

Análisis multivariante del uso de espacios virtualizados por estudiantes pregraduados en Ciencias de la Salud

M. Pilar Álvarez-Vázquez, Ana M. Álvarez-Méndez, Carmen Bravo-Llatas, M. Teresa Angulo-Carrere

Introducción. Los entornos virtuales de aprendizaje permiten crear espacios dinámicos y facilitadores del aprendizaje. Investigar el uso dado por los estudiantes puede identificar patrones de comportamiento y detectar tempranamente alumnos en riesgo de abandono, y se han descrito correlaciones entre su uso y el rendimiento académico.

Materiales y métodos. Se estudiaron siete espacios virtualizados correspondientes a cuatro asignaturas de tres grados de Ciencias de la Salud impartidas en los cursos 2017/2018 y 2018/2019, con un total de 517 estudiantes. Previamente se extrajeron, depuraron y anonimizaron los registros de cada espacio. Las variables analizadas fueron: número de visitas al campus virtual, de accesos a recursos y a URL, y uso del foro. Se aplicó un análisis de correspondencias múltiples, seguido de un análisis de conglomerados jerárquico.

Resultados. Se obtuvieron cuatro clústeres, con tamaños de entre el 20,9 y el 29,4% de los estudiantes, caracterizados por comportamientos diferenciales en cuanto al uso del campus virtual, y se establecieron relaciones con las calificaciones finales, las notas teóricas y las prácticas de las asignaturas. Se observa que, a menor interacción en el campus virtual, menor rendimiento académico, mientras que, a mayor actividad registrada, mejores calificaciones.

Conclusiones. Nuestro estudio revela grupos de estudiantes con comportamientos homogéneos según su uso del campus virtual y establece relaciones con el rendimiento académico.

Palabras clave. Análisis de conglomerados. Analítica del aprendizaje. Ciencias de la Salud. Educación superior. Moodle. Rendimiento académico.

Multivariate analysis of the use of virtualized spaces by health sciences undergraduate students

Introduction. Virtual learning environments enable users to create dynamics and learning facilitator spaces. To investigate the students' use can identify patterns and help to an early detection of students at high risk of dropping out since correlations between its use and the academic performance have been described.

Materials and methods. Seven virtualized spaces corresponding to four courses from three Health Sciences degrees taught in 2017/18 and 2018/19, with a total of 517 students were studied. Previously, logs had been extracted from every space, debugged and anonymized. Number of logins, of access to resources and to URL as well as the forums use were considered. A multiple correspondence analysis was applied followed by a hierarchical clustering analysis.

Results. 4 clusters, with sizes between 20.9% and 29.4% of the students, were obtained and characterized by differential behaviors of the virtual campus use. Relationships can be established with final grades as well as theory' and practical' grades. Results pointed out that the lower interaction in virtual campus, the lower the grades, while the higher interaction, the higher the marks.

Conclusions. Our study pinpoints different student clusters with homogeneous virtual campus behavior and establishes relationships with the academic performance.

Key words. Academic performance. Cluster analysis. Health Sciences. Higher education. Learning analytics. Moodle.

Introducción

Uno de los objetivos del Espacio Europeo de Educación Superior (EEES) es centrar la enseñanza en el aprendizaje del estudiante, en su formación y capacitación continuas. El EEES enfatiza que el estudiante sea activo y se responsabilice de su propio aprendizaje mientras que el docente le guíe en un

proceso dinámico, colaborativo y constructivista entre iguales. Los entornos virtuales de aprendizaje han permitido crear espacios seguros, dinámicos y facilitadores del aprendizaje. Según Gros Salvat, los entornos virtuales de aprendizaje deben proporcionar un espacio social en donde alumnos y profesores construyan conocimiento de forma conjunta [1].

Sección Departamental de Biología Celular. Facultad de Medicina (M.P. Álvarez-Vázquez). Departamento de Enfermería. Facultad de Enfermería, Fisioterapia y Podología (A.M. Álvarez-Méndez, M.T. Angulo-Carrere). Área de Gobierno de Tecnologías de la Información y Apoyo Técnico al Usuario. Universidad Complutense de Madrid. Madrid, España (C. Bravo-Llatas).

Correspondencia:

Dra. M. Pilar Álvarez Vázquez. Sección Departamental Biología Celular. Facultad de Medicina. Pza. Ramón y Cajal, s/n. Ciudad Universitaria. E-28040 Madrid.

E-mail:

pilar@med.ucm.es

Agradecimientos:

Esta comunicación es fruto de proyectos Innova concedidos por el vicerrectorado de Calidad de la UCM.

Recibido:

15.07.21.

Aceptado:

17.11.21.

Conflicto de intereses:

No declarado.

Competing interests:

None declared.

Cómo citar este artículo:

Álvarez-Vázquez MP, Álvarez-Méndez AM, Bravo-Llatas C, Angulo-Carrere MT. Análisis multivariante del uso de espacios virtualizados por estudiantes pregraduados en Ciencias de la Salud. FEM 2021; 24: 317-21. doi: 10.33588/fem.246.1159.

© 2021 FEM

Tabla 1. Descripción de los espacios analizados.

	Asignatura	Curso académico	Grado en	Curso de la titulación	N.º de alumnos
Prácticas		2017/2018			79
Prácticas	OMH	2018/2019	Medicina	2.º de 6	94
Teoría					94
Teoría y prácticas	AH	2017/2018	Fisioterapia	2.º de 4	60
Teoría y prácticas	BM	2017/2018			70
Teoría	SL	2017/2018	Podología	4.º de 4	62
Teoría		2018/2019			58
7	4	2	3	2	517

AH: Anatomía humana; BM: Biomecánica; OMH: Organografía microscópica humana; SL: Salud laboral.

Los registros generados en entornos digitales se han empleado desde hace décadas en el mundo empresarial para conocer mejor las necesidades del cliente y tomar decisiones estratégicas [2-4]. El mundo educativo empezó más tarde a utilizar el análisis masivo de datos. Así, las primeras conferencias sobre minería de datos de educación y analítica de aprendizaje y conocimiento se celebraron en Canadá en 2008, y el primer libro sobre minería de datos en *e-learning* se publicó en 2006 [5]. La analítica de aprendizaje se vislumbraba como un nuevo método para mejorar la docencia, el aprendizaje, la eficiencia de las organizaciones y la toma de decisiones fundadas [6]. El análisis masivo de datos a nivel educativo se plantea en tres niveles: descriptivo, predictivo y prescriptivo [7].

A diferencia de lo que sucede en otros países [8,9], las instituciones educativas españolas apenas explotan esa fuente de información. Recientemente, tres universidades madrileñas han creado, en colaboración con una empresa, un portal de datos abiertos sobre educación superior [10]. Son escasos los docentes que investigan el uso que hacen sus estudiantes del campus virtual (CV), identifican patrones de comportamiento y detectan tempranamente alumnos en riesgo de abandono [5,11-13], a pesar de haberse hallado correlaciones entre el uso del campus virtual y el sentimiento de pertenencia a una comunidad conectada [14], y el rendimiento académico [15-17].

El objetivo del trabajo es analizar si el comportamiento de estudiantes procedentes de distintas asig-

naturas, cursos, titulaciones, en diferentes espacios del campus virtual y estudiados de forma global revela la existencia de conglomerados de estudiantes y su posible relación con el rendimiento académico.

Materiales y métodos

Se realizó un análisis multivariante sobre siete espacios correspondientes a cuatro asignaturas, impartidas en los cursos 2017/2018 y 2018/2019, con un total de 517 estudiantes, de los cuales el 69,6% eran mujeres. Las asignaturas formaban parte de los grados en Medicina, Fisioterapia y Podología, y se impartían tres en segundo y una en cuarto curso de sus respectivos grados. Se analizaron espacios virtuales creados en la plataforma Moodle, que podían estar dedicados a la parte práctica, a la teórica o a ambas (Tabla 1). Se informó tanto el alumnado como el Vicerrectorado de Tecnologías de la Información y aceptaron que los registros almacenados en Moodle se analizaran con fines académicos.

Las variables analizadas fueron el número de visitas realizadas al campus virtual, de accesos a recursos y a URL, y el uso del foro. Para homogeneizar la variabilidad entre asignaturas, se transformó cada variable en cualitativa dividiendo las frecuencias en cuartiles (Q) en cada asignatura. Respecto del rendimiento, se eligieron: notas finales de prácticas y de teoría en primera convocatoria (prácticas 1 y teoría 1), notas finales en ambas convocatorias (actas 1 y actas 2) y notas de las pruebas de evaluación continua en prácticas (minitest), en función de cada espacio. Las categorías de estas variables fueron: suspenso, aprobado, notable y sobresaliente.

Se aplicó un análisis de correspondencias múltiples [18] a las variables de uso del campus virtual, seguido de un análisis de conglomerados ascendente jerárquico aplicado a las puntuaciones factoriales, utilizando el criterio de Ward con la distancia euclídea. Se obtuvieron clústeres al encontrarse una disminución importante de la variabilidad interclúster al formarse una nueva partición. Los clústeres se describen por las características de uso del campus virtual y de las calificaciones con mayor representación respecto del conjunto total y por las calificaciones con promedios diferenciados del promedio general, cuando se rechaza la hipótesis de una distribución azarosa de los estudiantes a los clústeres (aproximaciones de la distribución hipergeométrica y de la media muestral a la distribución normal, respectivamente). Se utilizó el *software* SPAD 9.1. El nivel de significación se fijó en $p < 0,05$.

Tabla II. Caracterización de los clústeres obtenidos. A. Características significativas de mayor uso del campus virtual en cada clúster: porcentajes de estudiantes de cada clúster y total en los cuartiles (Q) de visitas, recursos, foros y URL. B. Calificaciones significativas de los clústeres (medias \pm desviación estándar).

Clúster	A																B			
	Visitas				Recursos				Foros				URL				Calificaciones			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Prácticas 1	Teoría 1	Actas 1	Minitest
1 (n = 152; 20,9%)	91% ^a	ns	ns	ns	82% ^a	ns	ns	ns	57% ^a	24% ^a	ns	ns	56% ^a	ns	ns	ns	6,2 \pm 2,4 ^a	4,4 \pm 0,5 ^a	5,2 \pm 2,8 ^a	6,2 \pm 2,4 ^a
2 (n = 108; 29,4%)	ns	75% ^a	ns	ns	ns	61% ^a	ns	ns	53% ^a	ns	ns	ns	45% ^a	ns	ns	ns	ns	ns	ns	ns
3 (n = 142; 27,5%)	ns	ns	79% ^a	ns	ns	ns	54% ^a	ns	ns	ns	ns	51% ^a	ns	26% ^c	37% ^a	ns	7,6 \pm 1,4 ^b	6,2 \pm 2,3 ^c	7,1 \pm 2,3 ^b	ns
4 (n = 115; 22,2%)	ns	ns	ns	88% ^a	ns	ns	ns	76% ^a	ns	ns	ns	48% ^b	ns	ns	ns	63% ^a	ns	ns	7 \pm 2 ^c	7,7 \pm 1,3 ^b
Total (n = 517; 100%)	25%	25%	26%	24%	24%	26%	26%	24%	42%	11%	20%	37%	33%	20%	21%	26%	7,1 \pm 2	5,7 \pm 2,4	6,5 \pm 2,5	7,1 \pm 1,7

ns: no significativo. ^a $p < 0,001$; ^b $p < 0,01$; ^c $p < 0,05$.

Resultados

El análisis establece cuatro clústeres, que suponen el 49% de la variabilidad total explicada por la variabilidad interconglomerados. La tabla II recoge las características de cada uno. La figura representa los estudiantes en los dos primeros ejes factoriales agrupados en cuatro clústeres, las variables de uso del campus virtual, rendimiento académico y sexo. Nos permite diferenciar los estudiantes según su nivel de actividad y participación y su relación con el rendimiento académico. Esta representación acumula el 27,9% de la variabilidad inicial.

El clúster 1 tiene mayor proporción de estudiantes del Q1 en número de visitas y accesos a recursos y a URL y del Q1 y el Q2 de accesos a foros. Agrupa el 70% de los estudiantes menos activos en accesos a recursos y el 75% de los que menos visitas realizaron de todos los estudiantes. Contiene mayor proporción de suspensos en teoría 1, suspensos y aprobados en actas 1, aprobados en prácticas 1, y suspensos y notables en actas 2, así como de varones (38,9 frente a 30,4%).

El clúster 2 tiene una mayor representatividad de estudiantes con visitas y accesos a recursos en el Q2, así como de accesos a URL y a foros en el Q1.

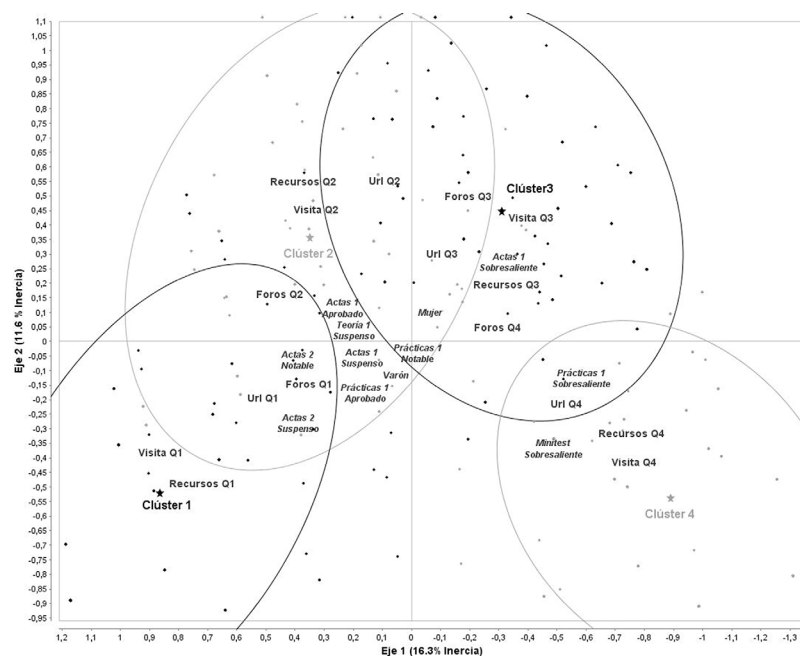
Aglutina el 70% de los estudiantes con accesos a recursos y el 88% de visitas del Q2 del total de estudiantes. No hay diferencias estadísticamente significativas en el rendimiento académico, salvo una mayor proporción de notables en prácticas 1.

El clúster 3 agrupa al 84% de los estudiantes con número de visitas en el Q3, representado por estudiantes del Q3 en acceso a recursos, del Q2 y el Q3 en acceso a URL y del Q4 en acceso a foros. Presenta mayor proporción de sobresalientes en actas 1 y de mujeres (79,6 frente a 69,6%).

El clúster 4 incluye mayoritariamente a los más activos (Q4) en número de visitas y accesos a recursos, URL y foros. Del total de estudiantes, aglutina el 82% de los de mayor número de visitas y el 71% con mayor número de accesos a recursos. Tiene mayor proporción de sobresalientes en prácticas 1 y minitest.

Al analizar el rendimiento medio de los estudiantes de los clústeres, encontramos que las variables características de los clústeres 1, 3 y 4 reflejan claras diferencias del promedio general. Así, los individuos de los clústeres con mayor uso del campus virtual alcanzan mejores notas (medio punto en promedio) en teoría 1, prácticas 1, actas 1 y minitest que el promedio de todos los estudiantes, en

Figura. Estudiantes, variables de uso del campus virtual, rendimiento académico y sexo proyectados sobre los dos primeros ejes factoriales. Elipses y centroides de los clústeres.



tanto que los estudiantes más pasivos obtuvieron peores notas medias en teoría 1, prácticas 1, actas 1 y minitest (un punto menos en promedio).

Discusión

De acuerdo con Buckingham Shum y Ferguson, el enfoque del análisis de datos puede hacerse en tres niveles [19]: a) micro, esto es, en un curso o asignatura; b) meso, es decir, en una titulación, y c) macro, si se refiere a un centro de estudios. Este estudio se sitúa entre los niveles micro y meso al analizar varios espacios virtualizados de distintos grados de Ciencias de la Salud.

Un entorno virtual de aprendizaje requiere que el espacio virtual sea también un espacio social [20] en el que los usuarios tengan una mayor libertad de elección, puedan escoger distintos trayectos y ritmos, y decidir las interacciones que realizan. En la bibliografía se discute la relevancia de los datos de las plataformas digitales educativas. Algunos autores argumentan que una parte significativa de las interacciones entre estudiantes acontecen en las redes sociales [21] y otros dudan de que el mero análisis de los datos cuantitativos permita interpretar el

comportamiento de los estudiantes si no se conoce el contexto [22]. Por ello, algunos trabajos proponen clasificar las interacciones [21,23,24], seleccionar los registros [25,26] y vincular las acciones de los estudiantes con condiciones internas y externas [17].

Sin embargo, numerosas evidencias indican una relación clara entre entorno virtual de aprendizaje y rendimiento académico en distintas áreas de conocimiento. Aunque se han descrito usos erráticos y dispares [27,28], y algunos autores consideran que puede ser un factor de distracción [29], muchos trabajos han demostrado un efecto positivo del campus virtual en las calificaciones [30-33] y que la actividad registrada puede proporcionar una medida indirecta válida del compromiso y la motivación del alumnado [34].

Conclusiones

Este trabajo realiza un análisis retrospectivo global de una muestra de estudiantes de asignaturas muy diferentes en cuanto a duración, número de créditos y peso relativo de teoría y prácticas, curso académico de la titulación, diseño del campus virtual, herramientas docentes empleadas y sistema de evaluación. Esta aparente diversidad no ha impedido encontrar un denominador común en el alumnado: el uso del campus virtual se relaciona directamente con los resultados académicos, de forma que, a menor interacción, peores calificaciones de todas las variables en primera convocatoria, mientras que, a mayor actividad registrada, mejores calificaciones en todas ellas. Así pues, nuestros resultados demuestran, también al analizar globalmente muestras de estudiantes universitarios procedentes de varias asignaturas, titulaciones y cursos, la existencia de una relación directa entre el uso de Moodle por el alumnado y el rendimiento académico.

La identificación de patrones de aprendizaje a partir de los registros generados en entornos virtuales de aprendizaje proporciona información valiosa para la mejora de la docencia, ya que permite a los profesores realizar un seguimiento y adoptar estrategias adaptativas, y también para los estudiantes, que pueden requerir asesoramiento para mejorar sus calificaciones y redefinir sus metas [35]. A nuestro juicio, los retos principales que debe enfrentar la analítica de aprendizaje son fomentar el aprendizaje personalizado, incluir en el análisis de variables externas al propio uso de entornos virtuales de aprendizaje, como pueden ser datos socioeconómicos del alumnado, y mejorar la retroalimentación que recibe cada estudiante.

Bibliografía

- Gros Salvat B. La construcción del conocimiento en la red: límites y posibilidades. *Education in the Knowledge Society* 2004; 5.
- Büchner, AG, Mulvenna MD. Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining. *ACM SIGMOD Record* 1998; 27: 54-61.
- Groth R. *Data mining building competitive advantage*. Hoboken: Prentice Hall; 2000.
- Castañeda J, Rodríguez M. La minería de datos como herramienta de marketing: delimitación y medidas de evaluación del resultado. Granada: Universidad de Granada, Departamento de Comercialización e Investigación de Mercados; 2003.
- Romero C, Ventura S. Educational data mining and learning analytics: an updated survey. *WIREs Data Mining and Knowledge Discovery* 2020; 10: e1355.
- Long P, Siemens G. Penetrating the fog: analytics in learning and education. *EDUCAUSE* 2011; 4: 5.
- Daniel B. Big data and analytics in higher education: opportunities and challenges. *Br J Educational Tech* 2015; 46: 904-20.
- Porter S. Survey research policies: an emerging issue for higher education. *New Directions in Institutional Research* 2005; 127: 5-15.
- Kennedy G, Corrin L, Lockyer L, Dawson S, Williams D, Mulder RA, et al. Completing the loop: returning learning analytics to teachers. Rhetoric and reality. *Critical perspectives on educational technology*. Proceedings ASCILITE Dunedin 2014: 436-40.
- DIMETRICAL, The Analytics Lab, S.L. *Universidata*. 2020. Disponible en <https://www.universidata.es>. Fecha última consulta: 10.07.21.
- Caminal de Mingo A, Vila B, García P, Puigcerver M. Análisis del uso de Moodle por parte de alumnos de la asignatura Matemáticas, Ciencias Experimentales y Educación, del Grado de Educación Infantil de la Universidad de Barcelona. En Calvo J, Pagès T, eds. *El aula Moodle: aprender y enseñar en la Universidad de Barcelona*. Barcelona: Octaedro; 2015. p. 105-12.
- Caminal de Mingo A, Puigcerver M. Diseños de campus virtual: descripción de la oferta actual en la enseñanza de las ciencias experimentales y su influencia en el uso por parte del alumnado universitario. *Didacticae* 2017; 2: 119-33.
- Hernández Estrada A, Martínez Rodríguez ME, Casado de Lucas D, Peñaloza Figueroa JL, Pérez Martín M, Arteaga Martínez B, et al. Big data en educación: tipologías de los estudiantes a partir del estudio de las interacciones dentro del triángulo pedagógico. 2017; <https://eprints.ucm.es/44899/>. Fecha última consulta: 28.12.2020.
- Black EW, Dawson K, Priem J. Data for free: using LMS activity logs to measure community in online courses. *Internet and Higher Education* 2008; 11: 65-70.
- Hidalgo Cajo BG. Minería de datos en los sistemas de gestión de aprendizaje en la educación universitaria. *Campus Virtuales* 2018; 7: 115-28.
- Jenaro Río C, Castaño Calle R, Martín Pastor ME, Flores Robaina N. Rendimiento académico en educación superior y su asociación con la participación activa en la plataforma Moodle. *Estudios sobre educación* 2018; 34: 177-98.
- Gros B, Cano E. *Educación 2018-2020. Retos, tendencias y compromisos*. Barcelona: IRE-UB; 2020.
- Lebart L, Piron M, Morineau A. *Statistique exploratoire multidimensionnelle: visualisation et inférences en fouilles de données*. 4 ed. Paris: Dunod; 2006.
- Buckingham Shum S, Ferguson R. Social Learning Analytics. *Educational Technology & Society* 2012; 15: 3-26.
- Brown JS, Duguid P. *La vida social de la información*. Buenos Aires: Prentice Hall; 2001.
- Scheffel M, Drachler H, Stoyanov S, Specht M. Quality indicators for learning analytics. *Educational Technology & Society* 2014; 17: 117-32.
- Ellis RA, Han F, Pardo A. Improving learning analytics - combining observational and self-report data on student learning. *Educational Technology & Society* 2017; 20: 158-69.
- Chaparro Peláez J, Iglesias Pradas S, Pascual Miguel F. Uso del registro de actividad de Moodle para un estudio del rendimiento académico de alumnos en entornos en línea y presencial. *4th International Conference on Industrial Engineering and Industrial Management 2010*: 753-60.
- Agudo-Peregrina ÁF, Iglesias-Pradas S, Conde-González MÁ, Hernández-García Á. Can we predict success from log data in VLEs? Classification of interactions of learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior* 2014; 31: 542-50.
- Gasevic D, Dawson S, Siemens G. Let's not forget: learning analytics are about learning. *Tech-Trends* 2015; 59: 64-71.
- Liu DY-T, Froissard JC, Richards D, Atif A. An enhanced learning analytics plugin for Moodle: student engagement and personalized intervention. In Reiners T, von Kinsky BR, Gibson D, Chang V, Irving L, Clarke K, eds. *Globally connected, digitally enabled*. Perth: Proceedings ASCILITE; 2015. p. 168-77.
- Vinueza Morales SX, Morocho Macas ÁA. Análisis del rendimiento académico en la cátedra de Fisiología y Fisiopatología usando como herramienta el aula virtual. *Cuadernos de Desarrollo Aplicados a las TIC* 2017; 6: 43-60.
- Angulo Carrere MT, Álvarez-Méndez AM, Bravo-Llatas C, Cristóbal Barrios J, Álvarez Vázquez MP. Diferencias de comportamiento entre estudiantes de dos asignaturas de Ciencias de la Salud. *Actas de IN-RED* 2020; 2020: 71-82.
- Granados-Zúñiga J. Relación entre el uso del aula virtual y el rendimiento académico en estudiantes del curso de Bioquímica para Enfermería de la Universidad de Costa Rica. *Revista Educación* 2019; 43: 561-76.
- Cantabella M, Martínez-España R, Ayuso B, Yañez JA, Muñoz A. Analysis of student behavior in learning management systems through a Big Data framework. *Future Generation Computer Systems* 2019; 90: 262-72.
- Álvarez-Méndez A, Angulo Carrere MT, Cristóbal Barrios J, Bravo-Llatas C, Álvarez Vázquez MP. Application of data mining in Moodle platform for the analysis of the academic performance of a subject of Podiatry Degree. *Proceedings of INTED 2020*; 2020: 984-92.
- Álvarez Vázquez MP, Angulo Carrere MT, Cristóbal Barrios J, Bravo-Llatas C, Álvarez-Méndez AM. Learning analytics in Human Histology reveals different student' clusters and different academic performance. *Proceedings of INTED 2020*; 2020: 66-72.
- Álvarez Vázquez MP, Angulo Carrere MT, Bravo-Llatas MC, Cristóbal Barrios J, Álvarez-Méndez AM. Análisis de los registros en Moodle para estudiar la actividad desarrollada por estudiantes de Histología humana y su relación con el rendimiento académico. *Actas de IN-RED* 2020; 2020: 48-60.
- Motz B, Quick J, Schroeder N, Zook J, Gunkel M. The validity and utility of activity logs as a measure of student engagement. *The 9th International Learning Analytics & Knowledge Conference* 2019; 300-9.
- Ferguson R, Barzilai S, Ben-Zvi D, Chinn CA, Herodotou C, Hod Y, et al. *Innovating pedagogy 2017: Open University innovation report 6*. Milton Keynes: The Open University; 2017.