010). Web Analytics. *Redline Heidelberg*, *3*(1), 1–14. http://doi.org/10.1080/19322900802660292

Hauge, J. B., Berta, R., Fiucci, G., Manjon, B. F., Padron-Napoles, C., Westra, W., & Nadolski, R. (2014a). Implications of Learning Analytics for Serious Game Design. In *2014 IEEE 14th International Conference on Advanced Learning Technologies* (pp. 230–232). IEEE. http://doi.org/10.1109/ICALT.2014.73

Hauge, J. B., Berta, R., Fiucci, G., Manjon, B. F., Padron-Napoles, C., Westra, W., & Nadolski, R. (2014b). Implications of Learning Analytics for Serious Game Design. In *Proceedings of the 14th International Conference on Advanced Learning Technologies* (pp. 230–232). IEEE. http://doi.org/10.1109/ICALT.2014.73

Hauge, J. B., & Riedel, J. C. K. H. (2012). Evaluation of Simulation Games for Teaching Engineering and Manufacturing. *Procedia Computer Science*, *15*, 210–220. http://doi.org/10.1016/j.procs.2012.10.073

Hays, R. T. (2005). The Effectiveness of Instructional Games: A Literature Review and Discussion. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA441935

Hernández, J. A., Ochoa, A., Muñoz, J., & Burlaka, G. (2006). Detecting cheats in online student assessments using Data Mining. In *Conference on Data Mining | DMIN* (p. Vol. 6, 205).

Hien, N. T. N., & Haddawy, P. (2007). A decision support system for evaluating international student applications. In *Proceedings - Frontiers in Education Conference, FIE* (p. F2A–1–F2A–6). IEEE. http://doi.org/10.1109/FIE.2007.4417958

Hollins, P., Westera, W., & Manero, B. (2015). Amplifying applied game development and uptake. Retrieved from http://dspace.ou.nl/handle/1820/6140

Hsia, T.-C., Shie, A.-J., & Chen, L.-C. (2008). Course planning of extension education to meet market demand by using data mining techniques – an example of Chinkuo technology university in Taiwan. *Expert Systems with Applications*, *34*(1), 596–602. http://doi.org/10.1016/j.eswa.2006.09.025

Ibrahim, R., Semarak, J., Lumpur, K., & Jaafar, A. (2010). Using educational games in learning introductory programming: A pilot study on students' perceptions. In *2010 International Symposium on Information Technology* (pp. 1–5). IEEE. http://doi.org/10.1109/ITSIM.2010.5561414

IMS Global Consortium. (2016). Caliper Analytics. Retrieved April 1, 2016, from http://www.imsglobal.org/activity/caliperram

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., & Freeman, A., and Hall, C. (2015). *Horizon Report: 2016 Higher Education Edition*. *Reading*. http://doi.org/ISBN 978-0-9906415-8-2

Johnson, L., Adams, S., & Cummins, M. (2012). *NMC Horizon Report: 2012 K-12 Edition*. Austin, Texas: The New Media Consortium.

Johnson, W. L., & Wang, N. (2007). Experience with serious games for learning foreign languages and cultures. In *Proceedings of the SimTecT Conference*. Retrieved from http://alelo.com/files/SIMTECT07-experience_with_serious_games.pdf

Jonassen, D. H., & Rohrer-Murphy, L. (1999). Activity Theory as a Framework for Designing Constructivist Learning Environments. *Educational Technology Research and Development*, *47*(I), 61–79. http://doi.org/10.1007/BF02299477

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *Internet and Higher Education*, *8*(1), 13–24. http://doi.org/10.1016/j.iheduc.2004.12.001

Kim, J., Chern, G., Feng, D., Shaw, E., & Hovy, E. (2006). Mining and Assessing Discussions on the Web through Speech Act Analysis. *Proceedings of the ISWC06 Workshop on Web Content Mining with Human Language Technologies*, (April 2016). Retrieved from http://www.isi.edu/~jihie/papers/SA-analysis-cmnts-2006-Kim.pdf

Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B., Pagulayan, R. J., & Wixon, D. (2008). Tracking real-time user experience (TRUE). In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* (p. 443). New York, New York, USA: ACM Press. http://doi.org/10.1145/1357054.1357126

Koedinger, K. R., Baker, R. S. J., Cunningham, K., & Skogsholm, A. (2010). A Data Repository for the EDM community : The PSLC DataShop. *Handbook of Educational Data Mining*, 43–55. http://doi.org/doi:10.1201/b10274-6

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218. http://doi.org/10.1207/s15430421tip4104

Lee, S. J., Liu, Y., & Popovic, Z. (2014). Learning Individual Behavior in an Educational Game : A Data-Driven Approach. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM)* (pp. 114–121). Retrieved from http://grail.cs.washington.edu/wp-content/uploads/2015/08/lee2014lib.pdf

Liñán, L. C., & Pérez, Á. A. J. (2015). Educational data mining and learning analytics: differences, similarities, and time evolution. *Rusc*, *12*(3), 98–112. Retrieved from http://0-search.proquest.com.aupac.lib.athabascau.ca/docview/1701132604?accountid=8408%5Cnhttp://aupac.lib.athabascau.ca:4550/resserv?genre=article&issn=&title=RUSC&volume=12&issue=3&date=2015-07-01&atitle=Educational+Data+Mining+and+Learning+Analytics:+d

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious Games Analytics: Theoretical Framework. In *Serious Games Analytics* (pp. 3–29). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-05834-4_1

Loh, C. S., Sheng, Y., & Li, I.-H. (2015). Predicting expert–novice performance as serious games analytics with objective-oriented and navigational action sequences. *Computers in Human Behavior*, *49*, 147–155. http://doi.org/10.1016/j.chb.2015.02.053

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *Educause Review*, *46*(5), 31–40. Retrieved from http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume46/PenetratingtheFogAnalyticsinLe/235017

Malliarakis, C., Satratzemi, M., & Xinogalos, S. (2014). CMX: Implementing an MMORPG for learning programming. In *Proceedings of the European Conference on Games-based Learning* (Vol. 1, pp. 346–355). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84923604194&partnerID=tZOtx3y1

Manero, B., Fernández-Vara, C., & Fernández-Manjón, B. (2013a). E-learning a escena: De La Dama Boba a Juego Serio. *Vaep Rita*, *1*(1), 51–58.

Manero, B., Fernández-Vara, C., & Fernández-Manjón, B. (2013b). E-learning takes the stage: From La Dama Boba to a serious game. *Revista Iberoamericana de Tecnologias Del Aprendizaje*, *8*(4), 197–204. http://doi.org/10.1109/RITA.2013.2285023

Manero, B., Torrente, J., Serrano, Á., Martínez-Ortiz, I., & Fernández-Manjón, B. (2015). Can educational video games increase high school students' interest in theatre? *Computers & Education*, *87*, 182–191. http://doi.org/http://dx.doi.org/10.1016/j.compedu.2015.06.006

Marchiori, E. J., Torrente, J., Del Blanco, Á., Moreno-Ger, P., Sancho, P., & Fernández-Manjón, B. (n.d.). A narrative metaphor to facilitate educational game authoring. *Computers & Education*. http://doi.org/10.1016/j.compedu.2011.09.017

Marne, B., Wisdom, J., Huynh-Kim-Bang, B., & Labat, J.-M. (2012). The six facets of serious game design: a methodology enhanced by our design pattern library. In *21st Century Learning for 21st Century Skills* (Vol. 7563, pp. 208–221). http://doi.org/10.1007/978-3-642-33263-0

Medler, B. (2009). Generations of Game Analytics, Achievements and High Scores. *Eludamos. Journal for Computer Game Culture*. Retrieved from http://www.eludamos.org/index.php/eludamos/article/view/vol3no2-4/127

Medler, B., & Magerko, B. (2011). Analytics of play: Using information visualization and gameplay practices for visualizing video game data. *Parsons Journal for Information Mapping*, *3*(1), 1–12. Retrieved from http://piim.newschool.edu/journal/issues/

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A Brief Introduction to Evidence-Centered Design. *Russell The Journal Of The Bertrand Russell Archives*, *1522*(July), 1–37. http://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J. L., & Fernández-Manjón, B. (2008). Educational game design for online education. *Computers in Human Behavior*, *24*(6), 2530–2540. http://doi.org/10.1016/j.chb.2008.03.012

Mozilla. (2011). Open Badges for Lifelong Learning. *White Paper*, 1–14. Retrieved from https://wiki.mozilla.org/File:OpenBadges-Working-Paper_092011.pdf

Nwana, H. (1990). Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, *4*(4), 251–277. http://doi.org/10.1007/BF00168958

Owen, V. E., Ramirez, D., Salmon, A., & Halverson, R. (2014). Capturing Learner Trajectories in Educational Games through ADAGE (Assessment Data Aggregator for Game Environments): A Click-Stream Data Framework for Assessment of Learning in Play. Retrieved September 8, 2014, from http://www.academia.edu/7759291/Capturing_Learner_Trajectories_in_Educational_Games_through_ADAGE_Assessment_Data_Aggregator_for_Game_Environments_A_Click-Stream_Data_Framework_for_Assessment_of_Learning_in_Play

Plaza, B. (2011). Google Analytics for measuring website performance. *Tourism Management*, *32*(3), 477–481. http://doi.org/10.1016/j.tourman.2010.03.015

Price, C. B. (2007). The usability of a commercial game physics engine to develop physics educational materials: An investigation. *Simulation & Gaming*, *39*(3), 319–337. http://doi.org/10.1177/1046878108319579

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., … Kafai, Y. (2009). Scratch : Programming for all. *Communications of the ACM*, *52*(11), 60–67. http://doi.org/10.1145/1592761.1592779

Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers and Education*, *68*, 458–472. http://doi.org/10.1016/j.compedu.2013.06.009

Romero, C., & Ventura, S. Educational Data Mining: A Review of the State of the Art, 40IEEE Transactions on Systems Man and Cybernetics Part C Applications and Reviews 601–618 (2010). IEEE. http://doi.org/10.1109/TSMCC.2010.2053532

Rud, O. (2009). Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy. Retrieved December 7, 2014, from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470392401.html

Sancho, P., Torrente, J., & Fernández-Manjón, B. (2009). Do Multi-User Virtual Environments Really Enhance Student's Motivation in Engineering Education? San Antonio, TX, USA: IEEE digital library.

Scalise, K., & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *The Journal of Technology, Learning, and Assessment*, *4*(6), 45. Retrieved from http://napoleon.bc.edu/ojs/index.php/jtla/article/view/1653

Schmitz, H.-C., Wolpers, M., Kirschenmann, U., & Niemann, K. (2007). Contextualized Attention Metadata. *Human Attention in Digital Environments*, *13*, 1082–9873. http://doi.org/10.1045/september2007-wolpers

Scott, J. (2011). Social network analysis: developments, advances, and prospects. *Social Network Analysis and Mining*, *1*(1), 21–26. http://doi.org/10.1007/s13278-010-0012-6

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M.-W. (2016). Advances in the Science of Assessment. *Educational Assessment*, *21*(1), 34–59. http://doi.org/10.1080/10627197.2015.1127752

Siemens, G., & Baker, R. S. J. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 29 April – 2 May, Vancouver, BC, Canada.*

Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529.

http://doi.org/10.1177/0002764213479366

Snell, J., Atkins, M., Norris, W., Messina, C., Wilkinson, M., & Dolin, R. (2011). JSON Activity Streams 1.0. *Activity Streams Working Group*, *22*(8), 2013.

Squire, K. (2003). Video games in education. International Journal of Intelligent. *Simulations and Gaming*, *2*(1), 49–62.

Squire, K., Barnett, M., Grant, J. M., & Higginbotham, T. (2004). Electromagnetism supercharged!: learning physics with digital simulation games. In *International Conference on Learning Sciences* (pp. 513–520). International Society of the Learning Sciences. Retrieved from http://portal.acm.org/citation.cfm?id=1149126.1149189

Torrente, J., Borro-Escribano, B., Freire, M., Del Blanco, ??ngel, Marchiori, E. J., Mart??nez-Ortiz, I., … Fern??ndez-Manj??n, B. (2014). Development of game-like simulations for procedural knowledge in healthcare education. *IEEE Transactions on Learning Technologies*, *7*(1), 69–82. http://doi.org/10.1109/TLT.2013.35

Torrente, J., Del Blanco, Á., Marchiori, E. J., Moreno-Ger, P., & Fernández-Manjón, B. (2010). e-Adventure: Introducing Educational Games in the Learning Process. In *IEEE Education Engineering (EDUCON) 2010 Conference* (pp. 1121–1126). http://doi.org/10.1109/EDUCON.2010.5493056

Torrente, J., Del Blanco, Á., Moreno-Ger, P., & Fernández-Manjón, B. (2012). Designing Serious Games for Adult Students with Cognitive Disabilities. In T. Huang, Z. Zeng, C. Li, & C. Leung (Eds.), *Neural Information Processing, Lecture Notes in Computer Science Volume 7666* (Vol. 7666, pp. 603–610). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-34478-7_73

van der Graaf, J., Segers, E., & Verhoeven, L. (2016). Discovering the laws of physics with a serious game in kindergarten. *Computers & Education*, *101*, 168–178. http://doi.org/10.1016/j.compedu.2016.06.006

Ventura, M., Shute, V. J., Methods, P., Games, E., Worlds, V., Barzilai, S., … Boyer, D. M. (2013). Stealth assessment: Measuring and supporting learning in video games. *IEEE Transactions on Learning Technologies*, *4*(June), 135–137. http://doi.org/10.1109/TLT.2010.21

Ventura, S., & Romero, C. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*. http://doi.org/10.1016/j.eswa.2006.04.005

Watson, H. J., & Wixom, B. H. (2007). The Current State of Business Intelligence. *Computer*, *40*(9), 96–99. http://doi.org/10.1109/MC.2007.331

Williams, S., & Williams, N. (2003). The Business Value of Business Intelligence. *Intelligence*, *8*(301), 30–39. Retrieved from http://www.decisionpath.com:8180/DPC_revised/html/BIJarticle.pdf

Ye, F. (2014). *Validity, reliability, and concordance of the Duolingo English Test*. Retrieved from https://s3.amazonaws.com/duolingo-certifications-data/CorrelationStudy.pdf

Yudelson, M. V, Medvedeva, O., Legowski, E., Castine, M., Jukic, D., & Crowley, R. S. (2005). Mining student learning data to develop high level pedagogic strategy in a medical ITS. *Proceedings of AAAI Workshop on Educational Data Mining*.

Yusoff, A., Crowder, R., Gilbert, L., & Wills, G. (2009). A conceptual framework for serious games. In *Proceedings - 2009 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009* (pp. 21–23). http://doi.org/10.1109/ICALT.2009.19

Zoeller, G. (2010). Development Telemetry in Video Games Projects. In *Game Developers Conference*. Retrieved from http://gdc.gulbsoft.org/talk