



Universidad de Valladolid

TRABAJO FIN DE MÁSTER

MÁSTER EN PROFESOR DE EDUCACIÓN
SECUNDARIA OBLIGATORIA Y
BACHILLERATO, FORMACIÓN PROFESIONAL
Y ENSEÑANZAS DE IDIOMAS

Especialidad de Tecnología e Informática

Learning Analytics: Propuesta metodológica y caso de estudio en educación no universitaria

Autor:

D. Víctor Sanz Gómez

Tutor:

Dr. D. Manuel Barrio Solorzano

Valladolid, 12 de Julio de 2018



Universidad de Valladolid

MASTER'S THESIS

MASTER'S PROGRAMME IN TEACHER TRAINING FOR
SECONDARY EDUCATION AND VOCATIONAL TRAINING

The speciality of Computer Science and Technology

**Learning Analytics:
Methodological proposal and
Case Study for non-University
Education**

Author:

D. Víctor Sanz Gómez

Mentor:

Dr. D. Manuel Barrio Solorzano

Valladolid, July 12, 2018

Resumen

Learning Analytics es una disciplina de análisis de datos de y para el aprendizaje que actualmente está en proceso de desarrollo. En este trabajo vamos a presentar una visión general sobre Learning Analytics, o en castellano analíticas del aprendizaje. Vamos a conocer qué es, y cómo podemos ponerlo en práctica, dirigiéndonos principalmente a su utilización en la educación secundaria, proponiendo un método y herramientas para su aplicación. Por último, expondremos un caso de estudio usando este método y alguna de las herramientas propuestas con el fin de ilustrar este trabajo.

Abstract

Learning Analytics is a data analysis discipline related to education which is currently in development process. In this thesis we will present an overview of Learning Analytics. We will discover what it is, and how we can put it into practice. We will focus mainly on its use in secondary education, proposing a method and tools for its application. Finally, we will present a case study using this method and some of the tools proposed in order to illustrate this thesis.

Tabla de contenido

Tabla de ilustraciones	1
1 Introducción.....	3
1.1 Introducción	3
1.2 Motivación del trabajo	3
1.3 Objetivos y metodología	4
1.4 Estructura de la memoria.....	4
2 Una primera aproximación a Learning Analytics.....	5
2.1 ¿Qué son las Learning Analytics?.....	5
2.2 Learning Analytics en la actualidad	8
2.3 Las barreras de Learning Analytics.....	9
2.4 Conclusiones	11
3 Metodología propuesta	13
3.1 Identificar usuarios	14
3.2 Identificar necesidades y plantear objetivos.....	17
3.3 Recogida de datos.....	20
3.4 Analizar	23
3.4.1 Métodos predictivos.....	24
3.4.2 Descubridores de estructura.....	25
3.4.3 Minería de relaciones	26
3.5 Visualizar	26
3.6 Actuar	30
3.7 Conclusiones	32
4 Caso de estudio	35
5 Conclusiones y Líneas Futuras	43
5.1 Conclusiones	43
5.2 Líneas futuras	44
6 Referencias	47

Tabla de ilustraciones

Figura 1 Definición propuesta en la conferencia LAK'11 (2011)	6
Figura 3 Esquema de la propuesta	34
Figura 4 Código de Plickers.....	36
Figura 5 Visualización de Plickers	38
Tabla 1 (Siemens, Learning Analytics: Advancing the Science of Learning, 2015)...	8
Tabla 2 (Siemens, Learning Analytics: Advancing the Science of Learning, 2015)	13
Tabla 3 Comparación entre Long y Siemens (2012) y Buckingham (2012)	15
Tabla 4 (Siemens, Learning Analytics: Advancing the Science of Learning, 2015)	16
Tabla 5 Fuente: Gartner (octubre 2014)	18
Tabla 6 Diagramas de barras bidimensionales	27
Tabla 7 Histograma de calificaciones	27
Tabla 8 Caja y bigotes para un conjunto de ejercicios	28
Tabla 9 Mapa de calor	29
Tabla 10 Gráficos de dispersión y covarianza entre las variables	30
Tabla 11 Gráfico de Plickers	37
Tabla 12 Progresión del porcentaje de aciertos por pregunta	39
Tabla 13 Gráfico de barras.....	39
Tabla 14 Gráfico de caja y bigotes de las calificaciones de los alumnos por días	40
Tabla 15 Agrupación de preguntas por metodología utilizada	41

1 Introducción

1.1 Introducción

Learning Analytics (LA), es una disciplina de análisis de datos de y para el aprendizaje que actualmente está en proceso de desarrollo. Hace años comenzó su andadura con los cursos online masivos, continuó avanzando por distintas facultades y hoy en día está dando sus primeros pasos en la educación primaria y secundaria.

Mientras las plataformas virtuales van cobrando fuerza en el panorama educativo nacional, las analíticas de aprendizaje se aprovecharán de la posibilidad de obtener cada vez mayor cantidad de datos y tomarán más y más presencia como una herramienta de apoyo para facilitar el trabajo a los docentes.

Estas analíticas son capaces de explicar la realidad pasada y presente, otorgar sentido a esas realidades y hacer labores predictivas y prescriptivas. Esto permitirá lograr una mayor calidad en la educación adaptando y personalizando el aprendizaje a lo que necesita cada alumno en cada momento.

En este trabajo vamos a presentar una visión general sobre Learning Analytics, o en castellano analíticas del aprendizaje. Vamos a conocer qué es, y cómo podemos ponerlo en práctica, dirigiéndonos principalmente a su aplicación en la educación secundaria, proponiendo un método y herramientas para su aplicación. Por último, expondremos un caso de estudio usando este método y alguna de las herramientas propuestas con el fin de ilustrar este trabajo. Este estudio práctico fue realizado en una clase de un ciclo formativo de grado medio en un centro concertado de la localidad de Valladolid, en Castilla y León, España, por el autor de este trabajo.

1.2 Motivación del trabajo

El desarrollo de este trabajo responde al interés del alumno por la aplicación de nuevas técnicas de mejora de la calidad educativa a partir del uso de las nuevas tecnologías disponibles.

La gran expansión que están teniendo las tecnologías de la información y la comunicación en todos los ámbitos de la sociedad, tales como la Inteligencia Artificial y el Big Data, no han pasado inadvertidas en los centros educativos universitarios. Pese a esto, estas herramientas TICs, y con ellas Learning Analytics, aún se encuentran lejos de incidir en los centros de educación primaria y secundaria.

Por ello, a partir de este trabajo y con la inclusión de Entornos Virtuales de Aprendizaje basados en Moodle, esperamos motivar a los distintos actores educativos a considerar e impulsar el uso de Learning Analytics en sus centros.

1.3 Objetivos y metodología

El objetivo principal de este trabajo es establecer un método para la aplicación de Learning Analytics en un aula de secundaria por parte de cualquier docente, no especializado necesariamente en la analítica de datos. Este método se complementará con un ejemplo a través de un caso de estudio.

Como objetivos específicos, identificamos los siguientes:

- Dar a conocer Learning Analytics como un conjunto de herramientas cuya aplicación permite optimizar el desempeño en entornos educativos.
- Presentar el estado actual de Learning Analytics y los factores que han motivado su desarrollo, así como las barreras a las que se enfrenta.
- Desarrollar una metodología de actuación que permita aplicar Learning Analytics en educación secundaria.
- Ilustrar el marco metodológico propuesto a través de un caso de estudio realizado durante el prácticum del master.

La metodología aplicada para el logro de estos objetivos ha sido la revisión de literatura con el fin de conocer las posibilidades de Learning Analytics, el diseño del método de aplicación con su consecuente búsqueda de herramientas que lo complementen, y por último la realización del caso de estudio siguiendo los pasos previamente diseñados.

1.4 Estructura de la memoria

El contenido principal de la memoria lo hemos estructurado en tres capítulos definidos, precedidos por esta introducción.

El primero de ellos nos acercará a las Learning Analytics. Conoceremos los enfoques y la visión de esta disciplina, las razones principales por las que se ha incrementado tanto su investigación y desarrollo en los últimos años, su estado actual y por último las barreras que se le presentan y pueden entorpecer su crecimiento.

El segundo capítulo es, asemejando el trabajo a una obra narrativa, el nudo de la historia. Basándonos en la teoría presentada en el anterior epígrafe, aterrizaremos y será aquí donde plantearemos el método a seguir y propondremos una serie de herramientas y técnicas para utilizarlo.

El tercer capítulo, el desenlace, será donde pongamos en práctica este método en un caso concreto: en un módulo de un ciclo formativo en el que contamos con la dificultad añadida de que no usan ningún entorno virtual de aprendizaje.

Al finalizar expondremos las conclusiones y líneas futuras del trabajo en un último epígrafe.

Por último, enumeraremos todas las referencias bibliográficas consultadas durante la realización del trabajo.

2 Una primera aproximación a Learning Analytics

A lo largo de este capítulo vamos a conocer en más detalle qué son las Learning Analytics, con qué otras disciplinas limitan, sus usos actuales y potenciales, y por último las barreras que pueden dificultar el avance que está teniendo.

2.1 ¿Qué son las Learning Analytics?

El origen de las Learning Analytics (en adelante LA) será nuestro punto de partida, ya que se encuentra muy próximo al momento actual y nos servirá para conocer su desarrollo y realidad. Son varias las razones por las que la investigación en LA se ha visto fuertemente incrementada en los últimos años (Ferguson, Learning analytics: drivers, developments and challenges, 2012).

La primera de las razones detonantes fueron los MOOCs, (Massive Open Online Courses), cursos online masivos en el que una cantidad limitada de profesores, similar a la de un curso ordinario actual, tenía que hacer frente a la gestión de miles de alumnos. Mientras que en las clases universitarias tradicionales los profesores podían gestionar sus clases de forma más personalizada, adaptando los conocimientos al alumno, en los MOOCs resulta una tarea muy compleja y laboriosa. Por otro lado, este tipo de entornos cuentan con la ventaja de que se pueden recolectar multitud de datos de los estudiantes: interacciones entre alumnos, entre alumnos y profesores, tiempo dentro del entorno virtual de aprendizaje, fecha de entrega de las actividades, calificaciones obtenidas en las actividades...

Este aprendizaje en entornos online es precisamente la segunda de las principales razones del incremento de uso de LA y por la que se piensa que tiene un futuro muy prometedor. No sólo los MOOC emplean entornos virtuales educativos, sino que multitud de instituciones y universidades hacen uso de estas herramientas. En España, las aulas virtuales basadas en la plataforma Moodle son muy conocidas en la comunidad educativa, sobre todo en los entornos de la educación superior. Su utilidad ha propiciado un proyecto en Castilla y León para la instauración de aulas virtuales basadas en esta plataforma en los centros de secundaria de la comunidad a lo largo de los próximos años (Portal de Educación de la Junta de Castilla y León, 2017). Estas herramientas permiten recoger una gran cantidad de datos potencialmente analizables.

La tercera y última razón de la que hablaremos es del Big Data. El despunte de esta tecnología, sobre todo a través del marketing, ha llegado y afectado a muchas áreas de la sociedad y, como no podría ser de otra manera, a la educación también. Es por eso que gracias al desarrollo que ha tenido el Big Data en los últimos años, se ha podido desarrollar LA. La vastísima cantidad de datos recolectados permite usar técnicas de esta disciplina para encontrar patrones y a través del análisis de los datos poder predecir

eficazmente casos de éxito y fracaso en un determinado momento y poder actuar, automática o manualmente, para adaptar y mejorar el proceso de aprendizaje.

Los usuarios que parten de este análisis se pueden clasificar en tres tipos (Ferguson, Learning analytics: drivers, developments and challenges, 2012) (Buckingham Shum, 2012): un nivel mayor por estructuras gubernamentales regionales, nacionales e internacionales; un segundo nivel en el que los usuarios son las instituciones y centros; y por último un nivel más bajo en el que se encuentran los profesores y los estudiantes. Más adelante, en el apartado 3.1, hablaremos más detenidamente sobre los usuarios.

A continuación, lo que vamos a hacer es definir el concepto de LA. LA ha sido definida por muchos autores en los últimos años, aunque nosotros vamos a usar la que se está considerando como la principal definición de LA, expuesta en la primera conferencia internacional de Learning Analytics and Knowledge (LAK) en 2011 por SoLAR (Society of Learning Analytics Research). Desde entonces se considera LA como la medición, recolección, análisis e informe de datos sobre los estudiantes y sus contextos con el objetivo de comprender y optimizar el aprendizaje y el entorno en el que ocurre.

“Learning Analytics es la medición, recolección, análisis e informe de datos de los estudiantes y sus contextos, con el objetivo de comprender y optimizar el aprendizaje y el entorno en el que ocurre.”

Figura 1 Definición propuesta en la conferencia LAK'11 (2011)

Podemos extraer de esta definición que LA es un proceso en el que podemos distinguir al menos las siguientes fases:

1. Medición: establecimiento de objetivos
2. Recolección de los datos
3. Análisis y visualización de los datos
4. Sacar conclusiones y actuar en consecuencia.

Partiendo desde esta definición Long & Siemens (2011) establecen una clasificación de enfoque, diferenciando el análisis de datos en la educación en dos disciplinas distintas, Learning Analytics y Academic Analytics. Se refieren a la primera como la que está específicamente dirigida al proceso de aprendizaje mientras que la segunda tiene que ver con el análisis de datos a nivel institucional. Entraremos al detalle de esta clasificación en el apartado 3.1 Identificar usuarios.

Ahora, si queremos conocer mejor qué es LA también tenemos que saber qué no es, dónde están sus límites y diferencias con las disciplinas hermanas.

En este primer acercamiento a LA, en el que entre otras cosas hemos conocido su definición, los lectores habrán podido percibir que estaba propuesta por una comunidad de investigadores, asociados en SoLAR. Pero resulta que, en este creciente interés en los datos y analíticas en la educación, no se ha investigado únicamente desde esta comunidad,

sino que comparte campo de estudio con la comunidad de Educational Data Mining (EDM) o minería de datos educativos y su sociedad International Educational Data Mining Society (IEDMS). Ambas sociedades cuyo objetivo es la investigación en estas disciplinas son de carácter internacional, teniendo socios contribuyendo desde varios países del globo.

IEDMS define EDM de la siguiente manera: “La minería de datos educativos es una disciplina emergente preocupada por desarrollar métodos para explorar tipos únicos de datos provenientes de entornos educativos, y usar esos métodos para entender mejor a los estudiantes y los entornos en los que aprenden.” (International Educational Data Mining Society, s.f.).

Hemos presentado ambas comunidades para aclarar la visión de la analítica de datos desde LA, comparándola con EDM. Creemos que a través de esta comparación podremos acercarnos a LA desde otro punto de vista aportando riqueza a nuestro conocimiento. Siemens y Baker (2012) (Siemens & Baker, Learning analytics and educational data mining: towards communication and collaboration, 2012) analizaron las similitudes y diferencias haciendo una llamada de colaboración entre ambas comunidades. Es de este artículo del que vamos a extraer la visión de LA.

La similitud principal entre ambas vertientes es obvia, ambas tienen el mismo campo de estudio y su meta es la misma sobre todo cuando hablamos en términos de grandes escalas de datos. Esto sugiere que bastantes áreas de estudio se superponen, pudiéndose enriquecer mutuamente.

Pero no es en la similitud sino en la diferencia, donde vamos a poder conocer los detalles de LA. El primer punto del que hace distinción es el tipo de descubrimiento que se prioriza. Mientras que EDM se focaliza en descubrimientos automatizados, LA está más interesado en análisis dirigidos por el juicio humano.

La segunda diferencia a la que hace referencia es al tipo de adaptación y personalización educativa. Por un lado, los modelos de EDM suelen basarse en adaptación automatizada llevada a cabo por un sistema informático. En cambio, los modelos de LA normalmente están diseñados para informar y empoderar a los profesores y estudiantes para que sean ellos los que lleven a cabo las acciones que consideren.

El tercer y último punto clave es sobre el paradigma del marco de trabajo; mientras que típicamente EDM trabaja sobre un marco reduccionista, LA tiende hacia el paradigma holístico, viendo y trabajando con el sistema como un “todo”.

De forma sencilla y a modo de resumen, Ferguson (2012) hace una distinción en el reto que se plantea cada una de estas disciplinas. Afirma que EDM se centra en el reto técnico del análisis de datos educativos, LA en el reto educativo, y por último Academic Analytics en el reto político/económico.

2.2 Learning Analytics en la actualidad

Como ya hemos comentado anteriormente, los MOOCs hacen cada vez más fácil la recolección de datos sobre los alumnos y sus actividades, pero no son los únicos, ya que los entornos virtuales de aprendizaje también facilitan mucho esta tarea en las aulas tradicionales. A estos entornos o sistemas se les conoce por sus siglas en inglés como VLE (Virtual Learning Environment) o LMS (Learning Management System).

Precisamente, en Castilla y León se ha comenzado con un proyecto que pretende proporcionar aulas virtuales basadas en la plataforma Moodle a todos los centros de secundaria de la comunidad en los próximos años (Portal de Educación de la Junta de Castilla y León, 2017). Este hecho, sin lugar a duda, acercará y posibilitará el uso de LA a una gran cantidad de potenciales usuarios: estudiantes y profesores.

En la Tabla 1, Siemens presenta una progresión de maduración de las LA.

Para ponernos en contexto de esta tabla, Siemens propone este proceso de maduración para las LA en un entorno universitario, donde LA las desarrolle un equipo específico con al menos un experto en analítica de datos, un experto pedagógico, y un programador.

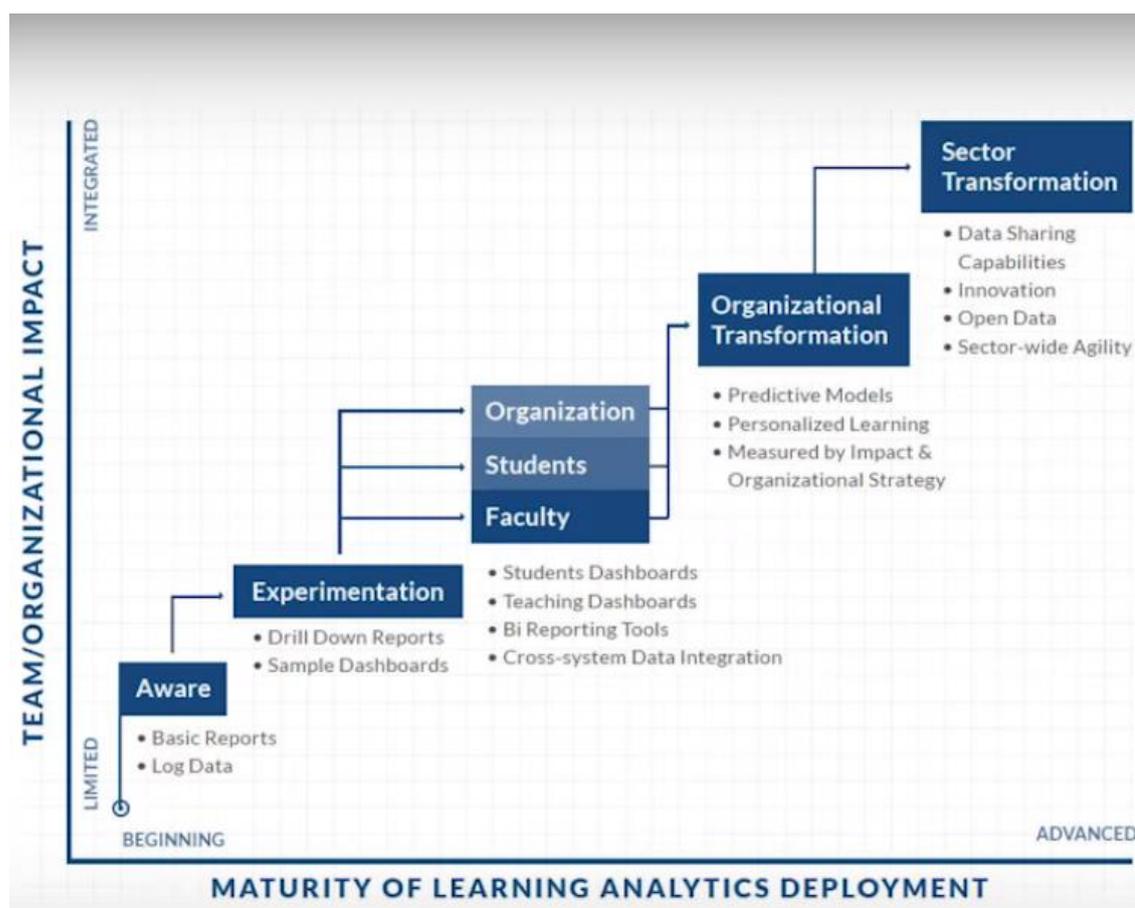


Tabla 1 (Siemens, *Learning Analytics: Advancing the Science of Learning*, 2015)

En esta conferencia de 2015, Siemens afirma que el estado actual de LA en los centros en los que se está desarrollando se encuentra en el tercer nivel, en el que la institución entera toma una conciencia real de las posibilidades de LA y se engrana el sistema para

que todos los roles puedan interactuar entre ellos. De este engranaje resultan paneles de control informativos (Dashboards) para los alumnos y los profesores y las analíticas afectan a la dirección del centro. Para el desarrollo efectivo de este tercer nivel, se precisa una buena organización institucional y un apoyo firme por parte del centro y la dirección.

Se espera que en un periodo de tiempo a corto-medio plazo se continúe progresando a los siguientes niveles madurativos. Pero, ¿hasta qué punto podemos considerar esto cierto? Desde el comienzo de LA, se ha hablado de un tiempo de desarrollo e implantación entre 2 y 5 años (Sharples, y otros, 2013) (Sharples, y otros, 2014) (Sharples, y otros, 2016). Tanto la investigación y desarrollo de esta disciplina ha avanzado, sobre todo cuantitativamente, en esos periodos temporales. Pero no hasta el punto de una gran implantación, ni siquiera de una puesta en marcha significativa.

Esto nos lleva directamente a una pregunta. Comenzamos el capítulo preguntándonos el porqué de LA, y ahora nos vamos a preguntar el porqué no: ¿qué está limitando el desarrollo de LA?

2.3 Las barreras de Learning Analytics

Hay una gran cantidad de barreras en el uso de las Learning Analytics y debemos ser capaces de conocerlas e identificarlas con el objetivo de que, en la puesta en marcha de nuestro sistema, o de cualquier otro, no nos veamos frustrados por ellas.

La primera barrera que nos podemos encontrar es la de **disponibilidad de datos**. Los datos que se pueden extraer en las actividades y entornos educativos no siempre van a ser voluminosos, sino que pueden ser pequeñas cantidades de datos las que se analicen. De todas formas, en algunos casos la barrera es no tanto que los datos educacionales sean escasos, sino más bien que sean inexistentes. El LEA's Box Project informó en un pilotaje en su campamento de verano (LEA's Box 1st Learning Analytics Summercamp, 2015) en el que los profesores estuvieron usando software de análisis para preparar informes sobre sus estudiantes, introduciendo los datos a mano ya que la red inalámbrica era insuficiente para conectar las tablets. Si este es el caso para un centro educativo entusiasta con el proyecto en el mundo desarrollado, podemos esperar que la conectividad y la infraestructura del software sean un problema mayor en circunstancias menos favorables.

Continuando en relación con la recogida de datos, otro problema que nos atañe cuando sí se dispone de ellos es la **accesibilidad**. Disponemos de los datos, pero no podemos acceder a ellos. Esto puede deberse tanto a cuestiones éticas de la institución como a leyes de privacidad y protección de datos.

Los usuarios pueden presentar cierta **resistencia personal**. Profesores, alumnos y familias pueden tener preocupaciones personales sobre la privacidad y la política de seguridad. Estas preocupaciones pueden derivar en campañas contra este sistema. Un ejemplo de esto es el fracaso del proyecto InBloom de la Gates Foundation. Este proyecto de LA contaba con una financiación de millones de dólares y fracasó al enfrentarse a una sociedad preocupada por la privacidad de los datos (Strauss, Student privacy activists win a big one, 2014) (Strauss, \$100 million Gates-funded student data project ends in failure, 2014).

Como vemos, las dos últimas barreras presentadas confluyen en lo mismo: la **ética** y la **privacidad**. Más adelante, en este mismo apartado, profundizaremos un poco más en este punto, que alcanza una gran relevancia en LA.

Los propios docentes pueden sentir un **impacto en los roles profesionales**. Los avances que se están produciendo en la disciplina de LA pueden provocar resistencia por parte de los profesionales, quienes temen una extensión del control sobre su autonomía profesional y pérdida de libertades.

Está claro que los métodos de analíticas tienen un gran potencial para generar cambios en la investigación, pedagogía y gestión educacional, y que ofrecen oportunidades empresariales. La mayoría de LA busca proporcionar vistas posibles basadas en la evidencia, de tal forma que uno pueda esperar que habrá una gran evidencia disponible para guiar a los usuarios a planear su propio uso de LA. En cambio, mientras que hay multitud de informes sobre la adopción, las evidencias sólidas acerca del impacto positivo son escasas, aunque se encuentran en aumento, como podemos encontrar en LAC Evidence Hub (<http://evidence.laceproject.eu/>). Como con muchas nuevas tecnologías, la adopción ha estado plagada de oportunidades tentadoras. Mientras que las oportunidades pueden ser reales, las **expectativas demasiado altas** y exageradas pueden derivar en un cinismo también exagerado si los beneficios prometidos no se materializan, creando una barrera para un futuro desarrollo.

Por último, uno de los mayores peligros potenciales de LA, ya avanzado previamente, es la **ética y privacidad**. De este tema específico hay una gran cantidad de investigaciones y discusiones detrás, y resultaría inabarcable tratarlo de la forma en la que merece. Desde el momento en el que LA puede determinar el comportamiento de un alumno basándose en datos incompletos o incorrectos hasta la intromisión en la privacidad de los usuarios. Un ejemplo del primero es cuando LA son usadas para recomendar una u otra carrera, especialidad o asignatura por las posibilidades de éxito en ellas y el alumno las elige en lugar de seguir opciones más retadoras y apropiadas.

Por otro lado, la privacidad de los datos está en el punto de mira de la mayoría de los países desarrollados, hasta el punto de la muy reciente entrada en vigor el pasado 25 de mayo del reglamento europeo de protección de datos (2016) más restrictiva para las organizaciones que custodian datos y que otorga más libertades a los usuarios. Esta ley puede implicar una gran cantidad de cambios en el desarrollo futuro de LA y origina la necesidad de investigaciones sobre la convivencia entre REPD y LA.

Tanto Rodríguez-Triana (2016), como Drachsler & Greller (2016) presentan en una tabla una lista de recomendaciones para tratar la privacidad en el desarrollo de analíticas de aprendizaje. Estos últimos establecen el acrónimo DELICATE:

- Determinación, tener claro la razón de uso de LA;
- Explicar, objetivos y límites del análisis, detallar cómo usaremos los datos;
- Legitimar las razones por las que tienes acceso a los datos;
- Involucrar a las partes interesadas, a los usuarios, dentro del proyecto;
- Consentimiento escrito de los sujetos de los datos con todos los puntos especificados y claros;
- Anonimizar los datos tanto como se pueda, impidiendo la identificación de las personas cuyos datos han sido recolectados;

- Técnico, tener controlado en el aspecto técnico todos los procesos para garantizar la privacidad, como monitorear quienes han accedido a los datos o que el almacenamiento cumpla los estándares de seguridad;
- Externos, si compartimos los datos o analíticas con agentes externos deberemos asegurarnos de que estos cumplen también con las medidas de seguridad.

2.4 Conclusiones

En este capítulo hemos introducido los conceptos necesarios para comprender qué es LA. Gracias a esto, podremos abordar el resto de este trabajo, en concreto lo que veremos a continuación, el diseño de una metodología de aplicación en entornos educativos.

También hemos analizado su situación actual y evaluado tanto los factores que impulsan su desarrollo, como aquellos que lo obstaculizan y ante los cuales tendrá que hacer frente. Este apartado nos servirá en el resto del trabajo para valorar en qué aspectos debemos prestar especial atención.

En base a todo ello, podemos concluir que LA constituye una herramienta de gran utilidad para la optimización del desempeño en la educación en general. Esta disciplina se encuentra en fase de investigación y desarrollo, y será una herramienta a tener en cuenta a medio plazo siempre que consiga adaptarse a las normativas de protección de datos y demás barreras.

3 Metodología propuesta

En este capítulo vamos a presentar el marco metodológico que se propone y que posteriormente se usará en el caso de estudio. Este marco metodológico se ha diseñado basándonos en toda la literatura revisada en general, siendo reestructurada tantas veces como hemos creído necesario hasta obtener un método que nos convenciese y con el que nos sintiésemos cómodos trabajando, teniendo en cuenta que la cantidad de datos disponibles no sería muy abundante. De forma más específica se ha diseñado atendiendo a la definición de LA, en la Figura 1, y a la Tabla 2.

Por un lado, la definición establece al menos cuatro fases diferentes: medición, recolección, análisis y actuación. En cambio, la propuesta de Siemens determina hasta siete etapas: recolección, almacenamiento, filtrado, integración, análisis, visualización y acción.

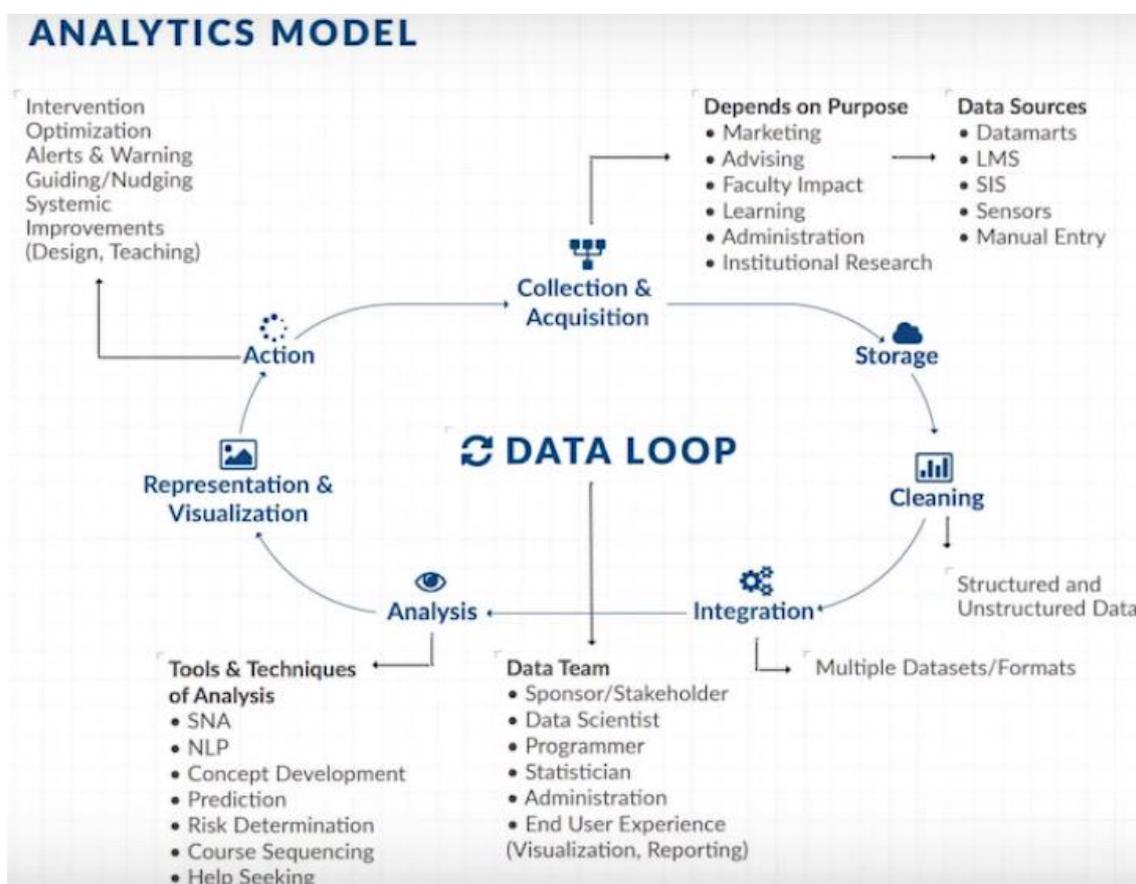


Tabla 2 (Siemens, Learning Analytics: Advancing the Science of Learning, 2015)

En la metodología diseñada proponemos en cambio seis apartados. El primero de ellos es identificar quién va a ser el **usuario** de las analíticas. En segundo lugar, es el establecimiento de **objetivos** en base a las necesidades y las medidas disponibles. A continuación, vendría la **recogida de datos** en base a esos objetivos que nos hemos marcado. En este paso veremos qué técnicas y herramientas podemos utilizar para la extracción de datos. Los dos siguientes pasos son la **analítica** y **visualización** de los datos: aprenderemos cómo podemos dar utilidad a los datos recogidos. Por último,

tendremos la **acción** motivada por los pasos anteriores y que será lo que marque la utilidad de este modelo analítico.

Podemos distinguir entre ellos una fase de planificación, que correspondería a los dos primeros y parte del tercero a la que daremos especial importancia, como veremos en los siguientes puntos.

Conviene aclarar antes de comenzar con la presentación de cada uno de los pasos, que estos no tienen por qué ser considerados de forma lineal, sino que se trata de un método flexible y adaptable a las circunstancias de cada usuario. También consideramos que es de carácter cíclico, y podemos usarlo en forma de iteraciones para mejorar los resultados.

3.1 Identificar usuarios

En este apartado vamos a definir quiénes van a ser los usuarios finales de los resultados revelados por las analíticas de aprendizaje. Históricamente los usuarios podrían ser, entre los más destacados: el gobierno, las autoridades, el centro, el departamento de orientación, los profesores, los tutores, las familias y los alumnos.

Los campos de actuación de LA pueden llegar a ser muy amplios. Previamente en este documento, en el apartado 2.1 ¿Qué son las Learning Analytics?, hemos hablado de los niveles que establece Buckingham, y de las disciplinas con las que limita LA. Ahora es cuando vamos a desarrollar ambas ideas.

Long y Siemens (2011) distinguen entre Learning Analytics y Academic Analytics, definiendo la segunda en contraste con la primera como la aplicación de inteligencia de negocio en educación, relacionando las analíticas a niveles institucionales, regionales e internacionales. De esta manera nos encontramos con una primera diferenciación, en la que Learning Analytics tendrá como objeto de las analíticas un curso o clase, identificando redes sociales, desarrollo conceptual, currículum inteligente; y a nivel departamental con modelos predictivos y patrones de éxito y fracaso. Por otro lado, Academic Analytics estará presente en los niveles institucionales, perfiles tipo de estudiantes, desempeño, flujo de conocimiento; niveles regionales con comparaciones entre sistemas; y niveles nacionales e internacionales. También hay una diferencia entre los beneficiados, es decir, los usuarios. Mientras que en LA (Learning Analytics) los principales beneficiarios serán los miembros del centro (alumnos, familias y profesores), en Academic Analytics serán los administradores, encargados de marketing, patrocinadores, autoridades educativas y gobiernos.

Por otro lado, (Buckingham Shum, 2012) defiende que LA puede dividirse en tres rangos de analíticas distintos: macro, meso y micro. Esta clasificación se asemeja a la de Siemens. En este caso, el nivel macro sería a lo que el primero llama Academic Analytics, mientras que el nivel Meso se correspondería con las LA dirigidas a la institución, al centro, y el nivel micro a un nivel bajo, de individuos y pequeños grupos. Este autor además afirma que pese a haber una distinción entre los niveles, están relacionados de tal forma que se enriquecen mutuamente: el aumento de la recogida de datos en los niveles micro proporcionaba más datos para procesar a los niveles superiores; y la amplitud y

profundidad que obtienen las analíticas macro y meso otorgan más poder de actuación a las micro.

Conviene recordar que LA comenzó a desarrollarse en entornos de educación superior y como tal, una gran cantidad de herramientas e investigaciones han sido especialmente desarrolladas para este contexto, en el que el uso de dispositivos conectados a la red es significativamente mayor que en la educación básica. Existen otras clasificaciones de enfoques y usuarios, como la propuesta por (Divjak & Vondra, 2016) específica para la educación básica. Pero nos centraremos en hablar de las dos primeras porque son las más aceptadas y relevantes, y resumen de forma simple el resto de las propuestas.

En la Tabla 3 podemos ver una comparación ilustrativa entre ambos paradigmas.

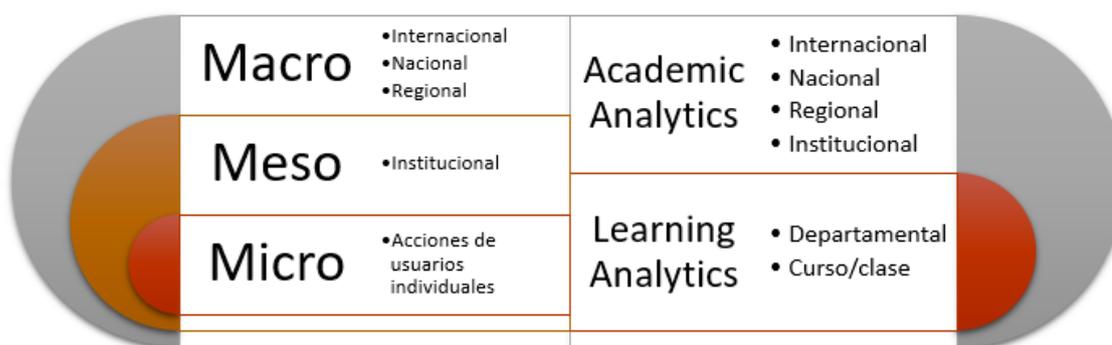


Tabla 3 Comparación entre Long y Siemens (2012) y Buckingham (2012)

En estas clasificaciones basadas en el rango de aplicación y objetivos de las LA, los usuarios son un elemento clave y diferenciador.

A la hora de comenzar a trabajar con Learning Analytics, tenemos que tener claro los usuarios objetivos, que dependerán del nivel en el que nos situemos. Vamos a comenzar identificando los usuarios desde los niveles más altos hasta los más bajos.

En el nivel más alto, de acuerdo con la clasificación de Buckingham, nos encontraremos como objeto de análisis la educación a nivel internacional, nacional y regional. En el caso específico de España cuando hablemos de nivel regional nos referiremos a las Comunidades Autónomas. En este caso, los usuarios son los gobiernos y las autoridades educativas encargadas de crear y modificar las políticas educativas.

Es por esto, por lo que cuando se habla de instituciones, resulta complejo pensar en el tratamiento de datos que puede hacer un centro de secundaria español de forma autónoma, cuando se está haciendo referencia a universidades con decenas de miles de alumnos y equipos especializados y dedicados a LA. Los usuarios de este nivel, en ambos paradigmas, son el sector encargado de administrar la institución, el marketing y los encargados de la institución como una empresa. En otras palabras, usuarios con miras administrativas y económicas.

En los niveles más bajos los usuarios directos de la visualización de los datos y sus analíticas, y su consecuente reacción, son los profesores y los alumnos, bien agrupados en clases y departamentos o bien de forma individual.

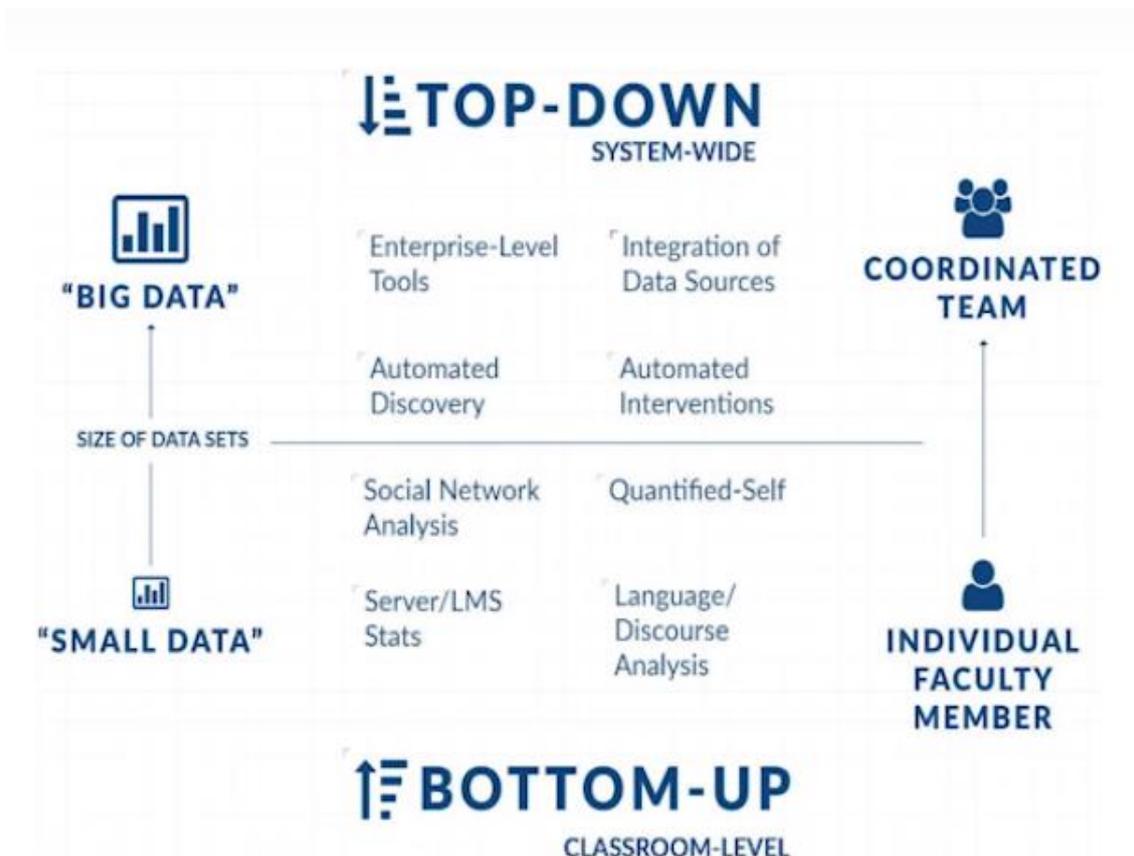


Tabla 4 (Siemens, *Learning Analytics: Advancing the Science of Learning*, 2015)

En el caso de este trabajo, vamos a dirigirnos principalmente a este último nivel. En los siguientes apartados se recogerán maneras de recogida, análisis y visualización de datos especialmente dirigidos a los estudiantes y a los docentes.

La razón de esta decisión es porque consideramos que es el enfoque más enriquecedor para el lector, acorde a los objetivos planteados. Este trabajo está pensado para que cualquier profesor independientemente de su especialidad o departamento pueda ponerlo en práctica a nivel individual, con mayor o menor ayuda si sus conocimientos sobre las TIC y estadística son escasos o amplios. Como ya reflejamos previamente, el uso de LA en primaria y secundaria será llevado a cabo por educadores con ganas de explotar esta nueva disciplina y rara vez se contará con un equipo de especialistas.

Hemos acotado en gran manera los posibles usuarios de nuestra propuesta. Ahora, en este primer paso de la metodología, debemos concretar hacia qué tipo de usuario vamos a dirigir nuestra LA. La pregunta que nos planteamos es **¿Quién va a hacer uso de las analíticas de aprendizaje?**

En el momento de responder a esta pregunta es probable que tengamos que recurrir al siguiente punto de forma simultánea. Antes de decidir quién va a ser el sujeto principal de la herramienta quiero saber qué puedo medir, de qué datos puedo disponer, y después, con esta información, establecer los usuarios y los objetivos.

3.2 Identificar necesidades y plantear objetivos

Lo siguiente que tendremos que hacer es identificar qué objetivo se pretende con el análisis de datos del aprendizaje. Para ello, tenemos que tener en cuenta las necesidades de nuestro aula, de los alumnos y del propio profesor. También tenemos que ser conscientes de qué datos disponemos y de qué datos podemos disponer.

La principal razón por la que hemos considerado este punto como un paso determinante en el método es porque el mensaje de los expertos en LA es que, si queremos evitar el fracaso, es necesario tener una visión clara de lo que queremos conseguir con LA. Esta visión tiene que ser revisada frecuentemente de tal forma que los usuarios siempre tengan claro el porqué del uso de LA (Ferguson & Clow, *Learning Analytics - Avoiding Failure*, 2017).

Las preguntas que trataremos de resolver en este paso del método parten de ¿qué necesidades tienen el aula, los alumnos y el profesor? para llegar a ¿Qué objetivo nos planteamos?

Nuevamente parece que, para poder responder estas preguntas, tendremos que avanzar un paso más al siguiente epígrafe para conocer qué datos podemos llegar a obtener para saber si tenemos la posibilidad de recolectar aquellos que necesitamos para cumplir nuestros objetivos. Pero primero veamos las necesidades de las que partimos, establezcamos unos objetivos conforme a ello, y en el caso de que no podamos recoger datos útiles para la meta que nos hemos marcado, reformulémosla.

Comencemos planteándonos la primera pregunta: ¿Qué necesidades tiene el aula, los alumnos o el profesor? Dependiendo del usuario final que hayamos escogido en el paso anterior, deberemos recoger estas necesidades de una u otra forma.

Si se trata de los alumnos o de las familias, podemos realizar encuestas cuantitativas en las que contesten el grado de relevancia de una serie de necesidades observadas (Divjak & Vondra, 2016), podemos entrevistarlos con ellos para encontrar necesidades no detectadas, o bien podemos destacar necesidades que haya detectado el profesor o el centro.

En el caso del profesor, si se trata de un usuario externo al que está diseñando la herramienta, podemos seguir la técnica anterior con el objetivo de encontrar las necesidades profesionales que tenga. En cambio, si el profesor es el mismo diseñador de la herramienta, directamente tendrá que reconocer sus necesidades a través de la observación y la experiencia.

Partiendo de las necesidades, tendremos que establecer de forma clara los objetivos específicos de la herramienta de LA. Estos objetivos han de ser concretos con lo que esperamos conseguir y habrán de dirigir el resto de nuestro plan.

Antes de continuar, vamos a hacer aquí un pequeño inciso en el que haremos una aproximación a los niveles de las analíticas de aprendizaje, que nos servirá para poder

establecer nuestros objetivos. Amo Filvà (2015), coincidiendo con el modelo de analíticas ascendentes de Gartner, clasifica los niveles de LA en cuatro:

1. **Descriptivo.** Esta fase puede responder a las preguntas ¿qué ha pasado? y ¿qué está pasando? Por lo tanto, trabaja con análisis del pasado y del presente. Es una fase inicial en la que los datos recolectados describen la realidad y su uso se limita a proporcionar visualizaciones a los usuarios ampliando sus puntos de vista, pero deja toda la labor de análisis a estos.
2. **Diagnóstico.** En este punto aparecen los análisis. Si en la fase anterior se buscaba el “qué” en el pasado y presente, aquí se trata de responder a las preguntas “cómo” y “por qué”. En esta fase el análisis de datos puede otorgarnos conclusiones legítimas para nuestra investigación.
3. **Predictor.** Responde a las preguntas ¿qué puede pasar? y ¿cómo puede pasar? En base a análisis de los datos recogidos y a través de algoritmos estadísticos predictivos, facilita identificar a los usuarios de comportamientos de riesgo no deseados, dejando la decisión y acción al propio usuario.
4. **Prescriptor.** Si en el segundo nivel se detecta el problema y en el tercero se predice su aparición, en este cuarto nivel se responde como atajarlo. Responde a las preguntas ¿cómo podemos actuar? y ¿cómo prevenir lo negativo y potenciar lo positivo? Trata de definir las actuaciones a llevar a cabo. En este nivel es la propia herramienta la que puede, de forma autónoma, intervenir en el proceso de aprendizaje, de tal forma que reduce al mínimo el trabajo de análisis y decisión del docente.

La dificultad de aplicación de las técnicas de análisis crece con el nivel en el que nos situemos. Conocer esta clasificación es un buen punto de partida para tratar de establecer nuestros objetivos y conocer la ambición del proyecto. Si somos poco expertos en LA, deberemos centrarnos en desarrollar objetivos descriptivos o diagnósticos, y según crezca nuestros conocimientos sobre LA y estadística, podremos realizar modelos predictivos.

También es reseñable que, para el desarrollo de herramientas de niveles superiores, en particular para el prescriptivo, se hace necesario un equipo coordinado de expertos en análisis de datos y pedagogía, ya que implica un conocimiento profundo de los datos y de las analíticas para el diseño de modelos educativos.

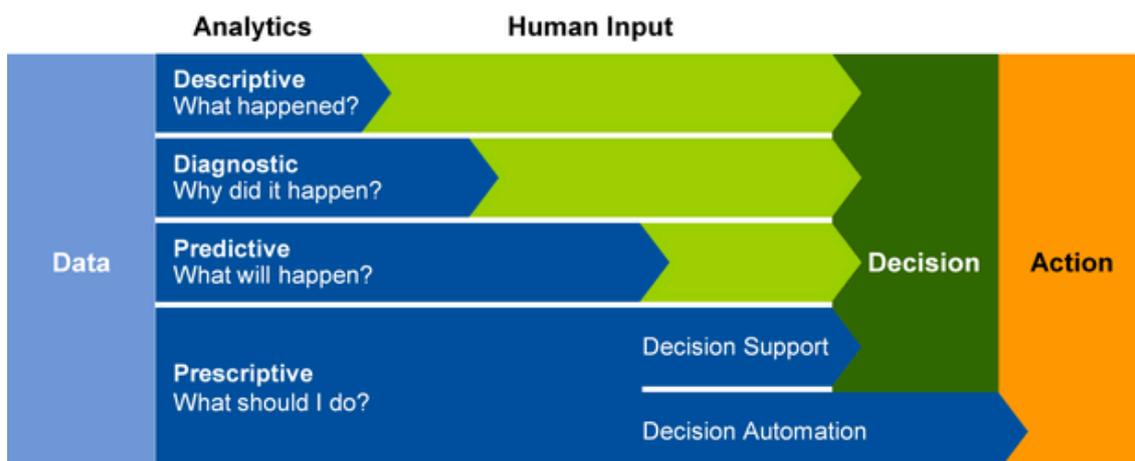


Tabla 5 Fuente: Gartner (octubre 2014)

Una vez conocemos esta clasificación de los niveles de LA, vamos a volver al estudio de las necesidades de los usuarios y posterior establecimiento de los objetivos.

Comenzando con las necesidades, Divjak & Vondra (2016) a través de encuestas cuantitativas, extraen información sobre las necesidades de los profesores y los alumnos en el entorno de la educación básica en Croacia con el objetivo de crear una *dashboard* informativa. Como resultado de las encuestas, obtuvieron que a los profesores les gustaría usar la información de su *dashboard* para aprender sobre:

- Su desempeño (resultado de encuestas, calificaciones de sus estudiantes, visión histórica).
- Quiénes son los alumnos más virtuosos y aquellos que necesitan apoyo adicional.
- Actividades de aprendizaje de los estudiantes.
- Si los estudiantes continúan en la educación.

Mientras que a los alumnos les gustaría:

- Comparar las competencias conseguidas con las demandadas en el mercado laboral.
- Comparar sus logros con los necesarios para entrar en las carreras de su interés.
- El desempeño de los profesores.
- Conocer los riesgos que puedan tener para conseguir sus metas
- Cómo pueden mejorar su desempeño.

A continuación, vamos a hacer un breve repaso de los objetivos generales sobre los que distintos autores han escrito.

Verbert, Manouselis, Drachsler y Duval (2012) distinguen seis objetivos muy generales:

- Predecir el desempeño del estudiante y perfilar/modelar estudiantes tipos.
- Proponer recursos de aprendizaje relevantes.
- Incrementar la reflexión y la conciencia.
- Mejorar entornos sociales de aprendizaje
- Detectar comportamientos indeseables de los estudiantes
- Detectar efectos de los estudiantes

Dyckhoff (2013), en su artículo presentado en la conferencia LAK13, diferencia hasta 16 objetivos distintos para los que usar Learning Analytics, clasificados entre profesores y alumnos.

Conocer todos estos objetivos y necesidades puede resultarnos muy útil a la hora de concretar los nuestros, pero muchos de ellos, pueden ser demasiado generales y ambiciosos, y necesitarían de un equipo experto para poder ser llevados a cabo. El docente promedio que desee adentrarse en las LA, deberá ser mucho más conciso a la hora de establecer sus objetivos y valorar los recursos necesarios para la puesta en marcha con la ambición del proyecto.

3.3 Recogida de datos

La metodología comienza por la definición clara de los usuarios, necesidades y objetivos. A continuación, vamos a tratar la recogida de datos: qué tipos de datos podemos obtener y un pequeño compendio de herramientas que podemos utilizar para realizar su adquisición.

Lo primero, será especificar las métricas representativas para que el análisis de los datos recogidos proporcione resultados concluyentes. Las métricas son las medidas que nos proporcionarán datos. Por otro lado, cada uno de estos datos recolectados se denomina “dato crudo”.

Un dato crudo representa la unidad mínima de información recolectada y sin procesar de un dato específico (Amo Filvà, Edulíticas, 2015).

Muchas veces los analistas fuerzan los datos con algoritmos y técnicas con tal de obtener el resultado esperado, generando informes que concluyen en esos resultados deseados. Por esto, los datos crudos son de vital importancia porque representan la realidad tal y como es, no están procesados. Si vamos a almacenar algo de nuestra analítica de datos, deberían ser los datos crudos. Al no estar procesados, nos ahorramos los errores interpretativos que podamos hacer y con ello conclusiones erróneas y si en un futuro queremos nuevamente un informe, sólo tenemos que generarlo otra vez.

Ahora nos planteamos dos nuevas preguntas ¿qué tipos de datos crudos vamos a recoger? y ¿cómo los vamos a obtener?

Los dos grandes tipos de datos que tradicionalmente se contemplan en las disciplinas que trabajan con estos son los cualitativos y los cuantitativos.

Los datos cualitativos son aquellos que no son numéricos. Documentos entregados, imágenes compartidas, conversaciones entre alumnos y profesores o el contenido de un correo son algunos de los ejemplos que nos podemos encontrar.

Los datos cuantitativos, sin embargo, son numéricos. Pueden ser ordenados y utilizarse directamente en cálculos matemáticos. Algunos de los más comunes son: número de interacciones con compañeros o profesores, cantidad de veces que accede a un recurso o ve un vídeo, el tiempo dedicado de un alumno en el entorno virtual o las calificaciones obtenidas en un examen.

La cantidad de datos que podemos recolectar es amplísima, y por eso es de vital importancia establecer bien los objetivos y dejar claro qué queremos medir, qué queremos obtener de los datos. Pongamos un ejemplo ilustrativo:

En nuestra clase hemos abogado por complementar el aprendizaje en el aula a través de visionado de vídeos en casa. Nos marcamos como meta el conocer en qué grado resultan interesantes para nuestros alumnos esta serie de vídeos.

Algunos de los datos que podemos recoger son el número de veces que se accede al recurso a través del entorno virtual o los minutos de visualización. Estos datos por sí solos pueden darnos una visión descriptiva de la realidad: han accedido de media

1.14 veces a cada vídeo; un 19% de los alumnos no ha entrado a ningún recurso; se han visualizado 40 minutos de media; el vídeo con más visualizaciones ha sido este.

Sin lugar a duda, esto puede ayudar al profesor a sacar sus propias conclusiones y actuar en consecuencia, cambiando los vídeos con menos visualizaciones por otros, por ejemplo. Pero la analítica del aprendizaje tiene la capacidad de ir más allá. Con este mismo ejemplo, podemos recoger también las calificaciones obtenidas por los alumnos en las actividades y exámenes de la asignatura que tengan relación directa con la temática de los vídeos, y cruzarlo con el tiempo de visualización de estos.

No nos vamos a entrometer ahora en cómo establecer una hipótesis y contrastarla, pues corresponde a las técnicas de análisis que veremos en el próximo capítulo. En cambio, lo que queríamos ilustrar es la amplia cantidad de datos que podemos recoger y cómo poder hacer uso de LA.

A continuación, vamos a proponer una serie de herramientas TIC con la que recoger datos y el tipo de datos que estas nos van a proporcionar. Muchas de estas herramientas nos proporcionan ya posibilidades de visualización y análisis de datos, facilitando mucho la tarea del docente, pero también limitándola a su oferta. La virtud, por lo tanto, la encontraremos en la combinación de las herramientas en aras de la consecución de nuestros objetivos.

No vamos a hacer un análisis exhaustivo de estas herramientas, ya que eso nos podría llevar cientos de páginas. Tampoco vamos a realizar una guía de cómo extraer los datos de cada una, simplemente vamos a nombrar algunas de las más usadas y qué tipos de datos se pueden extraer de ellas.

Por supuesto, hay muchas más de las que aquí se proponen, y con el tiempo aparecerán nuevas y con mejores funcionalidades. El objetivo que tenemos es ilustrar al lector acerca de algunas de las posibilidades actuales.

La primera herramienta de la que debemos hablar es de los entornos virtuales de aprendizaje (o sistemas de gestión de aprendizaje). Estos recogen enormes cantidades de datos de la actividad de los usuarios. A día de hoy, estos sistemas suelen incluir ya integrados en su plataforma dashboards o paneles de control con información sobre los datos extraídos que puede sernos de gran utilidad. En el caso de Moodle, podemos añadir dashboards a través de plugins, por lo que no estamos limitados a una única herramienta. Por otro lado, algunos VLE como Moodle permiten descargar directamente logs con la actividad de los alumnos, datos crudos que podemos procesar posteriormente nosotros a voluntad.

Google ofrece herramientas para complementar nuestros cursos o clases. Cabe destacar el uso de Google Classroom (GC), que no se considera un LMS. El principal problema es que GC no proporciona acceso a logs o bases de datos para ser analizados por terceros. Sólo permite el acceso a un número reducido de datos si GC está instalado como una parte de Google Suit, y este acceso sólo se encuentra disponible para los

usuarios administradores y en forma de interfaz, sin posibilidad de acceder a los datos crudos (Martínez-Monés, Reffay, Hoyos Torío, & Muñoz Cristobal, 2017).

En su artículo, Martínez-Monés et al. (2017) ofrecen alternativas para la recolección de datos de uso de GC.

Los formularios de Google, utilizados en educación para realizar actividades y pruebas evaluativas, pueden ser una fuente más de datos, tanto cualitativos como cuantitativos. A parte de las respuestas obtenidas en dichos formularios, también podemos extraer el momento temporal en el que se realizaron.

Actualmente el uso de plataformas TICs en el aula para comprobar y evaluar el aprendizaje de los alumnos está en continuo crecimiento. Las herramientas que presentamos a continuación son ampliamente usadas por su componente lúdico, resultando atractivo para los alumnos.

Kahoot, Plickers, Socrative y Quizizz ofrecen la posibilidad de descargar los datos crudos en formatos .csv o .xlsx, junto a informes propios de la herramienta. Todas ellas ofrecen visualización de los datos bastante sencilla en sus propias interfaces.

La última aplicación que queremos nombrar es Edpuzzle, una herramienta que nos permite multiplicar las posibilidades del uso de vídeos en nuestras aulas, tanto de creación propia, como de plataformas como YouTube o Vimeo. Edpuzzle nos permite “trocear” los vídeos y añadir preguntas entre cada uno de esos trozos. De esta forma, una vez se visualice un trozo, se pausará y se pedirá al espectador que responda la pregunta. Hasta aquí, en relación con los datos que podemos recoger, no hay distinción con las herramientas anteriores. La diferencia está en que Edpuzzle además nos proporciona datos de visualización de vídeos: qué usuarios han visto qué vídeos, qué vídeos tienen más reproducciones, las veces que un usuario reproduce un vídeo determinado, e incluso las veces que se visualiza cada una de las secciones de vídeo. Esto otorga al docente nuevas posibilidades para monitorizar el aprendizaje de los estudiantes y para obtener datos de reproducciones de vídeos.

En el capítulo 22 del Handbook of Learning Analytics (Lang, Siemens, Wise, & Gašević, 2017), páginas 251-267, Mirriahi y Vigentini presentan una tesis muy interesante sobre el uso de vídeos en la educación y el análisis de los datos que estos proporcionan.

Por último, en este apartado, vamos a tratar sobre el almacenamiento de los datos. Si se tratase de un método pensado para aplicarlo usando técnicas de Big Data, sin duda tendríamos que dedicarle un epígrafe, pero, como no es el caso, bastará con dedicarle este pequeño apartado.

Almacenar los datos es una acción muy interesante para el desarrollo de LA. Con ella podremos estudiar históricos de datos y se vuelve necesaria para los objetivos más ambiciosos de prevención y prescripción. Pero, aunque recomendable, no siempre resulta necesaria. En el ejemplo que comentamos previamente en este epígrafe, el almacenamiento de datos no es necesario una vez se finalice el estudio. Aun así, seguiría

siendo recomendable ya que en un futuro podríamos replicar el estudio o utilizar sus datos.

Como ya justificamos antes, lo que deberíamos almacenar no serían los resultados de las analíticas, los informes o las conclusiones, sino, prioritariamente, los datos crudos.

En este proceso de almacenaje hemos de tener especial cuidado con la privacidad de los datos. La tendencia actual se dirige a políticas que ofrecen más libertad a los usuarios como propietarios de los datos, en las que estos han de autorizar el uso de sus datos y pueden, en cualquier momento, solicitar su eliminación completa. Bajo este contexto aparece la necesidad de nuevas investigaciones que medien en este conflicto aportando soluciones. Estas pueden pasar por des-identificar los datos o anonimizarlos en cuanto su uso ya no esté dirigido específicamente al alumno en cuestión, sino a la generación de modelos y búsqueda de patrones.

3.4 Analizar

El proceso de análisis de los datos es el paso determinante de nuestra metodología. Haciendo una analogía con la cocina: la identificación y selección de los usuarios y objetivos se correspondería con la selección de la receta para el gusto de los comensales, y marcará el resto de nuestros pasos; la recogida de datos con la adquisición, la compra de los ingredientes; el análisis con la propia acción de cocinar los alimentos; la visualización con el emplatado, la comida entra por los ojos; y por último actuar, el fin último del arte culinario, comer.

Ahora es el momento de cocinar nuestros datos crudos. Bien es cierto que en la cocina podemos hacer deliciosas ensaladas sin procesar ningún alimento, y lo mismo ocurre cuando trabajamos con datos. Podemos obtener conclusiones con las que actuar sin realizar ningún tipo de análisis estadístico de los datos. Sólo a través de su visualización podemos saber, por ejemplo, qué conceptos asimilan nuestros alumnos fácilmente y con cuáles tienen más dificultades, y en base a ello, establecer un programa de refuerzo.

Pero el realizar análisis nos permite ir más allá, averiguar el por qué. Nos permite escanear datos para descubrir patrones que ocurren en pequeñas cantidades de estudiantes o esporádicamente; investigar cómo diferentes alumnos eligen usar distintos recursos educativos y obtienen distintos resultados; realizar análisis detallados de fenómenos que ocurren en largos periodos de tiempo; y analizar cómo el diseño de entornos de aprendizaje puede impactar en variables de interés en el estudio de gran variedad de casos (Baker & Siemens, 2013).

En su artículo, Baker y Siemens (2013) exponen un compendio de métodos y herramientas de análisis, clasificadas en modelos: de predicción, descubridores de estructuras, minería de relaciones, destilación de datos para juicio humano (visualización) y descubrimiento con modelos. No vamos a entrar a detallar cada uno de estos modelos, métodos, técnicas y herramientas que proponen, sino sólo aquellos que consideramos más útiles y sencillos, aquellos de los que hay evidencia de su uso, proporcionando una descripción y casos de uso. Con este trabajo queremos democratizar la analítica de datos.

Queremos que esta sea accesible a cualquier docente y no se limite a aquellos con amplia formación en estadística o análisis de datos.

Si por el contrario, el lector quiere especializarse en el análisis de datos educativos, recomendamos encarecidamente los títulos Handbook of Educational Data Mining (Romero, Ventura, Pechenizkiy, & Baker, 2011) y, en especial, Handbook of Learning Analytics (Lang, Siemens, Wise, & Gašević, 2017), con los que se encontrará con largos manuales desarrollados por los mayores expertos de cada disciplina.

Siguiendo el artículo de Baker y Siemens (2013), vamos a hacer una primera aproximación a algunas técnicas de predicción, de descubridores de estructuras y minería de relaciones.

3.4.1 Métodos predictivos

En los **métodos predictivos** se pretende desarrollar un modelo que infiera una única variable a predecir a partir de otro conjunto de datos conocido, las variables predictivas. La variable a predecir es preestablecida por los diseñadores del modelo y se buscan las variables predictivas que la infieran. Los métodos predictivos más comunes son los modelos clasificadores y las regresiones lineales.

Normalmente, en el proceso de desarrollo del modelo predictivo, este se diseña para un conjunto pequeño de datos pasados, del que ya se conoce la variable a predecir. Algunas técnicas comunes de este tipo de diseños son: los árboles de decisiones, random forest, reglas de decisión, regresión por pasos, regresión logística para variables binarias o categóricas (que pueden tener un número limitado de valores); y la regresión lineal cuando la variable a predecir es continua.

Estos métodos asumen que las relaciones entre las variables que se han dado en el pasado y el presente van a continuar existiendo en el futuro, hecho difícilmente sostenible en la práctica en la que, por ejemplo, un cambio de profesor, de la metodología usada, o del currículo oficial de la asignatura puede dar un giro a esas relaciones e invalidar el sistema.

Una vez se ha diseñado el modelo, se valida usándolo en otro conjunto de datos del cual también conocemos la variable predicha, comparando los resultados del modelo con la realidad y comprobando su exactitud. Tras realizar este proceso de modelado y validación podríamos dar por concluido el diseño de la herramienta y comenzar su uso para predecir a partir de datos nuevos.

Un ejemplo sobre el proceso de diseño de un modelo predictivo lo encontramos en Aguiar et al. (2015), en el que se presenta el diseño de un predictor a partir de datasets provenientes de 11000 alumnos estadounidenses que se esperaba que terminasen el equivalente a bachillerato (high school), en 2013.

En este experimento se recogieron un conjunto de 15 variables cuantitativas y cualitativas entre las que se encontraban, entre otras: mes y año de nacimiento, nombre del centro educativo y un conjunto de calificaciones de distintas entidades evaluadoras.

Su objetivo era alertar de estudiantes en riesgo de no graduarse en el año que les correspondía.

Se diseñaron varios modelos de aprendizaje automático para comprobar cual de ellos funcionaba mejor, usando datos históricos de los alumnos, no únicamente los del último curso. Los resultados obtenidos fueron de predicciones de hasta el 75% de acierto usando técnicas de random forest y regresión logística, frente al 38% que se venía obteniendo en el pasado con modelos anteriores, evidenciando la utilidad de estas nuevas aproximaciones.

Xing et al. (2016) detallan el proceso de puesta en marcha de un proyecto de LA que utiliza analíticas predictivas con técnicas clasificadoras como el árbol de decisiones. El objetivo del proyecto es predecir el riesgo de abandono en MOOCs para que el profesor pueda actuar y prevenir.

3.4.2 Descubridores de estructura

Los algoritmos **descubridores de estructura** tratan de encontrar una estructura en los datos sin partir de ningún guion preestablecido, a diferencia de en la predicción donde elegimos la variable que deseamos predecir. En este caso el objetivo es determinar estructuras que emergen naturalmente de los propios datos. Las dos técnicas que veremos son clustering y Social Network Analysis (SNA).

El objetivo principal del clustering es agrupar los datos en grupos lógicos y naturales. Esta técnica podemos usarla, por ejemplo, para generar grupos de estudiantes a partir de ciertas variables, o bien clasificar sus acciones en distintas agrupaciones.

En el análisis de redes sociales (SNA), los modelos son desarrollados a partir de las relaciones e interacciones de los sujetos y los patrones que se generan en estas interacciones. En este análisis los usuarios se representan por nodos y las interacciones por uniones entre los nodos. Es interesante para detectar la efectividad de los grupos de trabajo, los tipos de comunicación, y cómo puede esto afectar al rendimiento académico. Además, también puede servirnos para identificar clusters (agrupaciones) naturales entre los usuarios.

Esta técnica ha sido estudiada e investigada en multitud de ocasiones en las últimas décadas, siendo su origen previo al desarrollo de LA. El pero principal de este tipo de análisis es que está especialmente desarrollado para comunidades de alumnos online donde todas las interacciones se realizan a través de la red, dejando un rastro de datos crudos analizables. Esto supone que en una clase convencional de educación secundaria esta técnica tenga que ser descartada debido a este déficit de interacciones digitales.

En Almeda et al. (2014), se realiza un proceso analítico usando la técnica de clustering para el diseño visual de las clases en 30 colegios estadounidenses con el objetivo de encontrar los tipos de diseños que más positivos resultan para el aprendizaje.

Por otro lado, Abdous, Wu y Yen (2012) usan la técnica de clustering sobre 298 alumnos de una universidad estadounidense que ofrecía cursos por vídeo streaming. En

estos cursos clasificó a los alumnos por las características comunes de sus preguntas en las interacciones alumno-profesor y entre pares. En este caso las agrupaciones formadas fueron de una complejidad más alta que en el estudio anterior ya que tenían agrupaciones con estructura jerárquica, multi-nivel.

3.4.3 Minería de relaciones

El último de los grandes modelos analíticos que vamos a exponer es la **minería de las relaciones**. Para el desarrollo de las técnicas basadas en la minería de datos, se precisa de una gran cantidad de variables para obtener resultados válidos. El objetivo que aquí se plantea es la búsqueda de relaciones entre variables de un dataset.

Las técnicas más destacadas son minería de la regla de asociación y de correlaciones. La primera busca reglas condicionales entre las variables (si X, entonces Y). La segunda suele buscar correlaciones lineales entre variables, la dependencia de una respecto a otra u otras. Dos ejemplos del uso de las correlaciones, acompañados de otras técnicas, los podemos ver en los estudios de Macfayden y Dawson (2010) y en Romero-Zaldivar et al. (2012).

3.5 Visualizar

La visualización hace que información compleja sea más sencilla de comprender, así como que las analíticas sean accesibles para la gran mayoría de los usuarios, ya que no es necesario ser un experto en análisis de datos para poder hacer uso de estas.

De esta forma, la visualización de los datos sin un análisis previa también puede ser Learning Analytics. La diferencia principal entre el apartado anterior y este está en el sujeto. En el primer caso es un algoritmo el que realiza el análisis y le ofrece los resultados al usuario. En las visualizaciones de datos sin análisis computacional previo, es el propio usuario el que realiza el análisis bajo su propio juicio.

A continuación, vamos a mostrar los gráficos más comunes y su uso en educación y en LA: los diagramas de barras, histogramas, gráfico de caja y bigotes y mapas de color.

Los diagramas de barras se utilizan para representar la distribución de los datos para variables categóricas, cuyos valores son categorías excluyentes. Con estos gráficos podemos representar una sola variable o usarlo para representar distribuciones bidimensionales (Tabla 6).

Son fáciles de generar a partir de una hoja de cálculo y de un solo vistazo podemos hacer comparaciones de forma muy visual, pero están limitados a pocas categorías para que las comparaciones sean útiles.

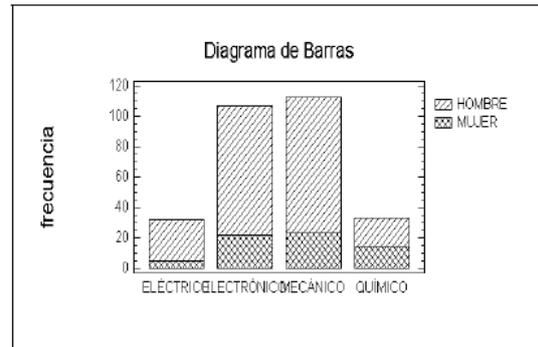
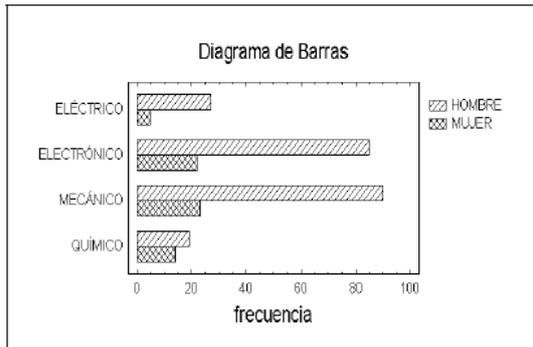


Tabla 6 Diagramas de barras bidimensionales

Los histogramas se diferencian de los diagramas de barras en que representan variables cuantitativas continuas, agrupadas en intervalos. De un solo vistazo podemos comprender la dispersión, las frecuencias relativas, y la tendencia.

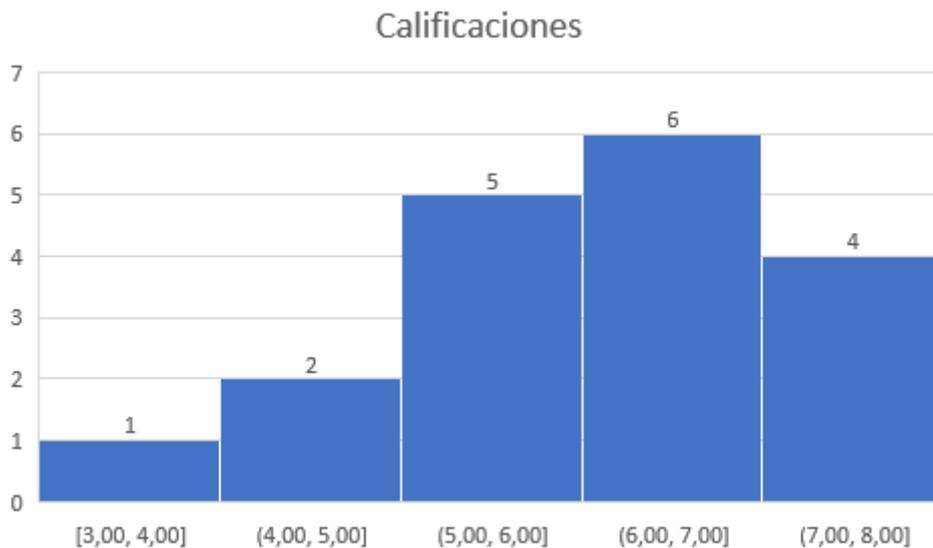


Tabla 7 Histograma de calificaciones

El diagrama de caja y bigotes es otro gráfico muy común para visualizar la distribución de los datos y sus características principales: asimetría, apuntamiento (concentración de las variables alrededor de la media), dispersión, los tres cuartiles Q1, Q2 y Q3 (los valores que dividen la distribución en cuatro partes de igual frecuencia) y la mediana.

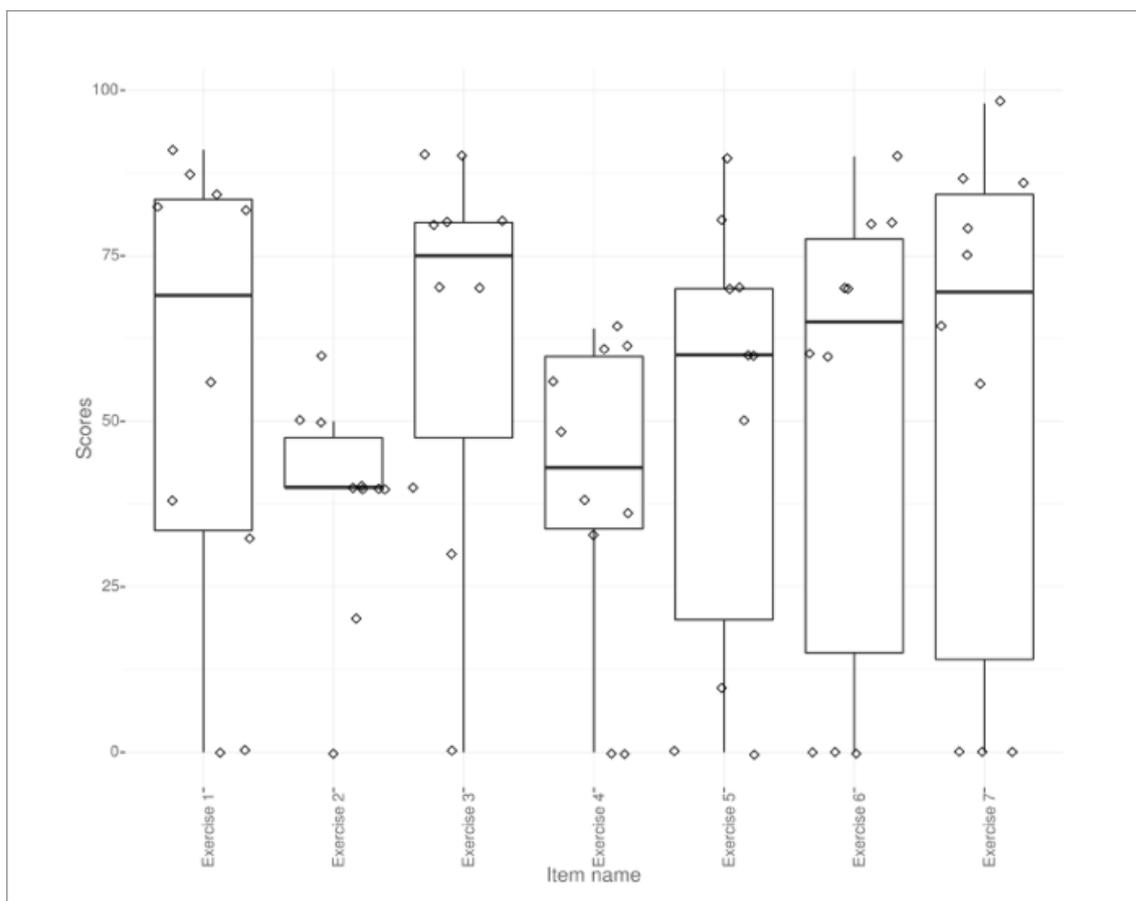


Tabla 8 Caja y bigotes para un conjunto de ejercicios (Fuente: https://help.blackboard.com/es-es/Moodlerooms/Teacher/Track_Progress/X-Ray_Learning_Analytics/Gradebook_Report)

En la Tabla 8 podemos observar las distribuciones de las calificaciones para 7 ejercicios distintos. Analizando este ejemplo podemos detectar para qué ejercicios han tenido más problemas los estudiantes, la tendencia y la dispersión de las calificaciones.

Para la visualización directa de los datos en tabla, cuando se trata de cantidades de datos de pequeño tamaño como las que se pueden disponer en un aula tradicional de secundaria, podemos hacer uso de colores con el fin de resaltar y encontrar datos de forma más cómoda y rápida. Los listados de calificaciones, por ejemplo, pueden ir complementados por un mapa de calor.

Nos encontramos un ejemplo de mapa de calor en la

Tabla 9, donde se presentan un conjunto de datos reducido que gracias a los colores ayuda a la visualización rápida por parte del docente, o del propio alumno, de los resultados obtenidos.

El último tipo de gráfico que vamos a proponer es el gráfico de dispersión, este diagrama representa en coordenadas cartesianas los valores de dos variables, en un eje cada una de ellas. Este gráfico nos permite inferir correlaciones entre variables sin necesidad de realizar el análisis, o como complemento a este.

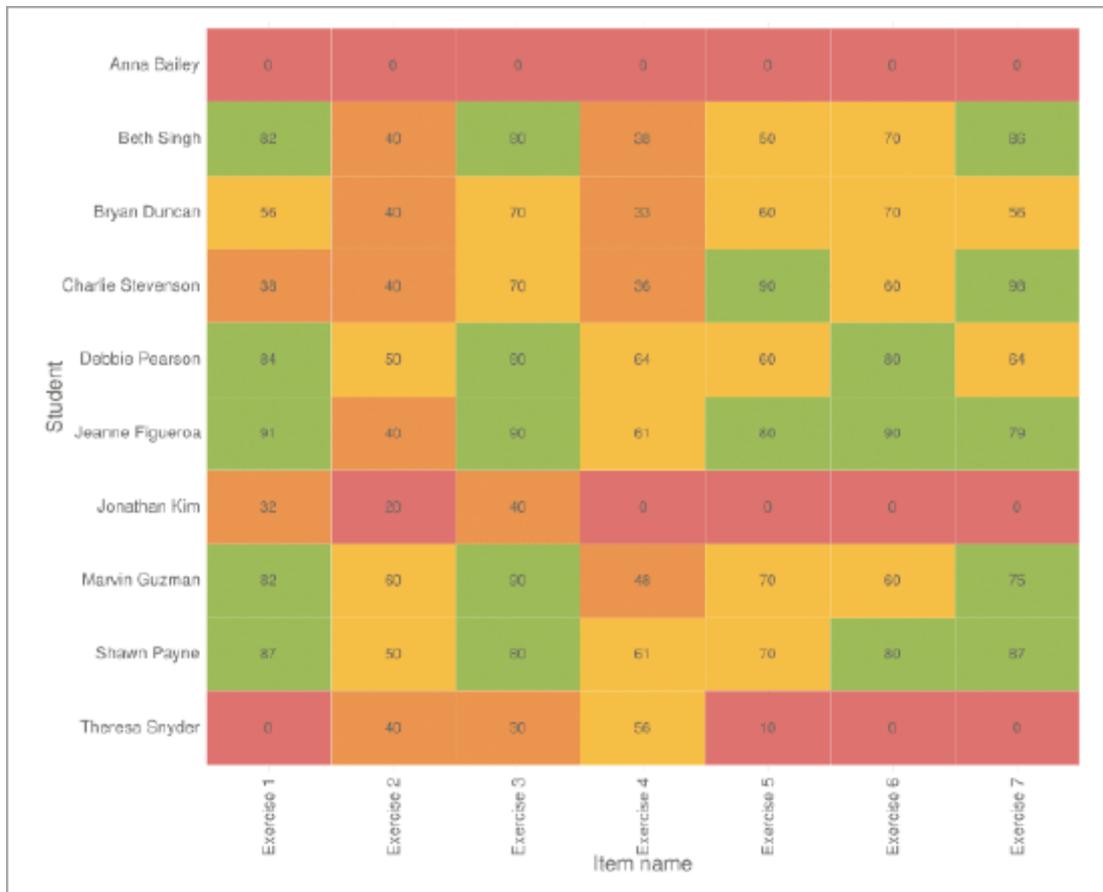


Tabla 9 Mapa de calor (Fuente: https://help.blackboard.com/es-es/Moodlerooms/Teacher/Track_Progress/X-Ray_Learning_Analytics/Gradebook_Report)

En la Tabla 10 podemos observar tres gráficos de dispersión y la covarianza entre las dos variables. Cabe destacar que en la tercera tabla la covarianza es casi nula, sin embargo, podemos observar que existe una dependencia exponencial entre las variables.

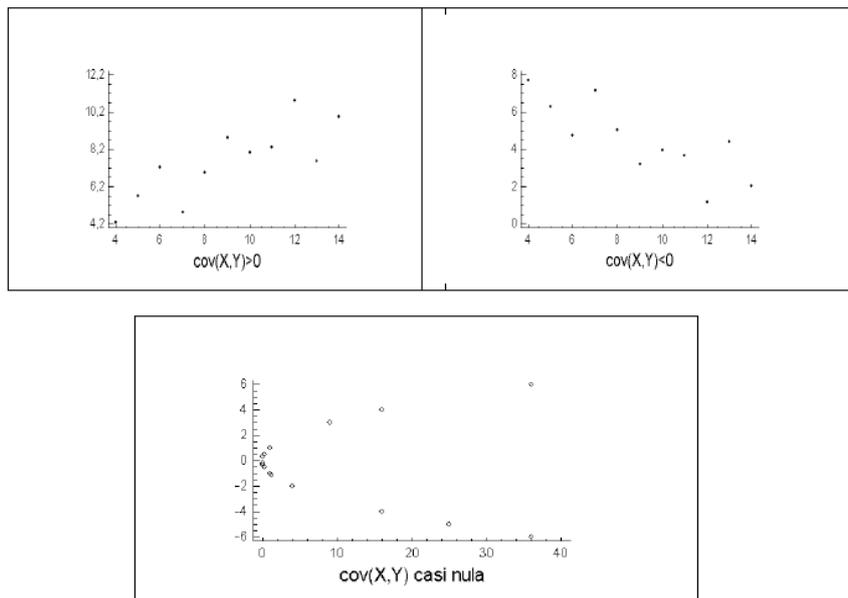


Tabla 10 Gráficos de dispersión y covarianza entre las variables

Conocer y saber cuándo usar cada uno de esta serie de gráficos es un arte que ayudará a los usuarios a visualizar e interpretar la realidad y a poder actuar en consecuencia. Se trata de la interfaz de comunicación entre los datos y las personas, ese es su sentido y ahí reside su importancia.

3.6 Actuar

El último punto es hacer uso del análisis realizado. Este análisis y reflexión de los datos, tiene que derivar en una acción o serie de acciones por parte del alumno y/o del profesor puesto que sino la utilidad de todo el proceso se vería comprometida.

La acción debe de ser consecuente con los objetivos marcados en los primeros pasos y con los análisis realizados.

En EDM la metodología se encauza hacia una actuación automática por parte del sistema: si un estudiante está siguiendo patrones de fracaso (modelos predictivos), se habilita automáticamente una hoja de ruta con actividades de refuerzo para este alumno; si los alumnos están agrupados (clustering) por técnicas de aprendizaje (visual, auditivo o kinestésico) se activan distintos recursos que faciliten su aprendizaje y retención, aprovechando las fortalezas de cada individuo.

Pese a que los métodos automáticos resultan muy interesantes y simplifican las tomas de decisiones y actuaciones de los usuarios, conllevan un gran trabajo en los primeros pasos de su elaboración y en un contexto como la educación básica en España, en la que

la duración media de las leyes educativas es inferior a los siete años, tendremos un plus de trabajo en la adaptación.

En cambio, en LA es el usuario el que decide la acción. El profesor decide en base a los análisis si unas actividades u otras son más beneficiosas para su clase y en consecuencia cambiarlas o reforzarlas; el alumno reflexiona, se concienza y tiene la capacidad de actuar en consecuencia, visualizando sus resultados, comparándolos con la media de la clase o viendo su progresión.

Cuando el usuario final de LA es el alumno, el sentido de LA es el de motivarle a la auto-reflexión. Sus acciones irán guiadas y serán consecuencia de la visualización de las analíticas. Servirán como feedback a su comportamiento. Hasta la fecha las retroalimentaciones básicas que reciben los alumnos de secundaria son calificaciones por sus exámenes, tareas y trabajos. LA puede ser un propulsor para aumentar el autoconocimiento de los alumnos hacia si mismos, su aprendizaje, sus conocimientos y su comportamiento, incidiendo positivamente en ellos, promoviendo acciones motivadas a la mejora.

Por otro lado, el profesor como profesional de la educación, deberá establecer un plan de acciones conforme a los resultados de las analíticas; impulsando un tipo de metodologías o actividades frente a otras; reformulando la programación y unidades didácticas; diseñando contextos y entornos educativos más apropiados; estableciendo formas de trabajo y de relacionarse más eficaces; motivando a los alumnos a la reflexión individual; al fin y al cabo, respondiendo al objetivo principal que se haya marcado el docente.

En el caso de que el objetivo y por lo tanto la analítica fuese descriptiva de la realidad, como por ejemplo “clasificar el gusto de los alumnos hacia los distintos contenidos”, el final de la metodología no puede quedarse ahí, no debe detenerse una vez hayamos descubierto esos gustos, sino que debemos explotar ese nuevo conocimiento potenciando las fortalezas de los alumnos.

El mejor momento para llevar a cabo la actuación no es al comienzo del curso siguiente o en la próxima Unidad Didáctica. El docente deberá actuar tan pronto como se tengan evidencias suficientes.

Un punto clave del método que aquí se propone y del que no hemos hablado es la evaluación. Todos los métodos necesitan feedback para mejorar en cada paso e iteración, y este no es una excepción. Idealmente, deberemos evaluar en cada paso que hagamos todo lo que hayamos avanzado hasta ese momento. En el establecimiento de objetivos, deberemos preguntarnos si realmente esos objetivos responden a las necesidades de los usuarios. Después, si los datos que estamos recogiendo y que vamos a analizar nos van a ayudar a lograr los objetivos. Si esto no sucede, tendríamos que reconsiderar si hemos planteado correctamente lo que queremos medir y analizar, o si los objetivos son suficientemente claros y realistas para LA. Lo mismo ocurriría con la visualización y por último con la actuación.

A lo que queremos llegar es a la gran importancia que tiene para el logro de los objetivos evaluar cada paso. Una vez desarrollemos todo nuestro proyecto, debemos entrar a valorar si ha funcionado o no, y de no ser así, encontrar el fallo para que, en la siguiente iteración, en el siguiente ciclo, podamos mejorarlo y cumplir con nuestras metas. Es común evaluar sólo al finalizar el ciclo, pero de esa manera podemos caer en el error de dar por buenas las conclusiones de la analítica y cargando el peso de la equivocación en el fallo humano de la acción.

Por eso, durante todo el proceso deberíamos ser capaces de responder a preguntas como ¿hemos sido coherentes en los objetivos con las necesidades y con los datos disponibles? ¿hemos tenido en cuenta las barreras en el desarrollo de la propuesta? ¿Tienen validez los datos? ¿El análisis y la visualización dan respuesta a las preocupaciones iniciales, u ofrecen un punto de partida para la reflexión? ¿Las acciones que este análisis ha sugerido han sido útiles?

3.7 Conclusiones

En este capítulo, hemos propuesto una metodología de aplicación de LA, basada en seis etapas, de las que se dan unas breves pinceladas a continuación:

1. **Identificar usuarios.** En esta fase se orienta todo el proyecto de LA. Seleccionar los usuarios determinará todos los demás pasos, desde los objetivos a la recolección y análisis de datos. Si los usuarios son todos los estudiantes de una universidad, la infraestructura material y humana necesaria para llevarse a cabo será muy diferente de si se tratase de un aula de secundaria. Por lo tanto, es vital tener claro el punto de partida.
2. **Identificar necesidades y plantear objetivos.** El establecimiento de objetivos claros y concisos a partir de las necesidades que presenten los distintos actores del contexto educativo es una fase clave dentro de la metodología. El buen cumplimiento de este punto facilitará el desarrollo de los siguientes.
3. **Recogida de datos.** En esta fase podemos diferenciar dos problemáticas distintas: el establecimiento de métricas, es decir, qué mediciones vamos a querer realizar, qué datos brutos deseamos obtener; por otro lado, cómo vamos a obtener esos datos, qué herramientas de recolección vamos a usar.
4. **Analizar.** El análisis de datos en LA es la parte central de nuestra metodología, sobre lo que todo gira. El problema es la gran cantidad de técnicas y herramientas disponibles en este proceso. Por ello es primordial llegar a este punto con una planificación clara sobre qué es exactamente lo que queremos hacer con los datos: qué variables deseamos predecir, qué agrupaciones queremos formar, o qué relaciones busquemos, por ejemplo. En definitiva, los objetivos de nuestro proyecto con LA.
5. **Visualizar.** La visualización es la forma en la que LA se comunica con los usuarios, transmitiendo la información deseada en esquemas y gráficas sencillas de entender. A veces, cuando el objetivo es descriptivo, podemos realizar análisis de los usuarios a través de una visualización de los datos. Es esta etapa la que más capacidad de empoderamiento otorga a los usuarios. En

esta memoria hemos descrito algunas gráficas comúnmente usadas en estadística para la visualización de los datos.

6. **Actuar.** Finalmente concluimos nuestra metodología con la acción motivada por el análisis y la visualización. Esta fase cierra el ciclo de LA, logrando, o no, los objetivos propuestos en mayor o menor medida, propone una reformulación de estos y una evaluación de las técnicas y herramientas usadas para mejorar en ciclos posteriores de LA.

Aunque hemos planteado las fases de forma lineal sobre el papel, en la práctica no tienen por qué llevarse a cabo secuencialmente, como en el ejemplo que se muestra en la Figura 2 Esquema de la propuesta, donde hemos querido ilustrar algunas de la multitud de opciones posibles.

En esta figura, representamos con las flechas más grandes el orden propuesto inicialmente, en el que al acabar el ciclo con la actuación, volvemos a revisar los objetivos y comenzamos una nueva iteración, dando por supuesto que no cambian los usuarios. El resto de flechas ilustran otras posibilidades en las que se invierte el orden de las etapas o tengamos que rediseñar etapas anteriores debido a barreras que nos encontremos.

Por ejemplo, habiendo establecido que los usuarios de nuestro proyecto van a ser los alumnos, formulando los objetivos queremos hacer también partícipes a las familias; o en la recogida de datos nos topamos con información privada que no podemos disponer y tenemos que replantear objetivos.

También es posible que tengamos que volver al principio del ciclo por diversos motivos en cualquier otra etapa. O que nos saltemos el análisis (entendiéndolo como análisis computacional), como ya hemos aclarado previamente.

Parece claro que podemos dividir las etapas en una primera fase de planificación donde estableceremos usuarios y objetivos, e incluso el establecimiento de métricas, y una segunda de puesta en marcha que englobe las demás.

No obstante, una vez establezcamos los objetivos y nos planteemos qué queremos conseguir con LA, puede darse el caso en el que no podamos responder a nuestras inquietudes porque los datos que disponemos son insuficientes, no los podamos recoger, o estén protegidos por las leyes nacionales o comunitarias.

En conclusión, el método que se plantea es flexible y adaptable a las necesidades y características de cada circunstancia, en el que la planificación tiene especial importancia para el buen devenir del proyecto.

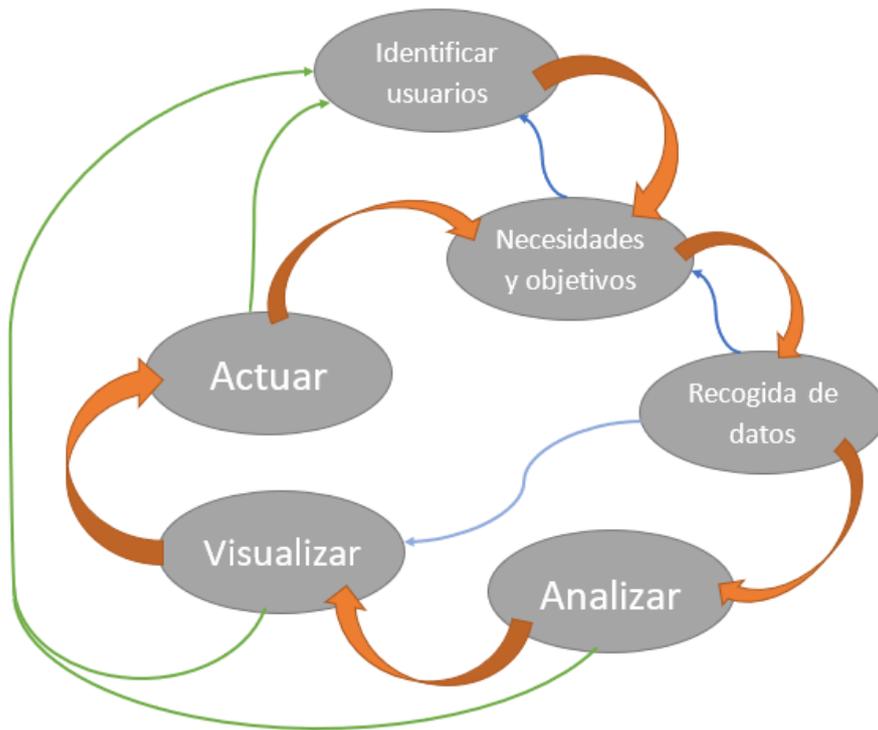


Figura 2 Esquema de la propuesta

4 Caso de estudio

El trabajo desarrollado en este caso de estudio se realizó en una única clase de formación profesional de grado medio, en un centro concertado urbano en la comunidad autónoma de Castilla y León. Dicho centro ofrece educación infantil, primaria, secundaria con ramas bilingües en inglés y numerosos títulos de formación profesional de la familia técnica y cuenta con cuantiosos recursos.

El docente a cargo de la clase y del caso de estudio es el autor de este documento, durante mi periodo de realización del prácticum. Estuve aproximadamente dos meses en el centro e impartí una unidad didáctica al grupo del que se trata en este apartado. El total de 6 sesiones de dos horas, de las cuales la última sesión se reservó para la realización de un examen.

Este grupo estaba formado por 18 alumnos estables de entre 16 y 18 años. Al comienzo de curso eran más de veinte, pero hoy en día acudían de forma ordinaria 18.

Se trataba de una unidad didáctica del módulo profesional Instalaciones Eléctricas Básicas (ELE02M0362). Mientras que el resto de las unidades de esta asignatura tenían una mayoría de carga práctica que motiva más al alumnado, la UD que impartí era de un contenido teórico. Realicé la programación de tal forma que se aplicaron 4 tipos de metodologías distintas para presentar nuevos conocimientos.

El principal fue el método expositivo debido fundamentalmente al carácter teórico de la unidad didáctica. En este método, yo exponía los contenidos e interactuaba con los alumnos de forma activa para captar su atención, resolver dudas y asegurar que me hacía entender. Los otros tres métodos utilizados fueron minoritarios, empleándose cada uno entre una y dos horas sobre el total de diez, y se detallan a continuación.

El primero fue el **aprendizaje cooperativo**. Se separó a los alumnos en grupos de entre 4 y 5 personas y eran ellos los que tenían que prepararse una sección del temario, cada grupo una parte, y posteriormente exponérsela al resto.

Otro método utilizado fue el de **aprendizaje basado en problemas**: se planteaba un problema, primero veíamos entre todos cómo poder resolverlo, y luego se planteaba un problema similar para que lo realizaran individualmente o por parejas.

Por último, una hora la dedicamos a hacer una **práctica en el exterior**, en la que, divididos en dos grandes grupos con roles individuales, se realizaba una instalación y se tomaban mediciones.

Como era mi primera experiencia como docente, tenía un gran interés por conocer qué métodos funcionaban mejor con mi manera de dar clase. Quería saber cuánto aprendían mis alumnos en las clases, cuán útil era su tiempo en el aula, descubrir qué les costaba más comprender y necesitaban repasar y qué entendían a la primera.

También deseaba que ellos fueran conscientes de qué puntos de lo que trabajábamos en clase retenían y cuáles les presentaban mayores dificultades. Que conocieran el grado de aprendizaje que estaban llevando a cabo.

Por lo tanto, distinguimos dos tipos de usuarios en este caso en el que aplicaremos LA: el profesor y los alumnos individuales.

Las necesidades del profesor son de una retroalimentación rápida y constante sobre su desempeño docente: quiere conocer cómo influyen sus métodos en el aprendizaje de los alumnos para poder mejorarlos. También quiere saber qué conocimientos presentan más y menos dificultades para dedicarles más o menos tiempo en próximas sesiones.

Respecto a las necesidades de los alumnos, no se realizó ninguna encuesta ni estudio, suponiéndose que sus objetivos principales eran aprender y aprobar. Para lograrlos, y más cuando se presentan ante un profesor desconocido, quieren saber específicamente qué han de aprender y en qué grado han interiorizado ese aprendizaje.

Con este análisis, establecemos los siguientes objetivos para uso de Learning Analytics en el aula; cuando el usuario es el docente:

- Conocer los conocimientos que causan más y menos problemas de aprendizaje a los alumnos.
- Conocer la efectividad de las clases y de los métodos de forma continuada.

Mientras que para los alumnos los objetivos serían:

- Conocer el grado de aprendizaje que está interiorizando.

Una vez hemos establecido los usuarios y objetivos de nuestro caso de Learning Analytics, hemos de establecer un plan de acción respecto a los datos: qué datos vamos a querer recolectar, cómo lo vamos a hacer, y cómo los vamos a procesar.

La mayoría de los objetivos que nos hemos marcado, han sido sobre la evaluación inmediata del aprendizaje, hasta qué punto unos contenidos han sido asimilados. Por ello, decidimos realizar pequeños test al principio de las clases que evaluaran el aprendizaje de los alumnos, salvo el primer día que no había ningún contenido que evaluar y el último que fue el examen. Por lo tanto, se realizaron cuatro test, entre 6 y 7 preguntas cada uno, de repaso y a modo de evaluación sobre lo visto el día anterior. Tras finalizar la quinta sesión también se realizó un último test de 7 preguntas con algunos de los conceptos en los que los alumnos habían obtenido menos calificación en los test previos.

Cada pregunta tenía 4 posibles respuestas y sólo una era correcta.

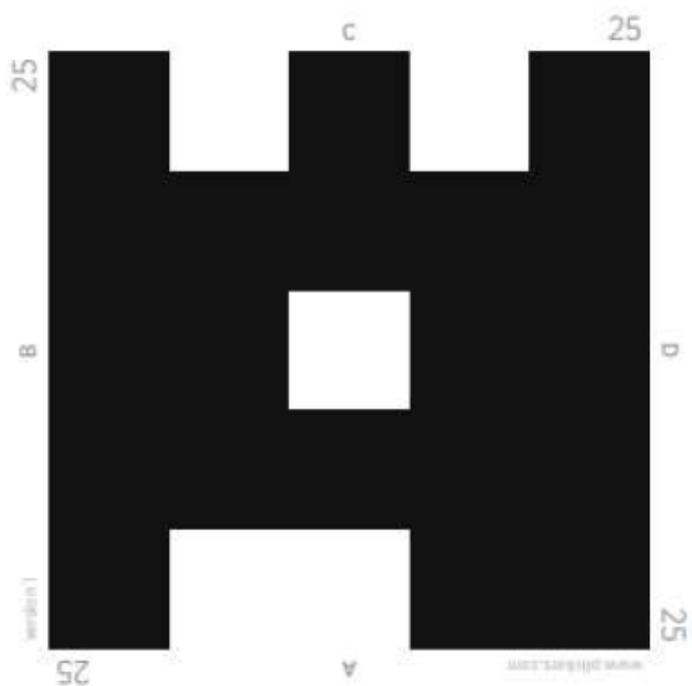


Figura 3 Código de Plickers

El método de recogida de datos fue a través de la plataforma Plickers. Elegimos esta plataforma en lugar de alguna semejante como Kahoot! porque no era necesario que los alumnos usaran el teléfono móvil en clase. Podíamos recoger en una plataforma web directamente sus respuestas sin los problemas que aparecen eventualmente en otras aplicaciones cuando se desconecta algún dispositivo. Con Plickers el único usuario que necesita tener conexión a internet es el profesor.

Se proyectaba cada pregunta del test y los alumnos tenían que responder colocando un papel con un código del estilo QR orientándolo con la letra que considerasen (A, B, C o D): en la Figura 4 podemos ver cómo era cada una de estas hojas de forma aclarativa. Cada alumno tenía un código distinto; de esta forma, el profesor desde la parte frontal de la clase podía escanear todas las respuestas, referenciadas a cada alumno, guardándolas en la aplicación y pudiendo acceder a ellas más tarde. Al finalizar cada una de estas pruebas, se proyectaba y discutía el test corregido.

El docente contaba también con un “diario de profesor”, que usaba antes, durante y después de las sesiones para preparar un esquema de las clases, anotar sucesos relevantes y un pequeño resumen de lo acontecido.

Aunque Plickers permite visualizar los datos en su plataforma en un entorno agradable, y con gráficos que detallan el número de personas que ha elegido cada opción de cada pregunta (Tabla 11), quiénes han sido esas personas, y la media que han obtenido tanto las personas como las preguntas (Figura 5), hemos descargado los datos en formato .csv y hemos cambiado por etiquetas los nombres de los alumnos, anonimizándolos.

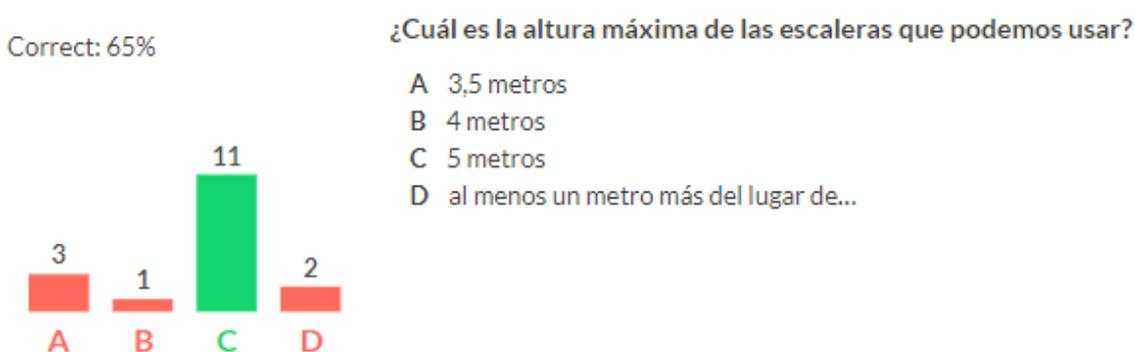


Tabla 11 Gráfico de Plickers

Los datos que tenemos en el archivo descargado son: etiqueta del alumno; porcentaje total de preguntas acertadas; número de preguntas respondidas correctamente; número total de preguntas respondidas; las preguntas completas; un enlace a la pregunta con las opciones; la opción que ha respondido cada alumno a cada una de las preguntas; la opción correcta de cada pregunta; y por último el porcentaje de acierto de cada pregunta.

Ahora volvamos a los objetivos y tratemos de realizar análisis y visualizaciones adaptadas a cada uno de ellos.

El primer objetivo que respondía a una necesidad del profesor era para poder adaptar las futuras sesiones conforme a estos conocimientos.

Conocer la efectividad de las clases y los métodos de forma continuada. Hay que ser muy concisos a la hora de establecer bien los objetivos. Aunque los agrupemos conscientemente en uno sólo, al realizar el análisis tendremos que trabajar dividiéndolos a su mínima expresión. De este, por lo tanto, podemos extraer dos sub-objetivos: conocer la efectividad de cada una de las clases y conocer la efectividad de los métodos aplicados.

El primero de ellos aún peca de demasiado genérico. ¿A qué nos referimos con efectividad?, ¿a que dé tiempo a ver todos los contenidos o a hacer todas las actividades?, ¿a que los alumnos estén en silencio?, ¿qué sean capaces de sacar la máxima nota en un examen?, ¿o quizás a que adquieran los conocimientos pretendidos?

¿Qué acción pretende motivar este objetivo?

Es importante hacerse preguntas y concretar qué queremos conseguir de tal forma que no quepa dudas de cómo vamos a focalizar nuestra herramienta.

En este caso, se trata de un profesor novel dando clase por primera vez y desea tener un seguimiento del aprendizaje de los alumnos para conocer si es necesario reforzar o aclarar conceptos. Por tanto, este sub-objetivo, contextualizado como lo hemos hecho, lo trataremos conjuntamente con el segundo objetivo del profesor, **conocer los conocimientos que causan más y menos problemas de aprendizaje a los alumnos.**



Figura 4 Visualización de Plickers

Para este primer sub-objetivo, o segundo objetivo principal, lo que buscamos es información de los conocimientos, del aprendizaje, por lo que realmente lo que queremos medir son solamente los datos evaluativos, las respuestas que ha tenido cada pregunta. Eso quiere decir que con la visualización proporcionada por Plickers, como la de la Tabla 11 o la Figura 5, podríamos sacar conclusiones acerca de cuáles no han sido interiorizadas por los alumnos y además con qué otras respuestas lo confunden.



Tabla 12 Progresión del porcentaje de aciertos por pregunta

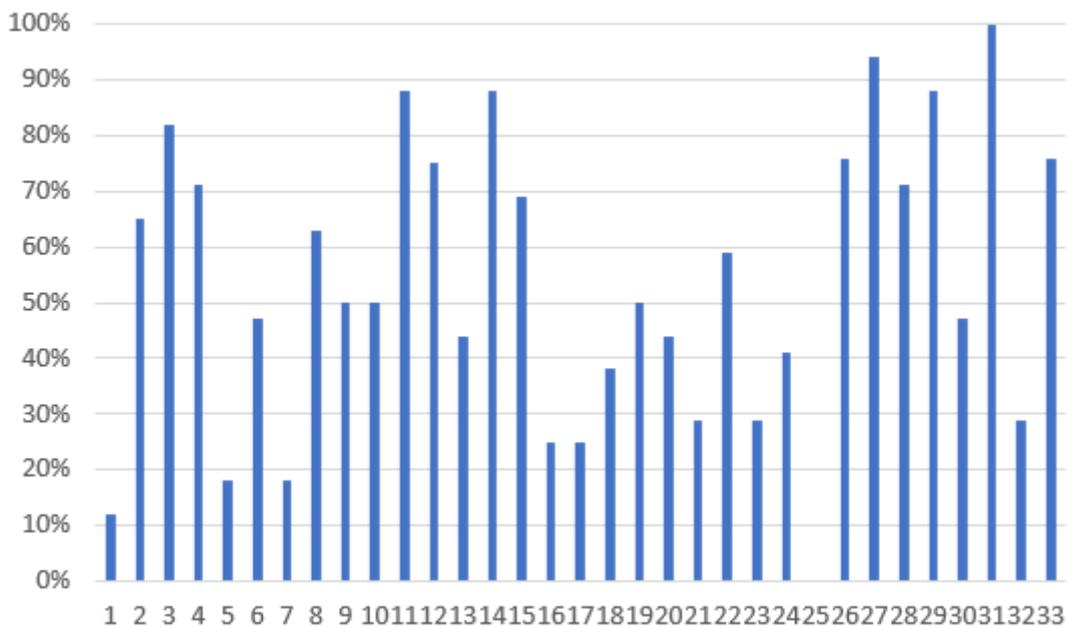


Tabla 13 Gráfico de barras donde podemos visualizar el porcentaje de acierto por cada pregunta

Para este objetivo descriptivo, podemos usar gran cantidad de tablas. La Tabla 12 nos servirá para ver la progresión general de la clase a lo largo de las preguntas. Con este

mismo formato podríamos realizar la progresión individual de cada alumno con las calificaciones obtenidas en las preguntas.

La Tabla 13 puede sernos realmente útil para encontrar rápidamente las preguntas que han presentado mayores dificultades e incidir en esos conceptos.

Por último, la Tabla 14 representa el diagrama de caja y bigotes con las notas diarias de cada alumno. Con esta tabla realmente podemos responder cómo de eficaz fue la sesión anterior, en general, para comprender e interiorizar los conceptos demandados en las preguntas.

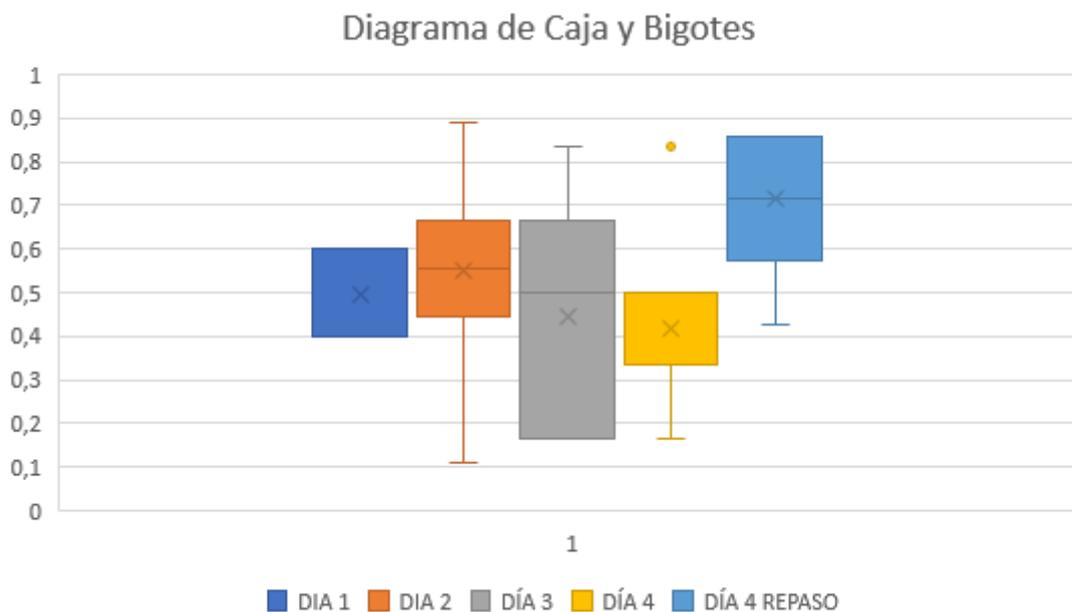


Tabla 14 Gráfico de caja y bigotes de las calificaciones de los alumnos por días

Lo cierto es que estamos observando el éxito o fracaso de la evaluación diaria de los alumnos sin entrar a analizar el por qué. La acción en la que derivará este análisis es paliar aquellos conceptos que hayan quedado menos claros. No servirá para mejorar la práctica docente del profesor, pero sí para que los efectos de las debilidades de su labor sean lo menores posibles.

En la Tabla 14 podemos observar también la importancia de los propios test en el aprendizaje. La primera vez que responden a preguntas nuevas, las distribuciones están localizadas en el centro y parte inferior de la gráfica. En cambio, en el test de repaso se observa que los alumnos ya han asimilado mejor los contenidos y cómo estructurarlos.

El segundo sub-objetivo del profesor era **conocer la efectividad de los métodos utilizados**. Este tipo de objetivo no pertenece al nivel descriptivo, sino al diagnóstico. Vamos a tratar de relacionar las metodologías empleadas por este profesor, en esta Unidad Didáctica y con este grupo, con el aprendizaje, o más específicamente, con la capacidad de responder correctamente las preguntas de nuestro test.

Para ello, agruparemos cada una de las 26 preguntas, sin contar las de repaso, según la metodología. Los tipos de metodología usados se presentaron al comienzo de este

epígrafe y se dividen en expositivo, cooperativo, resolución de problemas y práctica en el exterior. En la Tabla 15 Agrupación de preguntas por metodología utilizada podemos ver lo que obtuvimos.

	Expositivo	Cooperativo	Problemas	Práctica	Repaso
Etiqueta	1	2	3	4	5
Cantidad de preguntas	20	3	1	2	7
Media (sobre 1)	0,474	0,53	0,44	0,525	0,72142857
Desviación estándar	0,23386748	0,29810513	0	0,235	0,23948925

Tabla 15 Agrupación de preguntas por metodología utilizada

Lo primero de lo que nos damos cuenta es de las pocas preguntas que se han realizado de las metodologías no expositivas, en particular de la de aprendizaje basado en problemas, donde sólo se recogió una pregunta. Esto nos proporciona una muestra demasiado pequeña de cuyo análisis no podemos extraer resultados con significancia estadística.

El test estaba planteado como una actividad corta al inicio de cada sesión y, en el caso de los problemas, llevaba demasiado tiempo resolverlos como para añadir más de uno en los primeros minutos de la clase.

Para poder aplicar LA para conocer la efectividad de los métodos utilizados, debería haberse recogido un mayor número de preguntas de cada una de las técnicas. El problema por tanto ha sido de planificación del sistema de LA, específicamente el diseño de los objetivos que como ya hemos visto no han sido concisos y tomados en cuenta completamente en los siguientes pasos. De haberse querido desarrollar un sistema de LA que respondiese a estos objetivos, se debería priorizar la realización de más preguntas en cada una de las técnicas obteniendo así muestreos de los que podamos obtener resultados significativos.

El caso de estudio se realizó durante el periodo del prácticum, varios meses antes del diseño final de este método de aplicación de LA y redacción de la presente memoria. Por lo tanto, no disponía de tanto conocimiento e información acerca de LA, el método no estaba cerrado y no se planificó con suficiente profundidad.

Es nuestro deber incidir, nuevamente, en la importancia de planificar correctamente el método de LA: establecer usuarios, definir objetivos claros, y seleccionar métricas adecuadas para estos objetivos.

El último objetivo, que tenía como protagonistas a los alumnos, era **conocer el grado de aprendizaje que están interiorizando**. A través de este objetivo buscábamos motivar una reflexión individual por parte de los alumnos para que, tomando conciencia de su aprendizaje, los impulsase a mejorar su rendimiento.

Algo que se buscó desde el inicio de este caso de estudio era aprovechar el uso de las herramientas para responder a todos los objetivos posibles, por ello, utilizamos las mismas métricas y datos que habíamos usado cuando el usuario era el profesor.

La meta era conocer, concienciar, sobre el aprendizaje. Mostrarle qué sabe y qué no sabe. Los datos que recogíamos a través de Plickers nos daban precisamente esta información.

En las acciones realizadas por el profesor, nos limitamos a actuar en dos momentos diferenciados: al corregir cada pregunta en clase justo tras realizar el test informábamos sobre el porcentaje de acierto que había tenido la pregunta, pudiendo comparar su resultado con la distribución de la clase; el día de repaso antes del examen, en el que realizamos un test normal y un test de repaso, dimos la oportunidad a los alumnos de consultar todos sus resultados, generales y de las preguntas.

A través de estas visualizaciones, tratamos de empoderar a los alumnos hacia la mejora de su educación y aprendizaje.

5 Conclusiones y Líneas Futuras

5.1 Conclusiones

Hemos comenzado presentando Learning Analytics al lector: por un lado nos hemos detenido en explicar qué es esta disciplina; hemos hecho un repaso de su historia, sus inicios, su estado actual; y por último, estudiamos las barreras a las que se enfrenta actualmente y con las que tendrá que lidiar en el futuro.

A continuación, propusimos un marco metodológico para la aplicación de Learning Analytics en la educación secundaria donde distinguimos 6 fases distintas: identificar usuarios, identificar necesidades y plantear objetivos, recogida de datos, análisis de los datos, visualización y una actuación motivada por los pasos anteriores.

Por último, exponemos un caso de estudio utilizando el método propuesto. Este caso se realizó durante el prácticum del máster, en un módulo del área técnica del primer curso de un Ciclo Formativo de Grado Medio. Este caso de estudio se realizó cuando este trabajo aún no se había desarrollado en su totalidad por lo que su planificación y realización no resultaron tan consistentes como debieran.

A raíz de lo explicado hasta ahora, vamos a extraer un conjunto de conclusiones que sintetizen los resultados de este trabajo.

Learning Analytics es una disciplina versátil, aplicable a multitud de ámbitos tanto para “big data”, como para “small data”. Actualmente se encuentra en periodo de crecimiento y desarrollo y en un futuro a medio plazo será una herramienta de gran ayuda en el panorama educativo, personalizando la educación a las necesidades de los alumnos y los profesores, ofreciendo recursos adaptados a sus contextos. Hoy, nos encontramos en un buen momento para investigar y experimentar con esta disciplina.

La metodología propuesta ofrece una serie de pasos a los docentes para la puesta en marcha de herramientas de LA en contextos de educación primaria y secundaria. En estos centros las infraestructuras y recursos, tanto humanos como materiales, no son tan grandes ni especializados como en centros de educación superior. Por tanto, la metodología propuesta ha tenido en cuenta estas circunstancias.

Asimismo, esta serie de pasos no es sistemática, sino flexible, reestructurable y adaptable a cada circunstancia.

Este método otorga relevancia a todas las etapas del mismo, desde las fases de planificación a las propias de recogida, análisis de datos y acción motivada.

Tal y como se puede ver en el caso de estudio, es fundamental dedicar el tiempo necesario a la fase previa de identificación de objetivos, usuarios, y planificación del proceso, para obtener resultados representativos que nos permitan tomar decisiones. No dedicar el suficiente tiempo y reflexión a la planificación nos llevará al fracaso de nuestro proyecto en LA.

De igual manera, el estudio y conocimiento de las barreras es crítico para el logro de nuestras metas. Saber de antemano qué momentos serán los que hagan trastabillar nuestro proyecto, y anticiparnos a ellos, nos dará garantías de éxito. Probablemente no podamos predecir todas las barreras por lo que durante la puesta en marcha de cada una de las etapas deberemos evaluar fallos y posibles mejoras. De esta forma, en futuras iteraciones iremos mejorando nuestro ciclo ante distintos imprevistos.

5.2 Líneas futuras

Esta memoria constituye una fuente de posibles líneas de trabajo en el futuro, como consecuencia de la experiencia acumulada en el desarrollo de la misma. A continuación, proponemos algunas de ellas.

El primer campo de estudio que proponemos es el que concierne al nuevo reglamento europeo de protección de datos y su relación con la privacidad y LA. Sería de interés general para el crecimiento de esta disciplina la creación de un marco de convivencia entre el reglamento de protección de datos y las analíticas de aprendizaje. Este marco que vele por la privacidad de los datos de los usuarios y defina medidas, límites y propuestas para su uso en los entornos educativos.

Por otro lado, en general nos hemos encontrado con muy pocas investigaciones publicadas por parte de centros y profesores de la educación primaria y secundaria en relación con LA. La experimentación científica en este campo acarrearía más cantidad de propuestas y líneas de implementación para docentes que deseen embarcarse en la mejora de la educación a través de Learning Analytics.

Por último, sería conveniente un desarrollo más profundo de cada una de las fases de la metodología propuesta.

En especial nos referimos a las técnicas y herramientas de análisis de datos para la educación, en esta memoria presentadas en el apartado 3.4 Analizar. Estos métodos son descritos en multitud de documentos, pero sin esclarecer su funcionamiento a través de propuestas de uso, echando de menos transparencia en sus algoritmos.

Por ello, sería apremiante el contar con una guía de uso de las técnicas y las herramientas. Aún más ideal resultaría publicarlas y hacerlas accesibles a todos los docentes y

encargados de LA en los centros de tal forma que el trabajo técnico de esta fase no suponga un problema limitante.

6 Referencias

- (2015). Recuperado el 19 de Mayo de 2018, de LEA's Box 1st Learning Analytics Summercamp: <http://css-kmi.tugraz.at/mkrwww/leas-box/summercamp15.html>
- (2017). Recuperado el 26 de Mayo de 2018, de Portal de Educación de la Junta de Castilla y León: <http://www.educa.jcyl.es/es/informacion/aulas-virtuales-moodle-educacyl>
- Abdous, M., Wu, H., & Yen, C.-J. (2012). Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade. *Educational Technology & Society*, 77-88.
- Aguiar, E., Lakkaraju, H., Bhanpuri, N., Miller, D., Yuhas, B., & Addison, K. (2015). Who, when and why: A machine learning approach to prioritizing students at risk of not graduating high school on time. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge*, 93-102.
- Almeda, V., Scupelli, P., Baker, R., Weber, M., & Fisher, A. (2014). Clustering of Design Decisions in Classroom Visual Displays. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge*, 44-48.
- Amo Filvà, D. (Julio de 2015). Recuperado el 24 de Mayo de 2018, de Edulíticas: <http://www.eduliticas.com/2015/07/glosario/datos-crudos-en-la-analitica-del-aprendizaje/>
- Amo Filvà, D. (6 de Noviembre de 2015). *Edulíticas | Analítica del Aprendizaje - Learning Analytics*. Recuperado el 16 de Junio de 2018, de Niveles, fases y tiempos en la analítica del aprendizaje: <http://www.eduliticas.com/2015/11/divulgacion/niveles-fases-y-tiempos-en-la-analitica-del-aprendizaje/>
- Baker, R., & Siemens, G. (2013). Educational Data Mining and Learning Analytics.
- Buckingham Shum, S. (2012). *Learning Analytics: Policy Brief*. Institute for Information Technologies in Education. Moscú: UNESCO .
- Divjak, B., & Vondra, P. (2016). Learning Analytics: Meeting the Needs of Students and Teachers in Pre-tertiary Education. *Central European Conference on Information and Intelligent Systems* (pág. 117). Faculty of Organization and Informatics Varazdin.
- Drachslar, H., & Greller, W. (2016). Privacy and Learning Analytics - it's a DELICATE issue. *Proceedings of the Sixth International Learning Analytics and Knowledge Conference (LAK'16)*.
- Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., & Schroeder, U. (2013). Supporting Action Research with Learning Analytics. *Proceedings of the Third International Learning Analytics & Knowledge Conference (LAK13)*, 220-229.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304-307.

- Ferguson, R., & Clow, D. (2017). Learning Analytics - Avoiding Failure. *Educase Review Online*.
- International Educational Data Mining Society. (s.f.). *educationaldatamining.org*. Recuperado el 26 de Mayo de 2018, de <http://educationaldatamining.org/>
- LAK11. (2011). Proceedings of the 1st International Conference on Learning Analytics and Conference. *1st International Conference on Learning Analytics and Knowledge*. Banff, Alberta: The Association for Computing Machinery. Obtenido de <https://tekri.athabascau.ca/analytics/>
- Lang, C., Siemens, G., Wise, A., & Gašević, D. (Edits.). (2017). *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR).
- Long, P., & Siemens, G. (Septiembre/Octubre de 2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE REVIEW*, 46(5), 32-37.
- Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2), 588-599.
- Martínez-Monés, A., Reffay, C., Hoyos Torío, J., & Muñoz Cristobal, J. (2017). Learning Analytics with Google Classroom: Exploring the possibilities., (págs. 1-6).
- Reglamento (UE) del Parlamento Europeo y del Consejo relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE. (27 de abril de 2016).
- Rodríguez-Triana, M., Martínez-Monés, A., & Villagrà-Sobrino, S. (2016). Learning Analytics in Small-Scale Teacher-Led Innovations: Ethical and Data Privacy Issues. *Journal of Learning Analytics*, 3(1), 43-65.
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (Edits.). (2011). *Handbook of Educational Data Mining*. Chapman & Hall /CRC.
- Romero-Zaldivar, V.-A., Pardo, A., Burgos, D., & Delgado Kloos, C. (2012). Monitoring student progress using virtual appliances: A case study. *Computers & Education*, 58(4), 1058-1067.
- Sharples, M., Adams, A., Ferguson, R., Gaved, M., McAndrew, P., Rienties, B., . . . Whitelock, D. (2014). Learning design informed by analytics. *Innovating Pedagogy 2014: Open University Innovation Report 3*, 12-14.
- Sharples, M., de Roock, R., Ferguson, R., Gaved, M., Herodotou, C., Koh, E., . . . Wong, L. H. (2016). Formative Analytics. *Innovating Pedagogy 2016: Open University Innovation Report 5*, 32-34.
- Sharples, M., McAndrew, P., Weller, M., Ferguson, R., FitzGerald, E., Hirst, T., & Gaved, M. (2013). Learning Analytics. *Innovating Pedagogy: Open University Innovation Report 2*, 14-16.
- Siemens, G. (6 de Mayo de 2015). *Learning Analytics: Advancing the Science of Learning*. (CLEARUNT, Editor) Recuperado el 23 de Mayo de 2018, de Youtube: <https://www.youtube.com/watch?v=L62KafPI4AU>

- Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: towards communication and collaboration. (S. Buckingham Shum, D. Gasevic, & R. Ferguson, Edits.) *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252-254.
- Strauss, V. (21 de Abril de 2014). Recuperado el 30 de Mayo de 2018, de The Whashington Post: https://www.washingtonpost.com/news/answer-sheet/wp/2014/04/21/100-million-gates-funded-student-data-project-ends-in-failure/?noredirect=on&utm_term=.37574aedbc0d
- Strauss, V. (4 de Abril de 2014). Recuperado el 30 de Mayo de 2018, de The Whashington Post: https://www.washingtonpost.com/news/answer-sheet/wp/2014/04/04/student-privacy-activists-win-a-big-one/?utm_term=.a6f150cf446d
- Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, 119-129.