



TESIS DOCTORAL

2022

EL PERFIL DEL ESTUDIANTE EN LOS MOOC: IDENTIFICACIÓN DE PATRONES TEMPORALES DE USO A PARTIR DEL ANÁLISIS DE DOS MOOC DE LENGUAS

JUAN JOSÉ DEL PERAL PÉREZ

PROGRAMA DE DOCTORADO EN FILOGÍA. ESTUDIOS
LINGÜÍSTICOS Y LITERARIOS

DIRECTORA: DRA. M^a DOLORES CASTRILLO DE LARRETA-AZELAIN

AGRADECIMIENTOS

Esta tesis doctoral es el resultado de cinco años de trabajo arduo, de grandes esfuerzos y de sacrificios, en condiciones que no han sido en absoluto las ideales. Por este motivo quiero agradecer en primer lugar a Marta por su generosidad y su confianza eternas, por su paciencia durante tantos años acompañándome en este accidentado viaje que nos ha llevado a un rincón sosegado, sin marañas, bañado de luz y de promesas. Te perdono por todas las interrupciones. Te doy las gracias por todas las interrupciones.

Agradezco también a mis padres por su abnegación sin medida, por su amor incondicional, por darme siempre todo lo que tienen y un poco más, por tenderme la mano en cada nueva etapa, sin hacerme nunca dudar. Todo lo que soy, todo lo que puedo soñar ser, se lo debo a ellos.

También quiero agradecer a mi hermano por su apoyo a lo largo de los años, por su generosidad y su cariño sin límites, por su inmensa sabiduría. Gracias por unirme con los ojos cerrados a esta aventura. Y a mi primo, por creer siempre en mí, porque el orgullo que siente por mí me emociona.

En especial, me gustaría expresar mi más profundo agradecimiento y admiración a mi directora de tesis, la Dra. M^a Dolores Castrillo de Larreta-Azelain. Gracias por su inestimable ayuda, su cercanía y su paciencia a lo largo de estos años de investigación. Ha sido un enorme privilegio poder realizar este trabajo con su guía, sus valiosos consejos y su apoyo constante.

Por último, me gustaría agradecer a la coordinación de la escuela de Doctorado en Filología de la UNED por su magnífica labor y por su respaldo, el cual he sentido en todo momento. Me siento sumamente afortunado de haber podido formar parte de este programa.

TABLA DE CONTENIDO

<i>Listado de abreviaturas y siglas</i>	7
<i>Listado de tablas</i>	9
<i>Listado de figuras</i>	11
<i>Listado de imágenes</i>	13
<i>Listado de gráficos</i>	14
CAPÍTULO 1: INTRODUCCIÓN	16
1.1. Justificación	16
1.2. Objetivos y preguntas de la investigación.....	24
1.3. Organización y estructura de la tesis	26
CAPÍTULO 2: APROXIMACIÓN HISTÓRICA A LA ENSEÑANZA-APRENDIZAJE DE LENGUAS A DISTANCIA	29
2.1. Introducción	29
2.2. Definición de enseñanza-aprendizaje de idiomas.....	31
2.3. Definición de educación a distancia	35
2.4. Inicios de la enseñanza de lenguas a distancia.....	40
2.5. Enseñanza de lenguas asistida por ordenador (CALL).....	47
2.5.1. Definición de términos	48
2.5.2. CALL conductista.....	52
2.5.3. CALL comunicativo	54
2.5.4. CALL integrativo.....	56
CAPÍTULO 3: LOS MOOC: EVOLUCIÓN DE LOS RECURSOS EDUCATIVOS EN ABIERTO (REA)	64
3.1. Introducción	64
3.2. Recursos Educativos en Abierto (REA).....	64

3.3. Massive Open Online Courses (MOOC)	75
3.3.1. Definición	75
3.3.2. Historia de los MOOC	79
3.3.3. Taxonomías de MOOC	84
3.4. Language MOOC (LMOOC)	96
3.4.1. Definición y características	96
3.4.2. Principales críticas a los LMOOC	102
3.4.3. Diseño de un LMOOC	104
CAPÍTULO 4: LEARNING ANALYTICS (LA)	117
4.1. Introducción	117
4.2. Definición	118
4.3. Orígenes de LA	124
4.4. Aplicaciones de las LA	126
4.5. Metodología de LA	132
4.5.1. Recopilación y pre-procesado de datos	133
4.5.2. Análisis y acción.....	137
4.5.3. Post-procesado.....	152
4.6. Limitaciones de LA	152
4.7. LA y los LMOOC	159
CAPÍTULO 5: PERFILES DE PARTICIPANTES DE MOOC	162
5.1. Introducción	162
5.2. Perfiles de participantes de MOOC basados en la motivación y la intención	163
5.3. Perfiles de participantes de MOOC basados en la interacción con los elementos del curso	167
5.3.1. Clasificaciones Cualitativas.....	169
5.3.2. Clasificaciones Cuantitativas.....	173
5.3.3. Clasificaciones Mixtas	177

5.4.	Perfiles de participantes de MOOC basados en interacciones sociales	185
5.4	Perfiles temporales de participantes	190
CAPÍTULO 6: DISEÑO DE LA INVESTIGACIÓN		194
6.1.	Introducción	194
6.2.	Recopilación de datos y pre-procesado	195
6.2.1.	Fuentes de datos	196
6.2.2.	Recopilación de datos	206
6.2.3.	Pre-procesado de datos.....	209
6.3.	Análisis y acción	216
6.4.	Post-procesado	222
CAPÍTULO 7: ANÁLISIS DE RESULTADOS		227
7.1.	Introducción	227
7.2.	Identificación del número de grupos	227
7.2.1.	Agrupaciones para el LMOOC PA	228
7.2.2.	Agrupaciones para el LMOOC IP	231
7.2.3.	Comparación de las agrupaciones para los LMOOC IP y PA	232
7.3.	Grupo 1: pre-consumidores	238
7.4.	Grupo 2: participantes exhaustivos	241
7.5.	Grupo 3: participantes ocasionales	242
7.5.1.	Grupo 3a: participantes ocasionales rutinarios	243
7.5.2.	Grupo 3b: usuarios ocasionales flexibles.....	245
7.6.	Grupo 4: usuarios constantes.....	246
7.6.1.	Grupo 4a: participantes constantes rutinarios	248
7.6.2.	Grupo 4b: usuarios constantes flexibles	249
CAPÍTULO 8: CONCLUSIONES		251
8.1.	Introducción	251

8.2. Conclusiones del marco teórico y de la fase de investigación	251
8.3. Hipótesis y preguntas de la investigación.....	254
8.3.1. 1ª pregunta de la investigación.....	254
8.3.2. 2ª pregunta de la investigación.....	264
8.3.3. Confirmación de la hipótesis.....	276
8.4. Limitaciones, aportaciones del estudio y futuras líneas de investigación	278
8.4.1. Limitaciones.....	278
8.4.2. Aportaciones del estudio.....	281
8.4.3. Futuras líneas de investigación	286
BIBLIOGRAFÍA.....	289
ANEXO I. Código fuente del programa Java de pre-procesado	320

LISTADO DE ABREVIATURAS Y SIGLAS

AA	Academic analytics (analíticas académicas)
ABP	Aprendizaje Basado en Proyectos
BBC	British Broadcasting Company
BOOC	Big Open Online Courses
CAI	Computer-assisted instruction o computer-aided
CALI	Computer-assisted language instruction
CALL	Computer-assisted language learning
CBE	Computer-based education
CBI	Computer-based instruction
CBL	Computer-based learning
CCRTVU	Central China Radio and TV University
CEI	Computer-enriched instruction
CMI	Computer-managed instruction
CMS	Course/Content Management System
COOC	Community/Corporate Open Online Courses
DOCC	Distributed Open Collaborative Course
EAO	Enseñanza asistida por ordenador
ECTS	European Credit Transfer and Accumulation System
EDM	Educational Data Mining
gMOOC	Game-Based Learning Massive Open Online Courses
ECCA	Emisora Cultural Canaria
GROOC	Group Massive Open Online Courses
IP	<i>Inglés Profesional</i>
IV	Information Visualization
JSON	JavaScript Object Notation
KNN	k-vecinos más próximos
L1	Primera lengua, lengua materna
L2	Segunda lengua
LA	Learning Analytics
LCMS	Learning Content Management System
LDA	Latent Dirichlet Allocation
LE	Lengua extranjera
LM	Lengua meta
LMOOC	Language Massive Open Online Courses
LMS	Learning Management Systems
LOOC	Little Open Online Courses
MALL	Mobile Assisted Language Learning
MCERL	Marco Común Europeo de Referencia para las Lenguas

MDA	Mediana desviación Absoluta
MLE	Managed learning environment
MOOC	Massive Open Online Courses
MOOC-Ed	Massive Open Online Courses for Educators
MOOR	Massive Open Online Research
NHK	Nippon Hōsō Kyōkai
NOOC	Nano Open Online Courses
OEP	Open Educational Practices
OER	Open Educational Resources
OCW	OpenCourseWare
P2P	peer-to-peer o entre pares
PA	<i>Puertas Abiertas</i>
PLATO	Program Logic for Automated Teaching Operations
pMOOC	Project-Based Massive Open Online Courses
POOC	Participatory/Personalized Open Online Courses
REA	Recursos educativos en abierto
rMOOC	Rhizomatic Massive Open Online Courses
SL	Segunda lengua
SMOOC	Synchronous Massive Open Online Courses
SNA	Social Network Analysis
SPOC	Small Private Online Courses
SPOOC	Self-Paced Open Online Courses
SoLAR	Society for Learning Analytics Research
SQL	Structured Query Language
TBLT	Task-based Language Teaching
TELL	Technology-enhanced language learning
TIC	Tecnologías de la Información y la Comunicación
TICCIT	Time-shared, Interactive, Computer-Controlled, Information Television
UNED	Universidad Nacional de Educación a Distancia
VLE	Virtual learning environment
VOOC	Vocational Open Online Courses

LISTADO DE TABLAS

Tabla 1. Preguntas generales y específicas de la investigación.	26
Tabla 2. Crecimiento del número de grados en línea en el periodo 2017-2021.	83
Tabla 3. Crecimiento del número de micro credenciales en el periodo 2017-2021.	83
Tabla 4. Clasificación de los MOOC según las taxonomías de Lane (2012), Conole (2014), Pili y Admiraal (2016), Castrillo et al. (2018) y Seidametova (2018). Elaboración propia.....	95
Tabla 5. Clasificaciones cualitativas basadas en elementos de interacción. Elaboración propia.....	172
Tabla 6. Clasificaciones cuantitativas basadas en elementos de interacción. Elaboración propia.....	176
Tabla 7. Clasificaciones mixtas basadas en elementos de interacción. Elaboración propia.....	184
Tabla 8. Resumen de las actividades del LMOOC IP. Elaboración propia.	199
Tabla 9. Tabla de contenido del LMOOC Puertas Abiertas I	204
Tabla 10. Tamaño de cada grupo y valores medios de las variables para los datos del LMOOC PA.....	229
Tabla 11. Tamaño de cada grupo y valores medios de las variables para los datos del LMOOC IP.....	231
Tabla 12. Estadísticas de interacción con el LMOOC PA.	233
Tabla 13. Estadísticas de interacción con el LMOOC IP.	233
Tabla 14. Estadísticas de interacción con el LMOOC PA del grupo de pre- consumidores.	239
Tabla 15. Estadísticas de interacción con el LMOOC IP del grupo de pre- consumidores.	239
Tabla 16. Estadísticas relativas a interacción mínima en los cursos para los LMOOC PA e IP	240

Tabla 17. Estadísticas de interacción con el LMOOC PA de los participantes exhaustivos.	241
Tabla 18. Estadísticas de interacción con el LMOOC IP de los participantes exhaustivos.	242
Tabla 19. Estadísticas de interacción con el LMOOC PA de los participantes ocasionales rutinarios.	243
Tabla 20. Estadísticas de interacción con el LMOOC IP de los participantes ocasionales rutinarios.	244
Tabla 21. Estadísticas de interacción con el LMOOC PA de los participantes ocasionales flexibles.	245
Tabla 22. Estadísticas de interacción con el LMOOC IP de los participantes ocasionales flexibles.	245
Tabla 23. Estadísticas de interacción con el LMOOC IP de los participantes constantes.	247
Tabla 24. Estadísticas de interacción con el LMOOC PA de los participantes constantes rutinarios.	248
Tabla 25. Estadísticas de interacción con el LMOOC PA de participantes constantes y flexibles.	249
Tabla 26. Correspondencia de los perfiles de del Peral y el algoritmo k-medias para PA e IP	257
Tabla 27. Estadísticas de pre-consumidores.	259
Tabla 28. Estadísticas de participantes exhaustivos.	260
Tabla 29. Estadísticas de participantes ocasionales rutinarios	260
Tabla 30. Estadísticas de participantes ocasionales flexibles	261
Tabla 31. Estadísticas de participantes constantes del LMOOC IP	261
Tabla 32. Estadísticas de participantes constantes rutinarios del LMOOC PA..	262
Tabla 33. Estadísticas de participantes constantes flexibles del LMOOC PA....	262

LISTADO DE FIGURAS

Figura 1. Principales elementos de la justificación de la tesis doctoral.	24
Figura 2. Elementos definitorios de la educación a distancia	36
Figura 3. Relación entre los términos enseñanza, aprendizaje y educación a distancia. Elaboración propia.....	38
Figura 4. Factores del desarrollo de la educación a distancia.	39
Figura 5. Relación de los distintos términos relacionados con la enseñanza por ordenador. Elaboración propia.....	52
Figura 6. Etapas de CALL y sus características.....	58
Figura 7. Las 5R de los REA.	67
Figura 8. Las cuatro dimensiones de los MOOC. Fuente: Elaboración propia. ..	78
Figura 9. Crecimiento del número de MOOC en el periodo 2012-2022.	82
Figura 10. Distribución de los MOOC en 2021 por temática	84
Figura 11. Criterios para el diseño de los MOOC.	109
Figura 12. Tabla comparativa de las principales disciplinas que realizan análisis de datos educativos. Elaboración propia.....	121
Figura 13. Comparativa de LA y EDM. Elaboración propia.....	123
Figura 14. Objetivos principales de LA.	131
Figura 15. Etapas del proceso de LA.....	132
Figura 16. Tipos de datos educativos. Elaboración propia.	134
Figura 17. Principales riesgos derivados de los datos en proyectos LA. Elaboración propia.....	135
Figura 18. Esquema de los principales métodos y algoritmos de la minería de datos. Elaboración propia	143
Figura 19. Principales grupos de técnicas de análisis y herramientas en LA. Elaboración propia.....	151
Figura 20. Principales elementos de un MOOC. Elaboración propia.	168

Figura 21. Integración de los datos procedentes del LMOOC. elaboración propia.....	213
Figura 22. Operaciones realizadas en la fase de pre-procesado. elaboración propia.....	215
Figura 23. Posición jerarquizada del algoritmo k-medias dentro de los principales métodos de la minería de datos. Elaboración propia.....	217
Figura 24. Flujo de los datos de entrada del programa R. Elaboración propia.	218
Figura 25. Relación de los seis perfiles identificados a partir de los LMOOC PA e IP mediante el algoritmo k-medias. Elaboración propia.	250
Figura 26. Posibles adaptaciones que una identificación temprana de perfiles permite hacer de la ruta de aprendizaje de los participantes de MOOC	286

LISTADO DE IMÁGENES

Imagen 1. Vídeo del primer módulo del LMOOC IP.....	198
Imagen 2. Test de comprensión oral del primer módulo del LMOOC IP.	198
Imagen 3. Sección de progreso del LMOOC IP.....	200
Imagen 4. Ejemplo de navegación en una página del LMOOC Puertas abiertas l.....	202
Imagen 5. Transcripciones multilingües de los textos de los vídeos del curso...	202
Imagen 6. Glosario multilingüe en la sección vocabulario.	203
Imagen 7. Materiales descargables correspondientes al módulo 1.....	203
Imagen 8. Sección de foros de LMOOC PA.....	206

LISTADO DE GRÁFICOS

Gráfico 1. Clústeres obtenidos a partir de la función <i>eclust()</i> en R para los datos del LMOOC PA.....	219
Gráfico 2. Resultado de la función <i>gap statistic</i> en R para los datos del LMOOC PA.....	220
Gráfico 3. Resultado del cálculo del coeficiente de silueta en R para los datos del LMOOC PA.....	221
Gráfico 4. Clústeres obtenidos a partir de la función <i>eclust()</i> en R para los datos del LMOOC IP.....	223
Gráfico 5. Resultado de la función <i>gap statistic</i> en R para los datos del LMOOC IP.....	224
Gráfico 6. Resultado del cálculo del coeficiente de silueta en R con un valor de cinco clústeres para los datos del LMOOC IP.....	225
Gráfico 7. Clústeres obtenidos a partir de la función <i>eclust()</i> con valor 5 en R para los datos del LMOOC IP.....	225
Gráfico 8. Yuxtaposición de los gráficos que representan los clústeres obtenidos a partir de la función <i>eclust()</i> en R para los datos de los LMOOC PA (izquierda) e IP (derecha).....	234
Gráfico 9. Resultado del cálculo del coeficiente de silueta en R para los datos del LMOOC IP.....	236
Gráfico 10. Representación tridimensional de los valores de las muestras de los LMOOC PA (izquierda) e IP (derecha).....	237
Gráfico 11. Yuxtaposición de los distintos perfiles para los datos de PA (izquierda) e IP (derecha).	263
Gráfico 12. Porcentajes de prevalencia de cada perfil de participantes en el estudio original de del Peral (izquierda) y mediante el algoritmo k-medias (derecha).....	266

Gráfico 13. Porcentajes de prevalencia de cada perfil de participantes obtenidos mediante el algoritmo k-medias para PA (izquierda) e IP (derecha).	267
Gráfico 14. Calificaciones medias de cada perfil de participantes de PA (izquierda) e IP (derecha).....	268
Gráfico 15. Porcentajes del nivel educativo de los participantes de PA (izquierda) e IP (derecha).....	269
Gráfico 16. Variación en los tiempos de inicio de sesión para cada perfil de participantes de PA (izquierda) e IP (derecha).....	270
Gráfico 17. Número medio de sesiones de cada perfil de participantes de PA (izquierda) e IP (derecha).....	271
Gráfico 18. Tiempo medio dedicado al curso por los participantes de PA (izquierda) e IP (derecha).....	272
Gráfico 19. Número medio de acceso a los foros para cada perfil de participantes de PA (izquierda) e IP (derecha).....	272
Gráfico 20. Proporción de participantes elegibles para el certificado final para cada perfil de participantes de PA (izquierda) e IP (derecha).....	273

CAPÍTULO 1: INTRODUCCIÓN

1.1. JUSTIFICACIÓN

Las grandes innovaciones producidas en los campos de la informática y las telecomunicaciones desde finales del siglo XX han hecho posible que nuestra sociedad haya experimentado una revolución tecnológica que ha acelerado la transformación de un sistema económico basado en la producción a uno basado en la información (Karoly y Panis, 2004). En consecuencia, la época en la que vivimos es denominada la era de la información, ya que nuestra sociedad actual se caracteriza por el uso intensivo de la información y las comunicaciones. Esto ha provocado profundas modificaciones en nuestro modo de vida, en los procesos de trabajo, en nuestra forma de interaccionar con nuestro entorno y en la naturaleza de nuestras relaciones sociales (Gutiérrez Gutiérrez, 2017). Los desarrollos en el campo de las Tecnologías de la Información y la Comunicación (TIC en adelante) han posibilitado la inmediatez de las comunicaciones humanas, fomentado la interconexión global, tanto en el ámbito personal como en el laboral, y han hecho de las lenguas extranjeras, especialmente del inglés y el español¹, una herramienta fundamental en una sociedad crecientemente interconectada e interdependiente.

El origen de este trabajo de investigación parte del doble perfil académico y profesional de su autor, quien, por una parte, posee una Ingeniería Técnica en Telecomunicaciones en la especialidad de Telemática y cuenta con más de diez años de experiencia profesional en el sector de las TIC y, por otra, posee un Grado en Estudios Ingleses y un Máster de Formación del Profesorado de Educación Secundaria y actualmente forma parte del cuerpo de docentes de secundaria en la especialidad de inglés. Una de las principales motivaciones del autor es poder aplicar al sector de la educación los conocimientos técnicos que ha adquirido en el campo de las TIC, con el objetivo de potenciar la mejora de los procesos de enseñanza-aprendizaje, especialmente en la enseñanza de lenguas extranjeras a distancia.

¹ Según el anuario del Instituto Cervantes (2021), más de 24 millones de alumnos estudiaban español como lengua extranjera en 2021, mientras que el inglés sigue siendo de lejos la lengua más estudiada del mundo, con aproximadamente 1500 millones de estudiantes en todo el mundo.

En consecuencia, el avance en la investigación sobre nuevas tecnologías y metodologías que permitan implementar con éxito las TIC en las aulas es uno de los intereses prioritarios para este autor. Por este motivo, el objeto de este trabajo es profundizar en la aplicación de las TIC en el contexto educativo, más específicamente en la especialidad de la enseñanza de lenguas extranjeras.

La necesidad de profundizar en el estudio de esta área del conocimiento es incuestionable. La revolución tecnológica en el sector de la educación está siendo impulsada por las autoridades educativas y se va introduciendo paulatinamente en las aulas, pero nuestro sistema educativo aún no ha implementado el profundo cambio metodológico que reclama un creciente número de voces dentro del mundo académico (Cabrerizo Aparicio, 2018).

Antes de que irrumpiera la pandemia en 2020, ya existían grandes carencias de integración de las TIC dentro de las escuelas, además de una laguna importante en la formación del profesorado en el uso de las TIC y su aplicación en las aulas. En un estudio llevado a cabo hace seis años por Fernández-Cruz y Fernández-Díaz (2016), a través de una encuesta realizada a 1.433 profesores de 80 colegios de la Comunidad de Madrid, se identificó un perfil promedio de formación docente en TIC de nivel medio-bajo, siendo este nivel mucho inferior entre los profesores que tenían más edad (56-66 años) y mayor experiencia docente.

Dos años más tarde, el informe PISA de 2018 confirmaba las deficiencias que aún existían en cuanto al uso de las TIC en las escuelas. El informe revelaba que solo el 36% de los estudiantes de países de la OCDE asistían a escuelas con un programa específico para promover la colaboración docente en el uso de dispositivos digitales y un 46% a centros con una declaración escrita específica sobre el uso de dispositivos digitales con fines pedagógicos. Estos aspectos organizativos repercuten en el trabajo diario de los docentes, ya que en el mismo informe se alertaba de que alrededor del 45% de los estudiantes de países de la OCDE asistían a escuelas donde los profesores no contaban con incentivos o ayuda para integrar dispositivos digitales en su enseñanza, y 4 de cada diez alumnos tenían un profesor que declaraba no tener suficiente tiempo para preparar las lecciones integrando dispositivos digitales (OECD, 2020).

El cierre masivo de centros educativos a causa de la pandemia causada por la COVID-19 en la primavera de 2020 puso a prueba el nivel de implementación de las TIC en las escuelas de todo el mundo y el grado de desarrollo de las destrezas TIC por parte de los docentes, con resultados poco favorables. Según una encuesta realizada por el Banco Mundial, aunque los sistemas educativos nacionales proporcionaron a los docentes formación especial (el 73%); instrucciones específicas sobre educación a distancia (el 89%); material didáctico para el aprendizaje a distancia (el 80%), o herramientas TIC y conectividad (59%), no existe evidencia de la efectividad de las medidas implementadas (Johns Hopkins et al. 2021).

Por lo tanto, una de las lecciones aprendidas de la pandemia es la necesidad de incrementar la inversión en formación de calidad y el apoyo a los docentes en el uso de las TIC en el aula. Ello requiere cambios estructurales, y el desarrollo de nuevas habilidades pedagógicas y digitales que conviertan a los docentes en nativos de las tecnologías de aprendizaje a distancia disponibles (Johns Hopkins et al. 2021).

Este cambio metodológico en el sector de la enseñanza está siendo liderado por las universidades de todo el mundo. El informe *Trends 2015* (Sursock, 2015), que recoge el resultado de una encuesta realizada en 451 instituciones de 46 países europeos sobre el proceso de enseñanza-aprendizaje, encuentra en las universidades evidencia de un mayor desarrollo de las TIC, de la creciente importancia estratégica de la internacionalización y de numerosas iniciativas dirigidas a aumentar y ampliar la participación y a proporcionar a los estudiantes oportunidades para desarrollar habilidades transferibles.

Respecto a la universidad española en particular, el informe UNIVERSITIC (Gómez, 2017), que recoge los resultados de un cuestionario realizado en 49 universidades españolas sobre el uso de las TIC, destaca cómo la apuesta de las universidades por las TIC como soporte y apoyo a la docencia ha alcanzado cotas tan elevadas que obtiene altos valores en todos los indicadores analizados, además de causar un crecimiento interanual reducido en términos generales debido al menor margen de mejora. El informe revela que «las universidades prestan el 90% de los servicios de soporte TIC a la docencia incluidos en el catálogo. De estos servicios, la docencia virtual, la gestión de licencias *software* para docencia, el soporte a aulas de informática de uso

docente y el soporte a aulas de informática de libre acceso están implantados prácticamente en la totalidad de las universidades participantes en el informe» (p. 17).

Una muestra del liderazgo de las universidades en el área de las innovaciones tecnológicas en educación son los MOOC o *Massive Open Online Courses* (MOOC en adelante). Estos cursos masivos, en línea y abiertos se iniciaron en el seno de la universidad, más específicamente con el curso de 2008 *Connectivism and Connective Knowledge (CCK08)*, diseñado por George Siemens y Stephen Downes para la Universidad de la Isla del Príncipe Eduardo (Medina Salguero y Aguaded Gómez, 2014). Posteriormente, desde comienzos del año 2010, diferentes universidades de reconocido prestigio tales como Stanford, Harvard, MIT, la Universidad de Pennsylvania y la Universidad de Toronto impulsaron la creación de estos cursos (Vázquez Cano y López Meneses, 2014), cuyo número ha crecido de manera exponencial hasta alcanzar, en el año 2021, la cifra acumulada de 220 millones de matriculados en más de 19400 cursos ofrecidos por 950 universidades de todo el mundo (Shah, 2021).

Sin embargo, los MOOC presentan una serie de problemas específicos, siendo uno de los principales sus elevadas tasas de abandono, las cuales oscilan entre el 75 y el 95% (Poy y Gonzales-Aguilar, 2014; Liyanagunawardena, Adams et al., 2013; Jordan, 2014). En los esfuerzos por reducir el número de alumnos que no finalizan los MOOC, las analíticas de aprendizaje o *Learning Analytics* (LA en adelante) se han erigido como una de las herramientas más adecuadas a la hora de incrementar el compromiso de los estudiantes hacia los cursos, ya que las técnicas de LA posibilitan la identificación de puntos de intervención en los que los estudiantes dejan de participar y corren el riesgo de abandonar el curso (Sclater, 2017). Las LA permiten aplicar técnicas analíticas al estudio de datos educativos, de tal forma que sea posible la identificación de patrones de comportamiento entre los alumnos y así mejorar su proceso de aprendizaje (Maselena et al., 2018). Por este motivo, las LA se han postulado como un medio de profundizar en el conocimiento de los MOOC y mejorar su eficacia y, en los últimos años, han generado un creciente interés en el campo educativo y se han convertido en una de las tendencias de futuro en educación más citadas en los informes internacionales (León Urrutia et al., 2017; Holmes et al., 2019).

Una de las aplicaciones de las técnicas de LA es la identificación de perfiles de usuarios en los MOOC. De hecho, uno de los estudios más citados de los últimos años, según el análisis de publicaciones llevado a cabo por Khalil y Ebner (2016), es el realizado por Kizilcec et al. (2013), mediante el cual identificaron cuatro perfiles de usuarios de MOOC basados en el acceso a los recursos de los cursos. Por consiguiente, el uso de LA para analizar ciertas características de los MOOC permite, no solo investigar sobre el papel de las TIC en la educación, sino también utilizar las propias TIC para llevar a cabo el proceso de investigación.

Como consecuencia de lo anteriormente descrito, son cinco los aspectos que justifican la relevancia de este estudio:

1. La aparición de nuevas modalidades de enseñanza de lenguas, como son los recientes MOOC de lenguas o *Language Massive Open Online Courses* (LMOOC en adelante), los cuales suponen un campo de estudio fructífero (Sokolik, 2014). Los LMOOC son un tipo de cursos de lenguas extranjeras en línea (y por tanto a distancia) y abiertos, es decir, disponibles sin ningún coste². Los distintos aspectos de los LMOOC están generando un creciente interés por parte de la literatura científica, como lo demuestran los monográficos editados por Martín-Monje y Bárcena (2014) y por Dixon y Thomas (2015), los cuales se han convertido en obras de referencia en el estudio de los LMOOC, o el trabajo de Beirne (2017), que analiza los enfoques pedagógicos de los LMOOC a través de un cuestionario facilitado a 40 proveedores distintos. Sin embargo, los estudios sobre las características de los participantes en estos cursos y sus hábitos particulares de estudio y de interacción con las plataformas son aún escasos, por lo que investigaciones como la presente son necesarias para profundizar en las dinámicas existentes en los LMOOC y de este modo facilitar el diseño de cursos que incrementen el compromiso de los alumnos y reduzcan las elevadas tasas de abandono.
2. El empleo de las TIC y, más concretamente, el uso de LA, suponen una estrategia idónea para el análisis de los MOOC, ya que un conocimiento

² Aunque la matriculación en los MOOC es siempre gratuita, las plataformas y proveedores de cursos permiten obtener un certificado de finalización del MOOC en la mayoría de los casos sujeto a coste.

más profundo y exacto de las características de la interacción de los participantes de los cursos con los distintos tipos de contenidos permite la implementación y el diseño de modelos de MOOC que se adapten a las preferencias individuales de los estudiantes y acrecienten su grado de compromiso y vinculación. Las LA han experimentado un rápido desarrollo en los últimos años debido a, entre otros motivos, los poderosos métodos y técnicas estadísticos de las LA, que permiten encontrar patrones en los datos y construir modelos predictivos que puedan adaptarse fácilmente a los datos educativos; la presión a la que están sometidas las universidades para reducir las tasas de abandono y mejorar la calidad de los cursos; las crecientes demandas educativas de los países en desarrollo; la relativa facilidad para la generación de datos; la capacidad informática actual, o la posibilidad que ofrecen las LA para controlar la actividad académica con el objetivo de orientar las intervenciones (Calvet Liñán y Juan Pérez, 2015).

3. El creciente interés en el aprendizaje de lenguas extranjeras es un rasgo destacado de la sociedad moderna. El español y el inglés son dos de las lenguas más importantes en el mundo. El anuario del Instituto Cervantes (2021) indica que el español es la lengua materna o primera lengua (L1 en adelante) de 493 millones de personas en todo el mundo, a las que se suman casi 100 millones más de estudiantes de español y de hablantes de competencia limitada, lo que convierte al español en la segunda lengua materna del mundo por número de hablantes, solo por detrás del chino mandarín, y en la segunda lengua mundial en el cómputo global de hablantes, el 7,5% de la población. De hecho, el interés por el estudio del español está en claro auge. En Estados Unidos, el español es la lengua extranjera más estudiada en todos los niveles de enseñanza, mientras que en el Reino Unido el español es percibido como la lengua más importante para el futuro. Dentro de la Unión Europea, Francia, Suecia y Dinamarca destacan especialmente por el estudio del español como lengua extranjera. Esto hace que el español se dispute con el francés y con el chino mandarín el segundo puesto entre las lenguas extranjeras más estudiadas como segunda lengua (L2 o SL en adelante), por detrás del inglés. El inglés, la lengua franca de nuestros días, cuenta con 380 millones

de hablantes nativos, pero estimaciones recientes sitúan su cifra total de hablantes cerca de los 2.000 millones (Crystal, 2008), lo cual la convierte en la lengua indispensable para acceder a un gran número de puestos de trabajo. Según un estudio de 2019 de Spring Professional, consultora del Grupo Adecco, el conocimiento de otras lenguas es una condición prácticamente obligatoria para los perfiles más cualificados dentro del mercado laboral y, más allá del dominio del inglés, cada vez se valora más el conocimiento de una tercera lengua (Fernández, 2019). La importancia del aprendizaje de lenguas extranjeras en nuestra sociedad actual es uno de los motivos por los que se ha elegido para la presente investigación el LMOOC *Puertas Abiertas: curso de español para necesidades inmediatas I* (PA en adelante), un MOOC de español dirigido a personas desplazadas.

4. El aprendizaje de lenguas extranjeras es un objetivo prioritario en una sociedad democrática. En su contribución a la cumbre social celebrada el 17 de noviembre de 2017 en Gotemburgo, la Comisión Europea (2017) expuso la idea de un espacio europeo de educación en el que, para 2025, lo normal sea que todos los ciudadanos europeos hablen dos lenguas extranjeras, además de su lengua materna. De hecho, el preámbulo de la recomendación R (82) 18 del Comité de ministros del Consejo de Europa (1982) incluye tres principios básicos en referencia al aprendizaje de lenguas en Europa:
 - a. Que el rico patrimonio de las distintas lenguas y culturas de Europa constituye un recurso común muy valioso que hay que proteger y desarrollar, y que se hace necesario un importante esfuerzo educativo con el fin de que esa diversidad deje de ser un obstáculo para la comunicación y se convierta en una fuente de enriquecimiento y comprensión mutuos.
 - b. Que solo por medio de un mejor conocimiento de las lenguas europeas modernas será posible facilitar la comunicación y la interacción entre europeos que tienen distintas lenguas maternas con el fin de fomentar la movilidad en Europa, la comprensión mutua y la colaboración, y vencer los prejuicios y la discriminación.

- c. Que los estados miembros, al adoptar o elaborar políticas nacionales en el campo del aprendizaje y la enseñanza de lenguas, pueden conseguir una mayor convergencia a nivel europeo por medio de acuerdos adecuados para una continuada cooperación y coordinación de sus políticas (p. 1).

De hecho, la primera cumbre de jefes de Estado identificó «la reacción violenta de los xenófobos y de los ultranacionalistas como un obstáculo fundamental para la movilidad y la integración europea, así como una amenaza importante para la estabilidad de Europa y para el buen funcionamiento de la democracia» (Consejo de Europa, 2002, p. 3-4). En la segunda cumbre se propuso como objetivo educativo prioritario la preparación para la ciudadanía democrática, otorgando mayor importancia a «propiciar métodos de enseñanza de lenguas modernas que fortalezcan la independencia de pensamiento, de juicio y de acción, combinados con las destrezas sociales y la responsabilidad social» (Consejo de Europa, 2002, p. 4).

5. Otro de los objetivos modernos en educación es el aprendizaje a lo largo de toda la vida, el cual se define como los aprendizajes en todos los espacios y mediante todos los recursos disponibles, combinando aprendizajes en entornos formales, no-formales e informales (UNESCO-UIL, 2010). La UNESCO plantea el aprendizaje a lo largo de la vida como un enfoque para la política educativa y como un nuevo paradigma para la educación en el siglo XXI, ya que supone un medio de mejora de la calidad de vida de las personas y de contribución a la igualdad social (UNESCO, 2020). En este contexto, los MOOC se erigen como un fenómeno de globalización, democratización y apertura del conocimiento que cuenta con la potencialidad de satisfacer un derecho humano fundamental como es el derecho a la educación (Ruiz Bolívar, 2015).

FIGURA 1. PRINCIPALES ELEMENTOS DE LA JUSTIFICACIÓN DE LA TESIS DOCTORAL.



1.2. OBJETIVOS Y PREGUNTAS DE LA INVESTIGACIÓN

El objetivo general de esta investigación es la identificación, dentro de un nuevo LMOOC, de la taxonomía de perfiles de usuarios de LMOOC descrita por del Peral (2019), taxonomía basada en patrones temporales de acceso y uso por parte de los estudiantes. Derivado del primero, un segundo objetivo de este trabajo es el análisis de las posibles diferencias entre los distintos perfiles de participantes en función de las características particulares del nuevo LMOOC. El LMOOC para el que se definió la taxonomía de del Peral fue *Inglés Profesional / Professional English* (IP en adelante), un LMOOC dirigido a estudiantes con ciertos conocimientos de la lengua cuyo objetivo era mejorar su dominio del inglés con fines laborales. El nuevo LMOOC para el que se tratan de identificar los mismos perfiles de la taxonomía de del Peral es PA, un LMOOC dirigido especialmente a personas desplazadas, como refugiados y migrantes, cuyo

objetivo es ayudar a los estudiantes a emplear el español en situaciones comunes, además de enseñar aspectos básicos de la vida y cultura españolas. Las diferencias en el tipo de usuario de ambos LMOOC hacen prever diferencias en la prevalencia de los distintos perfiles.

Por lo tanto, los pasos a seguir en el presente estudio son los siguientes:

1. Partiendo de los perfiles de participantes de LMOOC descritos en el trabajo de del Peral (2019), emplear técnicas de LA con el objetivo de identificar los mismos perfiles en el nuevo LMOOC, PA.
2. Analizar los datos obtenidos, con el fin de determinar si los perfiles identificados en la taxonomía de del Peral son extrapolables al LMOOC PA y, por lo tanto, pueden considerarse comunes a todos los LMOOC.
3. Comparar los rasgos de los distintos perfiles en ambos LMOOC con el fin de determinar si las características propias inherentes de cada uno de los cursos y de sus participantes producen disparidades en la prevalencia y el nivel de desempeño de los diferentes perfiles.

En consecuencia, la hipótesis general que se plantea en este trabajo de investigación es la siguiente: es posible identificar mediante LA la taxonomía de perfiles de usuarios de LMOOC basada en patrones temporales de uso descrita por del Peral (2019) para el LMOOC IP en un nuevo LMOOC, PA, pero los niveles de prevalencia y desempeño variarán en función de las características particulares de cada LMOOC y de sus participantes. El objetivo de la presente investigación es, por lo tanto, la identificación de perfiles de usuario de LMOOC basados en patrones temporales de uso, de modo que se posibilite el diseño metodológico de LMOOC teniendo en cuenta los rasgos de cada tipo de usuario y procurando que los cursos se adapten a las preferencias de aprendizaje específicas de cada tipo de estudiantes.

A partir del objetivo y de la hipótesis de investigación, se deducen una serie de preguntas de investigación generales y específicas que se desglosan en la tabla 1.

TABLA 1. PREGUNTAS GENERALES Y ESPECÍFICAS DE LA INVESTIGACIÓN.

<p>PREGUNTA 1:</p> <p>¿EXISTE UNA TAXONOMÍA DE PERFILES DE PARTICIPANTES DE LMOOC BASADA EN PATRONES TEMPORALES DE USO?</p> <hr/> <ul style="list-style-type: none"> • ¿CUÁNTOS PERFILES PUEDEN IDENTIFICARSE? <hr/> • ¿CUÁLES SON LOS RASGOS IDENTIFICATIVOS DE CADA PERFIL? <hr/> • ¿ES DICHA TAXONOMÍA EXTRAPOLABLE A DIFERENTES LMOOC? <hr/> 	<p>PREGUNTAS GENERALES Y ESPECÍFICAS DE LA INVESTIGACIÓN</p>	<p>PREGUNTA 2:</p> <p>¿EXISTEN DIFERENCIAS EN LOS DISTINTOS PERFILES EN FUNCIÓN DEL LMOOC DEL QUE SE OBTIENEN LOS DATOS?</p> <hr/> <ul style="list-style-type: none"> • ¿QUÉ DIFERENCIAS HAY RESPECTO A ... <ol style="list-style-type: none"> 1. PREVALENCIA DE PERFILES? 2. GRADO DE DESEMPEÑO? 3. TIEMPOS DE ACCESO AL CURSO? 4. NÚMERO DE SESIONES TOTALES EN EL CURSO? 5. TIEMPO TOTAL INVERTIDO EN EL CURSO? 6. ACCESO A LOS FOROS DEL CURSO? 7. TASAS DE ELEGIBILIDAD PARA LA ACREDITACIÓN DEL CURSO? <hr/> • ¿QUÉ RASGOS INHERENTES AL DISEÑO DE LOS CURSOS PRODUCEN TALES DIFERENCIAS? <hr/> • ¿QUÉ RASGOS DEL ALUMNADO DE LOS CURSOS PRODUCEN TALES DIFERENCIAS? <hr/>
---	---	--

1.3. ORGANIZACIÓN Y ESTRUCTURA DE LA TESIS

Además de este primer capítulo introductorio, son tres los bloques principales de contenido en los que pueden dividirse los siete capítulos restantes de esta tesis doctoral: 1) la fundamentación teórica, que comprende los siguientes cuatro capítulos; 2) la descripción de la investigación, con un capítulo dedicado a la metodología empleada y otro al análisis de los datos resultantes, y 3) las conclusiones generales, junto con la respuesta a las preguntas de investigación y la verificación de la hipótesis inicial.

El primer bloque de contenido se corresponde, por lo tanto, con el análisis de la fundamentación teórica y en él se procederá a la revisión del estado de la cuestión de los campos académicos en los cuales se enmarca este trabajo. De este modo, en este bloque se profundizará en aquellos conceptos que son imprescindibles para el desarrollo de la presente investigación: la educación de lenguas a distancia, los MOOC, las LA y el perfilado de participantes de MOOC. Así pues, el segundo capítulo analiza, tanto desde el punto de vista conceptual

como histórico, la enseñanza de lenguas extranjeras como parte del desarrollo de la educación a distancia. Este capítulo explica los distintos términos y conceptos relativos a los procesos de enseñanza-aprendizaje y de la educación a distancia, con atención particular a la enseñanza de segundas lenguas, para posteriormente describir la evolución histórica de la enseñanza de lenguas a distancia y los medios y las metodologías empleados a lo largo de los años en tres principales fases tecnológicas. El tercer capítulo está dedicado a los MOOC, un tipo muy particular de recursos educativos en abierto (REA en adelante). Por lo tanto, este capítulo comienza con una descripción del término general REA, para posteriormente pasar a la definición de los MOOC y el análisis de sus características fundamentales, su evolución histórica y sus correspondientes taxonomías. La parte final del tercer capítulo analiza el tipo específico de MOOC que es objeto de esta tesis, los LMOOC, y describe las características particulares de su diseño. El cuarto capítulo está dedicado a la descripción y el análisis de las características de las LA, ya que es la metodología empleada en esta tesis. Para ello, el capítulo comienza con una necesaria comparación y delimitación de este campo con respecto a otros ámbitos y disciplinas afines para, posteriormente, ofrecer una breve descripción de los orígenes y la evolución histórica de las LA. A continuación, el capítulo analiza distintos aspectos de las LA, tales como sus aplicaciones educativas más importantes, sus principales metodologías y técnicas, o los retos y limitaciones a los que se enfrenta la disciplina. El capítulo concluye con un análisis de los proyectos y aplicaciones más relevantes de las LA dentro del campo de los LMOOC. El objetivo del último capítulo de este bloque es profundizar en las distintas taxonomías que clasifican a los participantes de MOOC en función de rasgos distintivos comunes. Para ello, las taxonomías de participantes de MOOC identificadas en la literatura científica serán divididas en cuatro grandes grupos de perfiles: perfiles basados en la motivación y la intención, perfiles basados en la interacción con los elementos del curso, perfiles basados en interacciones sociales y perfiles temporales, que es el tipo de clasificación en la que se basa este trabajo de investigación.

El segundo bloque de este trabajo, formado por los capítulos sexto y séptimo, describe las distintas fases del proceso de investigación. De este modo, el capítulo sexto está dedicado a la primera de estas fases, la metodología

empleada en el diseño de la investigación, la cual se corresponde con la de los proyectos de LA. Así pues, en este capítulo se detallan las operaciones llevadas a cabo durante los tres pasos de que consta un proyecto de LA. La primera de estas fases es la recopilación y pre-procesado de los datos educativos, por lo que, en primer lugar, se analizarán los dos LMOOC en los que se basa este trabajo y, en segundo lugar, se detallarán las técnicas y las herramientas empleadas para la obtención y el pre-procesado de los datos. La segunda fase de un proyecto de LA es la de análisis, por lo que en el capítulo sexto se describirá la metodología empleada en la fase de análisis de los datos y se explicará la forma en la que se ha aplicado el algoritmo k-medias, el algoritmo de descubrimiento de estructuras seleccionado para identificar los grupos de participantes. Finalmente, en este capítulo se procederá a describir la última fase del proyecto, el post-procesado. El capítulo séptimo analizará los datos obtenidos durante estas tres fases del proyecto de LA, con el objetivo de comparar los perfiles temporales identificados en los dos LMOOC. En consecuencia, en este capítulo se procederá a realizar un análisis de las agrupaciones obtenidas tras la ejecución del algoritmo k-medias y se estudiarán las características de los distintos grupos, con el fin de determinar si los rasgos propios de cada uno de los cursos y de sus participantes producen disparidades en el número y en la prevalencia de los grupos de usuarios y en sus correspondientes niveles de desempeño, como se ha planteado en las preguntas de la investigación.

Una vez finalizadas la interpretación y el análisis de los datos, el tercer y último bloque de la tesis, conformado por el capítulo octavo, está dedicado a la exposición de las conclusiones alcanzadas durante el trabajo de investigación. Este capítulo incluirá tanto las conclusiones derivadas de los dos bloques anteriores como la respuesta a las preguntas de investigación planteadas y la verificación de la hipótesis inicial. Por último, en el octavo capítulo se expondrán las aportaciones de este estudio al campo de la educación a distancia y de los MOOC, y se propondrán futuras líneas de investigación que permitan ampliar los descubrimientos realizados en esta tesis doctoral.

CAPÍTULO 2: APROXIMACIÓN HISTÓRICA A LA ENSEÑANZA-APRENDIZAJE DE LENGUAS A DISTANCIA

2.1. INTRODUCCIÓN

El éxito de la educación a distancia en las últimas décadas se debe principalmente a factores poblacionales y tecnológicos. Por un lado, en los últimos diez años se ha democratizado el acceso a internet y a las TIC en todo el mundo. Para finales del año 2021, 4.900 millones de personas, lo cual representa el 62% de la población mundial, tenían acceso a internet. Esto supone un aumento del 17% desde el año 2019, con 789 nuevos usuarios en tan solo dos años (ITU, 2022). Entre 2013 y 2018, el número de personas suscritas a servicios de telefonía móvil se incrementó en 1.000 millones, hasta alcanzar los 5.100 millones, lo cual representa el 65% de la población y un crecimiento anual del 5%. La estimación es que para el año 2025 se alcancen los 5.800 millones de suscriptores o el 73% de la población (GSM Association, 2019). Estos factores han propiciado que un creciente número de personas opten por la educación a distancia.

Por otro lado, el aumento de la población mundial, que ha pasado de 3.300 millones en 1965 a más de 7.900 millones en la actualidad, junto con las mejoras en los sistemas de educación primaria y secundaria de todo el mundo, ha incrementado significativamente la demanda de educación superior (Qayyum y Zawacki-Richter, 2019; Worldometer, 2022). Como resultado, desde 1995 hasta 2014, las matriculaciones en educación primaria han crecido un 9.5% a nivel mundial, las de secundaria lo han hecho un 34.3%, mientras que las universitarias han aumentado en un 60.9% (UNESCO, 2016).

En Australia, Brasil, Canadá, China, India, Rusia, Sudáfrica, Turquía y Estados Unidos, cerca de la quinta parte de todos los estudiantes universitarios sigue algún tipo de curso a distancia (Qayyum y Zawacki-Richter, 2019). Más particularmente, en Estados Unidos, el número de estudiantes universitarios que estudiaban de forma completamente presencial descendió en 1.737.955 de estudiantes (el 11.2%) entre los años 2012 y 2016, mientras que las inscripciones

en cursos a distancia crecieron durante catorce años consecutivos, alcanzando la cifra de 6.359.121 de estudiantes, el 31.6% del total de estudiantes universitarios (Seaman et al., 2018). En Brasil, las inscripciones en cursos a distancia crecieron un 63.8% anual entre los años 2003 y 2009, y un 9.9% anual desde 2009 a 2014. En Turquía, el crecimiento ha sido de un 20.1% de 2008 a 2014. En Australia, Brasil, Canadá, China, Alemania, India, Rusia, Sudáfrica, Corea del Sur, Turquía, Reino Unido y Estados Unidos, cuyas poblaciones representan el 51% de la población mundial, hay 23 millones de estudiantes universitarios matriculados en cursos a distancia, lo cual representa el 21,3% del total. Destaca el caso de Rusia, donde el 50% de los estudiantes universitarios sigue algún tipo de curso a distancia. En China, las matriculaciones en cursos a distancia crecieron una media del 8.8% anual entre 2004 y 2016 (Qayyum y Zawacki-Richter, 2019). Y en España, antes de la pandemia de 2020, el 15% de los estudiantes universitarios seguía total o parcialmente estudios a distancia (García Aretio, 2021). Este auge de la educación a distancia ha producido un aumento de la financiación, de manera que se prevé que el mercado mundial de la educación en línea llegue a una inversión global de 350.000 millones de dólares en 2025 (Research and Markets, 2019). De estos datos se desprende que el interés por la educación a distancia y la importancia de su papel se han ido incrementando progresivamente en los últimos años, en especial desde que el desarrollo de las TIC ha mejorado la facilidad de acceso a los recursos educativos.

Esta ingente inversión que gobiernos y escuelas han ido realizando a lo largo de los años para fomentar el uso de las TIC en la educación fue sometida a prueba en el año 2020, a raíz de la pandemia provocada por el COVID-19. Esta crisis sanitaria supuso a su vez la mayor crisis educativa de la historia, ya que jamás se había producido un cierre de instalaciones educativas presenciales de tal magnitud y de forma tan abrupta. En abril de 2020, en el pico de la pandemia, la educación presencial de 1.600 millones de estudiantes de primaria y secundaria de más de 190 países fue interrumpida (UNESCO, 16 de diciembre de 2020). Los cierres de las escuelas y otros espacios educativos tuvieron un impacto en el 94% de los estudiantes del mundo, porcentaje que aumentó al 99% en los países con rentas inferiores (UN, 2020). En este contexto, docentes, centros e instituciones educativas, gobiernos y organizaciones

privadas realizaron un descomunal esfuerzo por continuar impartiendo atención educativa a estos millones de estudiantes por todo el mundo mediante distintas tecnologías pertenecientes a la educación a distancia. Operadoras de telecomunicaciones y empresas privadas contribuyeron con medios propios y proporcionaron tanto el ancho de banda requerido como una serie de servicios, aplicaciones informáticas y plataformas necesarios para facilitar la docencia en línea. Los distintos sistemas educativos y sus respectivos centros adaptaron los currículos para impartirlos a través de medios alternativos, especialmente mediante aplicaciones informáticas y plataformas educativas, aunque también se recurrió a clases emitidas en directo por internet, radio y televisión (García Aretio, 2021). El 53% de los países con ingresos altos y el 9% de los de ingresos medios-bajos ofrecieron instrucción exclusivamente a través de internet, mientras que el 85% de los países con ingresos medios-bajos y el 100% de los países con ingresos más bajos optaron por una combinación de dos o más estrategias de aprendizaje a distancia. De este modo, los medios más frecuentemente utilizados por los distintos países fueron la instrucción a través de internet (91%), la televisión (85%), los materiales en papel (82%) y los teléfonos móviles (70%) (Johns Hopkins et al. 2021).

Estas tecnologías empleadas durante la pandemia son el fruto de siglos de desarrollo de la educación a distancia, el cual será el objetivo de este capítulo, con el foco específico en la educación de idiomas a distancia. De este modo, en las dos primeras secciones del capítulo se explicarán los términos y conceptos relativos a los procesos de enseñanza-aprendizaje y de la educación a distancia, con atención particular a la enseñanza de segundas lenguas. En los siguientes tres apartados, se pasará a detallar la evolución histórica de la enseñanza de idiomas a distancia y se describirán las tres etapas principales en las que esta se divide, con atención especial a las metodologías y tecnologías características de cada fase.

2.2. DEFINICIÓN DE ENSEÑANZA-APRENDIZAJE DE IDIOMAS

Históricamente, se ha entendido la enseñanza como la codificación y transmisión de informaciones, con un claro predominio del papel del docente sobre el del discente. Sin embargo, con el desarrollo de la ciencia de la

didáctica este enfoque ha quedado obsoleto, ya que prescinde del otro polo fundamental de la didáctica: el aprendizaje (Tejada Fernández, 2011). De este modo, Mallart i Navarra (2011) define la enseñanza como una actividad humana intencional que aplica el currículum y tiene por objeto el acto didáctico. Consta de la ejecución de estrategias preparadas para la consecución de las metas planificadas y se basa en la influencia de unas personas sobre otras, ya que enseñar es hacer que el alumno aprenda. De hecho, según Zabalza Beraza (1984), enseñar implica hacer posible, facilitar, monitorizar u orientar el establecimiento de relaciones comunicacionales con las personas. Tejada Fernández (2011) identifica tres rasgos distintivos de la enseñanza entendida como acto didáctico:

1. **Intencional y sistemática.** La enseñanza es un acto intencional, ya que su objetivo es producir el aprendizaje, y dicha intencionalidad la aporta el docente mediante la aplicación de sus creencias y competencias pedagógicas y la lectura de las situaciones específicas de enseñanza en las que interviene.
2. **Contextualizada.** El acto didáctico de la enseñanza se produce en un espacio social con relaciones sociales directas entre alumno y profesor, e indirectas dentro del contexto institucional.
3. **Compleja.** Desde la concepción original de enseñanza como transmisión de información, se han ido integrando cada vez más componentes en la acción de enseñar y la han transformado en un acto complejo.

Los conceptos de enseñanza y aprendizaje están estrechamente relacionados, aunque esta relación no sea una de causa-efecto, ya que existe aprendizaje sin enseñanza formal y enseñanza formal sin que se logre el aprendizaje (González Ornelas, 2001). Fenstermacher (1989) señala que enseñanza y aprendizaje son dos conceptos que difieren claramente, pero entre los que existe una dependencia ontológica o una relación de carácter semántico, ya que sin estudiantes no tendríamos el concepto de profesores y viceversa. De esta forma, existe una gama de actividades vinculadas con la condición de alumno que complementan las actividades de la enseñanza. Por ejemplo, los profesores explican, describen, definen, refieren, corrigen y estimulan; los alumnos, por su parte, repiten, practican, piden ayuda, repasan,

controlan, sitúan fuentes y buscan materiales de estudio. La relación ontológica del aprendizaje y la enseñanza yace en el hecho de que aprendizaje y enseñanza son conceptos complementarios, ya que el aprendizaje establece la dirección de la enseñanza. De este modo, la función principal del docente no es solo enseñar, sino hacer posible el aprendizaje de los alumnos. Aprendizaje y enseñanza son por tanto dos procesos distintos que los profesores tratan de integrar en uno solo: el proceso de enseñanza-aprendizaje (González Ornelas, 2001).

Aprendizaje es el proceso mediante el cual se origina o se modifica un comportamiento o se adquiere un conocimiento de una forma más o menos permanente (Mallart i Navarra, 2011). Es un proceso de adquisición cognoscitiva que consiste en una secuencia de acciones encaminadas a la construcción del conocimiento, al desarrollo de habilidades y a la formación de actitudes. En este proceso intervienen procesos psicológicos básicos y habilidades cognitivas, conocimientos específicos relativos al tema de trabajo o de aprendizaje, estrategias de aprendizaje y procedimientos, y el conocimiento de los propios procesos psicológicos implicados en la realización de una actividad. El proceso de aprendizaje aparece enmarcado dentro de una sociedad y cultura específicas, lo cual implica una apropiación de saberes e instrumentos culturales y una aportación personal desde la construcción individual y social como miembro de dicha cultura (Gómez, 1996). De acuerdo con Zabalza Beraza (1984), el aprendizaje se basa en el establecimiento por parte del alumno de relaciones con personas y situaciones, lo cual produce socialización y desarrollo experiencial; con ideas, acontecimientos, información cultural y técnica, lo cual produce desarrollo instructivo y culturización, y con objetos y su manejo, lo cual produce aprendizajes técnicos.

Según Díaz Alcaraz (2002), los nuevos planteamientos del aprendizaje han supuesto una reconceptualización del papel didáctico del docente, que pasa de ser la persona que enseña a ser la persona que facilita el aprendizaje. Se produce por tanto un cambio desde una concepción del aprendizaje como producto a una del aprendizaje como proceso. De este modo, se define el aprendizaje como el conjunto de estrategias de procesamiento de información, adaptación a las situaciones de aprendizaje y control de las propias

capacidades y efectos a través de las cuales los sujetos aprenden. La tarea didáctica ya no consiste solo en enseñar, sino en crear las condiciones para que los alumnos aprendan (Mallart i Navarra, 2011).

La disciplina que se ocupa del estudio de los procesos de enseñanza-aprendizaje es la didáctica, una ciencia auxiliar de la pedagogía (Tejada Fernández, 2011). Como indica Pastor Cesteros (2004), mientras que la didáctica general está enfocada al «establecimiento de técnicas y procedimientos de enseñanza vinculados directamente con el desarrollo del papel de profesor y alumnos en el aula» (p. 63), existen también didácticas específicas, entre las que se encuentra la didáctica de las lenguas extranjeras. Esta última concreta las técnicas que aseguran el conocimiento y uso de una lengua no nativa. Los procesos de enseñanza-aprendizaje de segundas lenguas son también el objeto de estudio de la lingüística aplicada, una rama de la lingüística. Esta disciplina estudia, por una parte, los fenómenos de adquisición de segundas lenguas desde una perspectiva psicolingüística y, por otra, los procesos de enseñanza-aprendizaje de segundas lenguas desde una perspectiva tanto lingüística como psicolingüística, sociolingüística y pedagógica (Pastor Cesteros, 2004).

Dentro del aprendizaje de idiomas es importante establecer la distinción entre los conceptos de adquisición y aprendizaje. Krashen (1981), en su influyente teoría de la monitorización, establece las diferencias fundamentales entre ambos. Adquisición es el proceso subconsciente a través del cual llegamos a conocer nuestra L1, mientras que aprendizaje es un proceso consciente que nos conduce a conocer una L2. Según esta distinción, las L1 se adquieren, mientras que las L2 se aprenden. De este modo y como indica Pastor Cesteros (2004), las L1 son aquellas que el individuo ha adquirido en primer lugar, con las cuales ha aprendido a hablar y a relacionarse con su entorno. Los individuos monolingües solo adquieren una L1, mientras que los individuos bilingües o trilingües adquieren respectivamente dos o tres lenguas al mismo tiempo. Las lenguas posteriores que un individuo llega a conocer una vez que ha adquirido su o sus L1 se derivan, por lo tanto, de un proceso de aprendizaje y se denominan lenguas extranjeras (LE en adelante) o L2. Aunque estos términos se utilizan en muchas ocasiones de modo indistinto, existe una diferenciación entre el aprendizaje de una lengua no nativa en contexto de LE

o en contexto de L2. Así pues, se define LE cuando la lengua no nativa que se aprende no está presente en el contexto en el que el hablante se desenvuelve diariamente, ya sea porque este pertenece a una comunidad monolingüe o porque la lengua de estudio no es una de las que se hablan dentro de la comunidad bilingüe a la que el hablante pertenece. Por el contrario, cuando la lengua se halla en el contexto inmediato del aprendiz y este puede entrar en contacto con ella fuera del aula, nos encontramos ante una situación de L2, tal como sucede en comunidades en las que coexisten dos o más lenguas. Por consiguiente, el proceso de enseñanza-aprendizaje de lenguas podrá ir dirigido tanto a estudiantes de LE como a estudiantes de L2 y las características particulares de cada tipo de estudiantes será un factor a tener en cuenta en el proceso de enseñanza-aprendizaje (Pastor Cesteros, 2004).

2.3. DEFINICIÓN DE EDUCACIÓN A DISTANCIA

Es difícil encontrar una definición absoluta del término «educación a distancia», ya que a lo largo de las décadas se han propuesto distintas definiciones que han ido incorporando al término las características de los últimos avances técnicos que permiten este tipo de educación. Desmond Keegan, en su obra *The Foundations of Distance Education* (1996) dedica un capítulo entero a la definición del término «educación a distancia» y analiza hasta ocho definiciones acuñadas a lo largo de tres décadas. De ellas infiere los siguientes seis elementos definitorios de la educación a distancia:

- La separación entre el alumno y el profesor, lo cual la distingue de las clases presenciales.
- La influencia de una organización educativa, lo cual la distingue del estudio particular.
- El uso de medios técnicos que unen al profesor y al alumno y contienen el contenido educativo.
- La provisión de un canal bidireccional de comunicación que permite que el estudiante se beneficie del diálogo e incluso lo pueda iniciar.
- La posibilidad de reuniones ocasionales con fines didácticos y de socialización.

- La participación en una forma industrializada de educación perteneciente a una categoría que separa de forma radical la educación a distancia de otros tipos dentro del espectro educativo (p.44).

FIGURA 2. ELEMENTOS DEFINITORIOS DE LA EDUCACIÓN A DISTANCIA



Nota. Adaptado de *The Foundations of Distance Education*, por D. Keegan, 1996. *The Foundations of Distance Education*. Copyright 1986, 1990, 1996 by Desmond Keegan.

Tomando en cuenta estos elementos, una de las definiciones más claras y concisas del término «educación a distancia» es la de Kaplan y Haenlein (2016, p.443), quienes lo definen como «cualquier forma que permita proporcionar educación a estudiantes que están separados por la distancia (es decir, que no están físicamente presentes en el mismo espacio) y en los que una institución educativa planifica y prepara el material pedagógico». A su vez, García Aretio, uno de los principales investigadores en España en el campo de la educación a distancia, afirma que «la educación a distancia se basa en un diálogo didáctico mediado entre el profesor (institución) y el estudiante que, ubicado en un espacio diferente al de aquel, puede aprender de forma independiente

y también colaborativa» (2011). Keegan (1996) considera, por lo tanto, «educación a distancia» como un «término genérico que comprende la variedad de estrategias de enseñanza-aprendizaje usadas por universidades a distancia, universidades abiertas, departamentos a distancia de universidades convencionales y unidades de formación a distancia de entidades y empresas» (p. 34).

A lo largo de los años, el término «educación a distancia» ha desplazado a otros significantes que carecían de precisión o que ignoraban alguno de los elementos definitorios de este tipo de educación. De este modo, el término «educación por correspondencia», utilizado de forma común en los inicios de la educación a distancia, ha quedado completamente obsoleto al haberse superado los medios técnicos a los que el término hace referencia; el término «estudio en casa», que fue común en los Estados Unidos para referirse a cursos de formación técnica y profesional, es inexacto, ya que la educación puede llevarse a cabo de forma parcial en centros distintos del hogar del estudiante; el término «estudio independiente», muy frecuente en los Estados Unidos para programas universitarios a distancia, tampoco es preciso, ya que sugiere independencia de la institución educativa. Los términos «enseñanza a distancia» y «aprendizaje a distancia» se presentan a su vez como problemáticos, ya que solamente ponen el foco en una parte del proceso de enseñanza-aprendizaje. Mientras que el término «enseñanza a distancia», que adquirió gran popularidad a raíz de la creación de la *Open University* en Reino Unido, describe la mitad del proceso, al centrarse en el papel de la institución e ignorar el del estudiante, el término «aprendizaje a distancia» invierte el foco y describe el proceso desde el punto de vista del estudiante, obviando a la institución educativa (Keegan, 1996). Como consecuencia a lo anteriormente expuesto, se concluye que el término «educación a distancia» es el más adecuado, al comprender la suma de ambos conceptos, el de la enseñanza y el aprendizaje, y al describir de forma más completa y exacta el fenómeno que se define.

FIGURA 3. RELACIÓN ENTRE LOS TÉRMINOS ENSEÑANZA, APRENDIZAJE Y EDUCACIÓN A DISTANCIA. ELABORACIÓN PROPIA.



Asimismo, la educación a distancia puede dividirse en dos grupos: la educación a distancia asíncrona y síncrona. La educación a distancia asíncrona se refiere a la circunstancia en que el proceso de enseñanza y el de aprendizaje se producen en distintos momentos en el tiempo, de tal modo que los estudiantes pueden aprender a su propio ritmo y de acuerdo con sus horarios. Por el contrario, en la educación a distancia síncrona o simultánea, el proceso de enseñanza y aprendizaje coincide en el tiempo, aunque las personas del docente y del discente no compartan el mismo espacio físico Kaplan y Haenlein (2016).

Según García Aretio (1999), son cinco los principales factores que han contribuido al desarrollo de la educación a distancia. El primero de ellos son los avances sociopolíticos logrados a lo largo del siglo XX, los cuales produjeron un aumento de la demanda social de educación. Un sector de la población previamente excluido de la educación, el cual incluía a adultos empleados; amas de casa; reclusos; residentes de zonas con escasas infraestructuras, y personas hospitalizadas o discapacitadas, comenzó a exigir la democratización de los estudios, lo cual contribuyó a la masificación de las aulas convencionales y al deterioro de la calidad de la educación. En consecuencia, fue necesaria la creación y adecuación de instituciones educativas que pudiesen satisfacer esta creciente demanda. El segundo factor es la creciente necesidad de aprender a lo largo de la vida, que hace que trabajadores a los que les es imposible acudir presencialmente a las aulas puedan combinar sus empleos con sus necesidades educativas de perfeccionamiento, reciclaje, cualificación y recualificación. El tercer factor, consecuencia de los dos anteriores, es la carestía de los sistemas educativos convencionales. Si un mayor número de individuos demandan el

acceso a la educación a lo largo de toda su vida, la formación presencial se torna inviable desde una perspectiva económica, mientras que la educación a distancia garantiza la extensión de una educación de calidad. El cuarto factor es el avance en las ciencias de la educación y de la psicología, que posibilitan una metodología que potencia el trabajo independiente y la personalización del aprendizaje sin la presencia directa del profesor, y que permite alcanzar un aprendizaje personal, reflexivo y significativo. Por último, los avances tecnológicos en el campo de las TIC han permitido la incorporación de equipos de grabación, reproducción y transmisión de texto, sonido e imagen en el campo educativo, lo cual posibilita tanto la comunicación vertical profesor-estudiante, como la horizontal entre participantes.

FIGURA 4. FACTORES DEL DESARROLLO DE LA EDUCACIÓN A DISTANCIA.



Nota. Adaptado de *Historia de la Educación a Distancia*, por L. García, 1999. RIED. Revista Iberoamericana de Educación a Distancia, 2(1). Copyright 1999 por Lorenzo García Aretio.

La historia del desarrollo de la educación a distancia puede dividirse en tres etapas distintas, vinculadas a tres grandes generaciones de innovación tecnológica que Garrison (1985) identifica como correspondencia, telecomunicación y telemática. Estas tres fases serán el objeto de análisis de la parte restante del presente capítulo, junto con las respectivas corrientes metodológicas de enseñanza-aprendizaje de lenguas extranjeras. De este modo, el siguiente apartado explicará el origen de la enseñanza de lenguas a

distancia y cubrirá las dos primeras etapas de Garrison: la enseñanza de idiomas por correspondencia, y mediante la radio y la televisión. El último apartado del capítulo se centrará en la enseñanza de lenguas asistida por ordenador o, en inglés, *Computer-assisted Language Learning* (CALL en adelante), la cual se corresponde con la tercera etapa de Garrison.

2.4. INICIOS DE LA ENSEÑANZA DE LENGUAS A DISTANCIA

El método de aprendizaje de idiomas generalizado en el siglo XIX era el método de gramática-traducción, el cual fue empleado originalmente para el estudio de los idiomas clásicos griego y latín (Chang, 2011). La metodología empleada se basaba en la traducción de frases de la lengua meta (LM en adelante) a la L1, con énfasis principal en el aprendizaje deductivo de las reglas gramaticales y en el vocabulario de la lengua estudiada (Kong, 2011). El objetivo de este método era el desarrollo de las destrezas de comprensión y expresión escritas, ignorando de forma sistemática las destrezas orales y utilizando la lengua materna como medio de instrucción (Richards y Rodgers, 1986).

La combinación de material escrito y del sistema postal fue el primer medio empleado en la enseñanza a distancia (Garrison, 1985). La enseñanza por correspondencia comenzó a popularizarse a lo largo del siglo XIX a manos de instituciones privadas. En 1856, el francés Charles Toussaint y el alemán Gustav Langenscheidt fundaron el *Institut Toussaint et Langenscheidt*, el primer instituto para la enseñanza de lenguas extranjeras por correspondencia, el cual se considera la primera institución de enseñanza auténticamente por correspondencia. Toussaint y Langenscheidt fueron los primeros en impartir clases con material diseñado para el autoestudio de personas adultas, mediante el método de gramática-traducción (García Aretio, 1999). El éxito del método fue tal que se extendió a un total de catorce lenguas y fue adoptado por editores fuera de Alemania, quienes tradujeron y adaptaron las unidades a otros idiomas (Methode Toussaint-Langenscheidt, 2020)

En los Estados Unidos, una de las primeras instituciones en impartir educación a distancia fue el *Chautauqua Institute* de Nueva York, cuyo origen se encontraba en las escuelas dominicales metodistas. En 1879 se creó una

escuela de idiomas dependiente del instituto y se estableció un programa que consistía en unas primeras clases presenciales seguidas de cartas semanales con ejercicios de lectura y traducción que se devolvían corregidos (Noffsinger, 1926).

A finales del siglo XIX, un conjunto de lingüistas europeos estableció nuevos principios para el aprendizaje de lenguas extranjeras que llegaron a conocerse como el enfoque natural. Su objetivo era replicar el proceso natural por el que los humanos aprenden las lenguas maternas. Estos principios fueron la base del método directo, el más conocido dentro de todos los métodos que usaron un enfoque natural y el cual comenzó a usarse en Europa a finales del siglo XIX. En las clases que seguían este método, las cuales se impartían exclusivamente en la LM, se aprendía la gramática de forma inductiva y se enseñaba solo vocabulario y frases de uso cotidiano, mediante objetos, ilustraciones o mímica. Al contrario que el método de gramática-traducción, el método directo daba una gran importancia a las destrezas orales, por lo que se trabajaba el desarrollo de la comprensión y la expresión oral y se daba atención a la correcta pronunciación de las palabras. Para la década de los años 20, el uso de este método en Europa decayó y fue remplazado por metodologías mixtas que usaban algunos de los principios y técnicas del método directo junto con la enseñanza explícita de gramática (Richards y Rodgers, 1986).

El método directo influyó en las metodologías usadas por las escuelas privadas a distancia tanto en Europa como en Estados Unidos, principalmente gracias al empleo del fonógrafo. A partir de 1905, el método Toussaint-Langenscheidt se complementó con discos para gramófono que permitían trabajar la pronunciación correcta del material y, en 1912, siguiendo el principio de dar prioridad al aprendizaje de vocabulario de uso cotidiano, se crearon libros de frases para viajeros (*Methode Toussaint-Langenscheidt*, 2020). En Francia, en los años previos a la I Guerra Mundial, aparecieron los primeros cursos privados de idiomas por correspondencia dirigidos al público adulto, los cuales presentaban claras carencias pedagógicas. A partir de los años 20, estos cursos se perfeccionaron, incluyendo principios del método directo a través del gramófono (Bourrel et al., 2008).

En los Estados Unidos, a finales del siglo XIX, se fueron creando las primeras instituciones privadas por correspondencia, las cuales fueron agrupándose en las décadas posteriores. En 1907, *La International Correspondence Schools* de Scranton, en Pennsylvania, estaba formada por 31 escuelas que ofrecían, con un coste de 100 dólares, cursos de idiomas similares a los del método Toussaint-Langenscheidt que incluían un fonógrafo, 25 discos y cuatro libros de texto (Watkinson, 1996). En 1926 se fundó el *National Home Study Council*, una institución federal de carácter privado que en 1931 aglutinaba a 36 centros privados de educación a distancia que ofertaban cursos de francés, alemán, español, latín y griego (Noffsinger, 1931).

La segunda etapa en la historia de la educación a distancia es aquella que utiliza la transmisión de mensajes a través de señales analógicas, mediante tecnologías como la radio, el teléfono o la televisión. De hecho, la radio fue el primer medio de telecomunicación usado de forma masiva en la enseñanza de lenguas extranjeras. La radio permitió cubrir las necesidades educativas de todos aquellos que no podían desplazarse a un centro presencial, en particular enfermos, ciudadanos confinados en instituciones, habitantes de poblaciones sin infraestructuras o ciudadanos cuyas solicitudes de matriculación eran rechazadas por las universidades (Demiray e İşman, 2003).

En los Estados Unidos, el primer curso radiofónico de idiomas fue ofrecido durante el curso de 1921-1922 por el Decano Charles Ernest Fay Wade, profesor de idiomas modernos del *Tufts College* de Massachusetts. La División de Extensión Universitaria del Departamento Estatal de Educación de Massachusetts ofreció desde el curso de 1923-1924 al curso de 1927-1928 cursos de idiomas extranjeros divididos en ocho lecciones, que contaban con el apoyo de material escrito para el estudio individual y culminaban en exámenes para la obtención del título correspondiente (Federal Writers' Project, 1939). También la Universidad de Wisconsin programó en 1931 un curso radiofónico de español con clases quincenales, con un primer trimestre dedicado exclusivamente a la gramática y un segundo a la lectura. El curso se apoyaba en libros de texto y constaba de ejercicios de repetición de frases y conversaciones, dictados y reproducción de clases fonográficas (Bolinger, 1934; García Aretio, 1999). Durante la década de 1920 y 1930, las escuelas de primaria y secundaria estadounidenses también

comenzaron a utilizar la radio como material complementario a las clases de lenguas extranjeras (Tyler, 1933). Algunos ejemplos fueron el curso semanal de español de Evansville, Indiana, o el curso de francés emitido por las escuelas de Cleveland (UNESCO, 1949).

En 1927, un curso de francés experimental a manos de la *British Broadcasting Company* (BBC en adelante) se convirtió en el primer curso de idiomas emitido en la radio británica. En 1930 se crearon cursos regulares de francés y alemán como complemento de las clases presenciales en las escuelas. Los cursos de francés y alemán estaban divididos en dos niveles. El primero de ellos se llamaba *Early Stages* y estaba diseñado para estudiantes con al menos un año de instrucción en el idioma. Cada lección contenía repetición de pasajes cortos leídos por el locutor, práctica de entonación, juegos escritos, canciones sencillas y diálogos representados por niños y niñas nativos. El segundo nivel se llamaba *Talks and Dialogues* y estaba dirigido a la mejora de la pronunciación y de la comprensión oral de los estudiantes de tercer o cuarto año mediante charlas sobre asuntos cotidianos. También incluía escenas dramatizadas y canciones, así como actividades de ampliación de vocabulario con la ayuda de cuadernillos (Parker, 1937; UNESCO, 1949). Aunque los cursos estaban principalmente dirigidos a escolares y diseñados para ellos, la BBC recibió numerosas solicitudes por parte del público general mostrando interés por adquirir los cuadernillos complementarios a los cursos. El número de cuadernillos vendidos muestra que, para el año 1935, cerca de 20.000 alumnos seguían el curso de francés inicial en 630 escuelas, mientras que el de alemán inicial era seguido por unos 6.000 alumnos de 162 escuelas (Parker, 1937). Para finales de la década de los 30, eran 10.000 las escuelas que usaban las retransmisiones de la BBC como complemento de las clases presenciales (Demiray e İşman, 2003).

Algunos países que comenzaron a utilizar la radio en la enseñanza de lenguas extranjeras durante la década de 1930 fueron Canadá, Australia, Brasil y Suecia. En Canadá, fueron varios los cursos radiofónicos para escolares que fueron seguidos por estudiantes adultos. En 1933, la Universidad de Alberta organizó un curso radiofónico experimental de fonética francesa dirigido a estudiantes de secundaria que despertó un gran interés por parte del público

adulto, el cual llegaba a adelantar la hora de finalización de su jornada para poder seguirlo. (Mirth, 1996). Durante el curso de 1948-1949, se emitió para las escuelas de Ontario un curso de francés que, aunque iba dirigido a un público escolar, tuvo una mayor acogida entre oyentes adultos. De manera adicional, en 1949 se comenzó a retransmitir con éxito un programa de aprendizaje de francés titulado *Écoutez* (Fleming y Toutant, 2008). En Australia, la *Australian Broadcasting Commission* inició en 1933 emisiones en escuelas de Nueva Gales del Sur y retransmitió cursos de idiomas, con especial predominio del francés (UNESCO, 1949). El número de escuelas que usaron las emisiones creció exponencialmente durante las décadas de 1930 y 1940, pasando de 760 escuelas en 1935 a 5.868 escuelas (el 55% del total) y 400.000 alumnos en 1948. En Brasil, el Ministerio de Educación estableció en 1936 una programación exclusivamente educativa de 14 horas que incluía cursos de portugués, francés, inglés y español. Los cursos estaban divididos en lecciones de media hora a modo de dramatizaciones para el nivel de primaria, diálogos para el nivel de secundaria y clases magistrales para el nivel universitario (UNESCO, 1949). En septiembre de 1929, en Suecia y a través de la emisora nacional *Radiotjänst*, el Consejo de Educación comenzó a emitir en cada año escolar cursos de noruego, danés, inglés, francés y alemán compuestos de charlas de 30 minutos que paulatinamente fueron sustituyéndose por diálogos entre hablantes nativos, los cuales los alumnos repetían. Los cursos se acompañaron de folletos con las ilustraciones en las que se basaban las emisiones, listas de vocabulario y transcripciones de los diálogos. Entre los años 1945 y 1951, Suecia experimentó con un curso de inglés para estudiantes de primaria que combinaba instrucción por correspondencia, por radio y de forma directa, y para el cual los profesores recibieron formación metodológica (UNESCO, 1949).

Desde los años posteriores a la II Guerra Mundial y hasta la década de 1960, el método enseñanza de idiomas que se popularizó en los Estados Unidos fue el método audio lingual, basado en los principios conductistas. Aunque este método tenía como objetivo el desarrollo de las destrezas orales, especialmente la expresión, favorecía la competencia lingüística sobre la competencia comunicativa. El método audio lingual empleaba ejercicios de entrenamiento, respuestas grupales y memorización de diálogos en aras de la adquisición secuencial de las distintas destrezas de la L2, siguiendo el siguiente orden:

comprensión oral, expresión oral, comprensión escrita y expresión escrita. Los cursos radiofónicos de esta época se adaptaron fácilmente a esta metodología, ya que permitían desarrollar la primera de las destrezas, la comprensión oral, junto con la repetición de diálogos y los ejercicios de entrenamiento (Nelson, y Wood, 1975; Richards y Rodgers, 1986). Por otra parte, el método de enseñanza de idiomas que se utilizó durante la década de los 50 y 60 en Reino Unido es el denominado enfoque oral o situacional, el cual presentaba numerosas similitudes con el método audio lingual. El enfoque oral está basado en principios sistemáticos de selección, gradación y presentación de los contenidos de un curso de idiomas, lo cual implica la utilización de listas de vocabulario y la selección de estructuras gramaticales. La práctica incluye repeticiones guiadas colectivas, ejercicios de entrenamiento y dictados mediante los cuales se busca alcanzar la exactitud en la pronunciación y en el uso de las estructuras gramaticales. (Richards y Rodgers, 1986).

En los años posteriores a la II Guerra Mundial, numerosos países utilizaron la radio en emisiones para países extranjeros. Algunos de los motivos fueron la satisfacción de las necesidades de alumnos en el extranjero, la cooperación en programas internacionales o la enseñanza de lenguas con fines políticos. De este modo, en 1942, el gobierno de los Estados Unidos fundó la emisora internacional *Voice of America*, con el objetivo de combatir la propaganda nazi (Voice of America, s.f.). En 1959, *Voice of America* comenzó a emitir material para el aprendizaje del inglés bajo el título *Special English* (Voice of America Learning English, s.f.). Entre el material se incluían emisiones adaptadas a un nivel principiante, con un vocabulario limitado a 800 palabras y una pronunciación clara y lenta que facilitaba la comprensión por parte de los estudiantes (Nelson, y Wood, 1975). En 1943, la BBC creó un departamento exclusivo dedicado a la enseñanza de la lengua inglesa, inicialmente denominado *English by Radio* y que actualmente continúa bajo el nombre de *BBC Learning English* (BBC, s.f.). Durante las siguientes décadas, y con la cooperación de los departamentos universitarios de inglés y lingüística y del British Council, este departamento creó numerosos cursos radiofónicos de inglés, que a su vez fueron adaptados a varios idiomas de la Commonwealth, entre ellos hindi, bembá, chichewa, tonga, fijiano, luganda, luo, malayo, maltés, pastún, hausa y urdu. Los cursos cubrían los distintos niveles de la lengua general, aunque también se diseñaron cursos

especializados de inglés profesional en los ámbitos del comercio, transporte aéreo, navegación, periodismo, medicina o turismo. También en la década de los 40, la *Canadian Broadcasting Corporation* participó en proyectos internacionales, preparando cursos radiofónicos de idiomas extranjeros para países del centro y el este de Europa (UNESCO, 1949). En los años 70, el servicio internacional de la radio sueca ofrecía cursos de sueco para alumnos en el extranjero, junto con folletos complementarios gratuitos que los alumnos podían enviar de vuelta para ser corregidos, además de grabaciones en discos fonográficos. En los años 70, otros países como Holanda, Sudáfrica, Japón y Checoslovaquia también ofrecieron, mediante los servicios internacionales de sus radios, cursos similares de los idiomas respectivos: holandés, afrikáans, japonés y checo, (Nelson, y Wood, 1975).

Una de las iniciativas radiofónicas más destacadas en la enseñanza de lenguas extranjeras en España proviene de la Emisora Cultural Canaria (ECCA en adelante), constituida en 1963. Radio ECCA creó un sistema educativo propio que conjuga esquemas impresos con clases radiofónicas y tutorías, tanto presenciales como a distancia. Radio ECCA, la cual sigue actualmente en activo, retransmitió en 1967 su primer curso de lengua extranjera, *Aula Abierta: inglés* (Radio ECCA, s.f.)

Dentro del uso de la televisión en la enseñanza de idiomas, una de las instituciones más destacadas es la BBC. Durante la década de 1970, el departamento de enseñanza de la lengua inglesa de la BBC cambió su nombre a *BBC English by Radio and Television* y a partir de 1962 comenzó a producir cursos de inglés para televisión, llegando a producir seis cursos distintos en los siguientes diez años. Tres de ellos fueron de inglés general; uno fue de inglés científico y tecnológico; otro fue de enseñanza de inglés como segunda lengua, y el último, ya en color, fue de inglés para negocios e industria (Stephens, 1972). Otros países que utilizaron de manera exitosa la televisión en la enseñanza de idiomas fueron China y Japón. En 1960 se fundó la facultad de educación mediante televisión de Pekín, la cual ofrecía un curso del idioma chino que combinaba el medio escrito con la televisión, y en los siguientes años se crearon facultades de modalidad televisiva en Shanghái, Cantón, Harbin y Shenyang. Sin embargo, la llamada Revolución Cultural de 1966 provocó el cierre de las

universidades hasta 1972 (Hawkridge y McCormick, 1983). En 1979 se creó la *Central China Radio and TV University* (CCRTVU en adelante), que ofrecía formación universitaria a través de la radio y la televisión, y en la cual, durante sus 40 años de historia, se han matriculado 19 millones de alumnos, 14,4 millones de los cuales han llegado a graduarse (He y Yuan, 2019). Desde su fundación, uno de los principales objetivos de la CCRTVU fue la producción de cursos de inglés que condujeran a la obtención de titulaciones universitarias, para lo cual colaboró estrechamente con instituciones de todo el mundo, muy especialmente con la *Open University* del Reino Unido (Xiao, 2008). Actualmente, la CCRTVU ostenta el nombre de *Open University of China* y se ha convertido en la mayor institución de educación a distancia del mundo (He y Yuan, 2019). En Japón, la emisora de radiodifusión pública *Nippon Hōsō Kyōkai* (NHK en adelante) fue creada en 1953 siguiendo el modelo de la BBC y ya en 1955 comenzó a emitir los primeros cursos radiofónicos de idiomas en inglés, francés, alemán y chino. En 1959, la NHK comenzó a emitir también cursos televisivos de inglés, francés y alemán; cursos de ruso y chino en 1973, y de coreano en 1984. Los cursos de inglés se usaron como complemento al currículo de las escuelas, ya que el inglés es una asignatura obligatoria, y adicionalmente se diseñaron cursos específicos para la preparación de los exámenes de inglés de acceso al último ciclo de secundaria y a la universidad (Abe, 2004)

2.5. ENSEÑANZA DE LENGUAS ASISTIDA POR ORDENADOR (CALL)

La tercera etapa en la historia de la educación a distancia está caracterizada por el uso de la informática y, a partir de los años 90, de internet. Warschauer (1996) divide, a su vez, esta generación en tres periodos distintos, en función de la metodología o de las teorías de aprendizaje en que están basados los programas y aplicaciones desarrollados. Son los periodos conductista, comunicativo e integrativo. El inicio de una de estas fases no implica un rechazo de los programas y métodos de la fase anterior, sino que cada periodo va ganando prominencia paulatinamente e incorporando las características más relevantes de la etapa previa.

Relacionados con esta tercera generación dentro de la educación a distancia, han surgido una multitud de términos y acrónimos que

frecuentemente llevan a una gran confusión. Por ese motivo, en el siguiente subapartado se pasará a definir los términos más frecuentemente empleados, tras lo cual se describirá cada uno de los tres periodos enunciados previamente en subsiguientes subsecciones.

2.5.1. DEFINICIÓN DE TÉRMINOS

Con el inicio del uso de los ordenadores con fines educativos, surgieron una multitud de términos que definían procesos de enseñanza y de aprendizaje a través de la informática. Algunos de estos términos son: *Computer-based Education* (CBE en adelante); *Computer-based Instruction* (CBI en adelante); *computer-based Learning* (CBL en adelante); *Computer-assisted Instruction* o *Computer-aided Instruction* (CAI en adelante) o, en español, enseñanza asistida por ordenador (EAO en adelante); *Computer-managed Instruction* (CMI en adelante), o *Computer-enriched Instruction* (CEI en adelante). Cotton (2008) proporciona las siguientes distinciones para estos términos:

- CBE, CBI y CBL son los términos más generales y se refieren a prácticamente cualquier tipo de uso del ordenador en entornos educativos. Estos términos se refieren tanto al uso autónomo del ordenador para actividades de aprendizaje como a actividades por ordenador que refuerzan el material introducido y enseñado por un profesor. Actualmente estos términos incluyen todo tipo de dispositivos electrónicos usados en una clase, desde ordenadores portátiles hasta teléfonos móviles, los cuales permiten interactuar con componentes gráficos, de audio y de texto de manera simultánea (Gu et al., 2016).
- CAI o EAO es un término más restringido y puede definirse como la interacción de los estudiantes, durante el proceso de aprendizaje, con cursos diseñados para ser seguidos por ordenador, en los cuales el ordenador proporciona el entorno de aprendizaje y el profesor actúa como guía en el proceso (Şeker, 2013). CAI normalmente se refiere a actividades de simulación, tutoriales o ejercicios de práctica y entrenamiento ofrecidos de forma aislada o como complemento de la instrucción tradicional por parte de un profesor.

- CMI comprende aquellas actividades en las que un ordenador evalúa el rendimiento de los alumnos en pruebas, les dirige a recursos educativos pertinentes y guarda un registro de su progreso.
- CEI se define como aquellas actividades en las que los ordenadores generan datos a petición de los alumnos para ilustrar relaciones en modelos relativos a la realidad física o social, o bien proporcionan enriquecimiento general en ejercicios relativamente desestructurados diseñados con el fin de estimular y motivar a los estudiantes.

En enseñanza de idiomas, se acuñaron diversos términos para referirse al uso de ordenadores como medio de instrucción. *Computer-assisted Language Instruction* (CALI en adelante) se derivó del término genérico CAL y en 1982 fue incorporado al nombre de la asociación profesional estadounidense CALICO. El término *Computer-assisted Language Learning* (CALL en adelante) se originó en Reino Unido y es el más frecuentemente empleado en la actualidad. Para 1982 ya era usado ampliamente en publicaciones y en 1983 la asociación de profesores de inglés como lengua extranjera TESOL adoptó el término para una de sus secciones. En los años 80 surgió otro término alternativo, *Technology-Enhanced Language Learning* (TELL en adelante), que también fue adoptado por diversas organizaciones y revistas científicas (Davies et al., 2013).

Como se analizará posteriormente, el surgimiento de internet supuso una enorme revolución tecnológica, lo cual generó otra oleada de términos relativos al aprendizaje en línea. A finales de la década de 1990 se popularizó rápidamente el concepto de *e-learning*, que la Comisión Europea define como «el uso de nuevas tecnologías multimedia e internet para mejorar la calidad del aprendizaje, facilitando acceso a recursos y servicios, así como intercambios y colaboración remota» (2001, p.2). Rosenberg (2001), define *e-learning* como la utilización de las tecnologías de Internet para ofrecer un conjunto de propuestas que permitan incrementar el conocimiento y la práctica. Según Area y Adell (2009), *e-learning* es «una modalidad de enseñanza-aprendizaje que consiste en el diseño, puesta en práctica y evaluación de un curso o plan formativo desarrollado a través de redes de ordenadores y puede definirse como una educación o formación ofrecida a individuos que están geográficamente dispersos o separados o que interactúan en tiempos diferidos del docente

empleando los recursos informáticos y de telecomunicaciones» (p. 392). Para estos autores, lo característico del e-learning es que «el proceso formativo tiene lugar totalmente o en parte a través de una especie de aula o entorno virtual en el cual tiene lugar la interacción profesor-alumno, así como las actividades de los estudiantes con los materiales de aprendizaje» (p. 392). Barberá (2008) considera que el e-learning se basa en «un proceso formativo que debe ser planificado, organizado, seguido, apoyado y valorado sirviéndose de los medios necesarios para hacer posible una comunicación lo más completa posible entre aprendiz y docente» (p.13) y afirma que este medio necesario es normalmente un ordenador conectado a una red, aunque cada vez más se usan otros dispositivos con conexión inalámbrica. Vemos, por lo tanto, que el uso de internet y las redes informáticas es un concepto clave del *e-learning*, lo cual lo distingue de otros conceptos como el CAI o el CBI.

Uno de los conceptos asociados al *e-learning* que apareció en la década de 1990 es el de los *Virtual Learning Environments* (VLE en adelante). Según Britain y Liber (2004), los VLE son sistemas *software* de gestión del aprendizaje que sintetizan la funcionalidad de los programas informáticos de comunicaciones y de los métodos en línea de distribución de materiales educativos. Según Dillenbourg et al. (2002), los VLE son espacios de información que poseen un diseño y que integran tecnologías heterogéneas y diversos enfoques pedagógicos. Son espacios sociales, ya que dentro del entorno se producen interacciones educativas y los estudiantes desempeñan un papel clave en su construcción. Asociado al concepto de VLE aparecen los términos *Managed Learning Environment* (MLE en adelante) y campus virtual, los cuales se considerarían términos hiperónimo e hipónimo, respectivamente. Weller (2007) indica que los MLE abarcan todos los sistemas pertenecientes a una institución universitaria. Esto incluye, además de los enfocados al proceso de aprendizaje, los sistemas administrativos y de registros de datos de los estudiantes. Los campus virtuales, por su parte, son un subgrupo dentro de los VLE. El objetivo de los campus virtuales es proporcionar cursos universitarios, normalmente como parte de un programa conducente a un título universitario. Los campos virtuales presentan por tanto unas restricciones de edad o de nivel que no tienen los VLE, los cuales pueden ser usados para configurar partes más

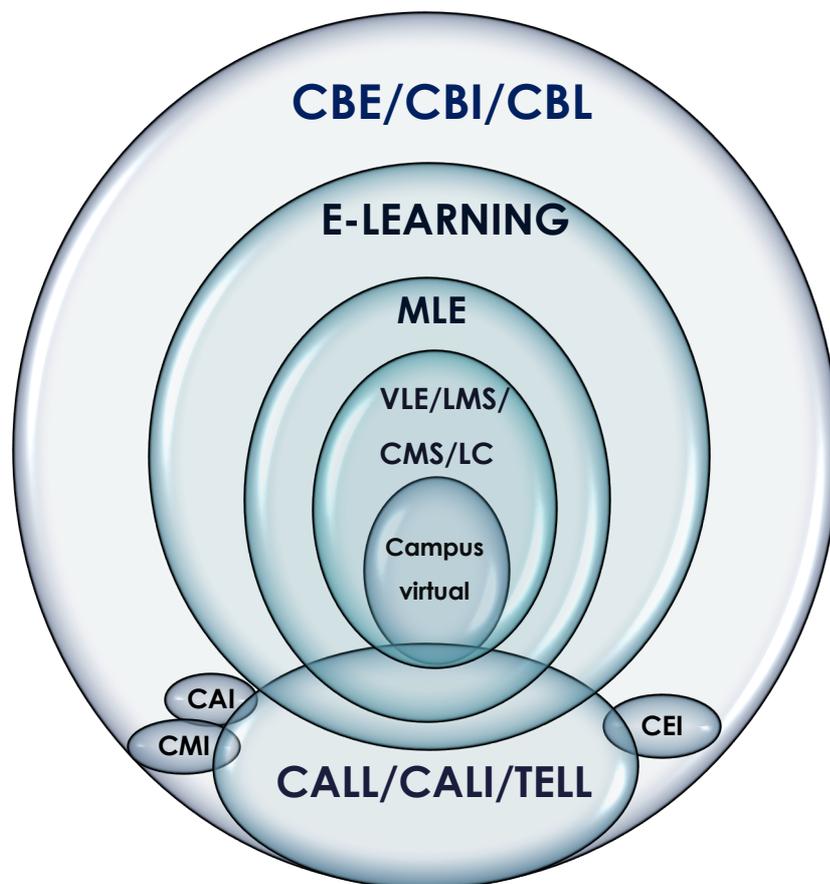
reducidas de un currículo. Por consiguiente, los campus virtuales se consideran un subconjunto de los VLE, que a su vez lo son de los MLE.

Los *Learning Management Systems* (LMS en adelante) son plataformas web diseñadas para la gestión, documentación, monitorización, registro de información y distribución de cursos, tanto para la educación superior como de otros tipos. Pueden proporcionar apoyo a las clases tradicionales, a la educación a distancia o una combinación de las dos (Ghilay, 2019). Weller (2007) considera que VLE y LMS son sinónimos, siendo LMS el término más usado en Estados Unidos y VLE el preferido en Europa, y los define como sistemas de *software* que combinan distintas herramientas con el objetivo de proporcionar contenido en línea y facilitar la experiencia de aprendizaje en torno a dicho contenido.

Otro término que se puede considerar sinónimo de VLE y LMS es el de *Course Management System* (CMS en adelante), el cual Morgan (2003) define como un sistema *software* diseñado y comercializado específicamente para la utilización por parte de docentes y estudiantes en el proceso de enseñanza y aprendizaje. La mayoría de los CMS permiten la organización y presentación del contenido del curso e incluyen herramientas para la comunicación y la evaluación de estudiantes y funciones para gestionar el material de clase y las actividades. Pero el acrónimo CMS también puede significar *Content Management System*, o sistema de gestión de contenidos en español, el cual se define como un sistema de métodos y técnicas para recoger, gestionar y publicar contenido en una empresa por medio de las TIC (Benevolo y Negri, 2007). Este último significado de CMS puede aplicarse a contenido de cualquier tipo, no solo educativo, por lo que a veces se emplea el término derivado *Learning Content Management System* (LCMS en adelante) para referirse, bien a un sistema que simplemente almacena y gestiona contenido relacionado con el aprendizaje, bien como sinónimo de VLE o LMS (Weller, 2007).

En la figura 5 se muestra una relación de los términos relacionados con la enseñanza de lenguas por ordenador que han sido analizados en esta sección.

FIGURA 5. RELACIÓN DE LOS DISTINTOS TÉRMINOS RELACIONADOS CON LA ENSEÑANZA POR ORDENADOR. ELABORACIÓN PROPIA.



2.5.2. CALL CONDUCTISTA

El origen del CALL se remonta a la década de 1950, con la utilización para el aprendizaje de idiomas de los primeros ordenadores: las unidades centrales, computadoras usadas por grandes organizaciones, entre las cuales se encontraban las universidades. A estas máquinas de enormes dimensiones se conectaban una serie de terminales con los que los estudiantes interactuaban a través de programas para el aprendizaje de idiomas (Daniel, 1999). La primera fase de CALL, por lo tanto, se extiende desde la década de 1950 hasta la década de 1970, y se denomina fase conductista, ya que estaba basada en las teorías conductistas, aún predominantes en la época, y en metodologías tales como el método audio lingual (Warschauer, 1996). En la época de 1950, se empezó a experimentar con el uso con fines educativos de los ordenadores y se desarrollaron programas básicos que intentaban replicar el conocimiento del

profesor y que mostraban poca flexibilidad en la dirección del proceso educativo (Fischetti y Gisolfi, 1990). Los primeros programas para el aprendizaje de idiomas estaban concebidos para ejercitar el vocabulario y la gramática mediante ejercicios de entrenamiento por repetición. Además, informaban de los errores de manera inmediata y mantenían un registro del progreso. La utilización de estos programas permitía liberar tiempo de instrucción en la clase para dedicarlo a tareas más comunicativas. (Davies et al., 2013).

Uno de los primeros proyectos educativos con ordenadores se desarrolló entre las décadas de 1950 y 1960, cuando educadores de la Universidad de Stanford en California colaboraron con IBM para introducir CAI en escuelas de secundaria (Arnold, 1997). En los años 70 hubo dos proyectos rivales a gran escala en el área de CAI: *Program Logic for Automated Teaching Operations* (PLATO en adelante) y *Time-shared, Interactive, Computer-Controlled, Information Television* (TICCIT en adelante). El sistema PLATO comenzó a implementarse en la Universidad de Illinois en 1960 y para 1970 había alcanzado una fase avanzada de desarrollo. El sistema estaba formado por un potente ordenador central que se comunicaba con hasta un millar de terminales más pequeños a través de la línea telefónica (Gibbons y O'Neal, 2014). El sistema ofrecía un gran número de servicios, entre los cuales se incluían tutoriales, pruebas, correo, mensajería electrónica, foros y juegos multijugador (Otto, 2017). Mediante el lenguaje de programación TUTOR, se desarrolló material en PLATO para el estudio de numerosos idiomas, entre ellos inglés, francés, alemán, hebreo, chino, latín, ruso, español, hindi, suajili y sueco (Davies et al., 2013). Para 1985, había cien sistemas PLATO operando en los Estados Unidos, que en tan solo siete años llegaron a contabilizar 40 millones de horas totales de uso educativo (Arnold, 1997). El proyecto TICCIT fue creado por la Fundación Nacional para la Ciencia de Estados Unidos con el objetivo de desarrollar sistemas de educación en ordenadores pequeños (Gibbons y O'Neal, 2014). El sistema TICCIT era un sistema de instrucción informático que combinaba la televisión y el ordenador y el cual se considera uno de los primeros sistemas de *courseware*. La tecnología de *software* y *hardware* inicial fue desarrollada por la corporación MITRE en McLean, Virginia, y sobre ella científicos de la Universidad de Brigham Young y de MITRE diseñaron un innovador sistema de *software*. Es sistema TICCIT difería de otros tipos de CAI en su enfoque hacia el

aprendizaje de conceptos y reglas mediante estrategias seleccionadas por el estudiante, en lugar de presentar simplemente problemas o ejercicios de entrenamiento. El primer prototipo de programa fueron cursos de matemáticas y de inglés. Posteriormente se desarrollaron cursos de francés, alemán y español, tanto para enseñanza secundaria como para enseñanza universitaria (Merrill et al., 1979).

El impacto de las aplicaciones CALL aumentó exponencialmente a finales de la década de 1970, cuando se comenzaron a fabricar los primeros ordenadores asequibles para el gran público y estos empezaron a usarse tanto en hogares como en instituciones educativas (Davies et al., 2013). En 1980, solo un 5% de las escuelas de primaria en Estados Unidos y un 20% de las escuelas de secundaria disponían de ordenadores como método complementario a la instrucción. Sin embargo, en 1983 ambos porcentajes se habían cuadruplicado y, para el final de la década de 1980, prácticamente todos los centros educativos de los Estados Unidos y de la mayoría de los países industrializados contaban con equipos informáticos con fines educativos (Arnold, 1997). Sin embargo, la mayoría de los programas creados durante la década de 1980 aún no se basaban en el enfoque comunicativo, sino que seguían siendo programas de entrenamiento de gramática y vocabulario a través de ejercicios. Los procesadores de texto y las hojas de cálculo también experimentaron un gran crecimiento durante la década y los profesores de idiomas las comenzaron a utilizar en sus clases de idiomas de forma innovadora para crear fichas de ejercicios. El programa para ordenador *LingoFox*, por ejemplo, procesaba textos auténticos y a partir de ellos permitía compilar diversos tipos de ejercicios de gramática, vocabulario, ortografía o comprensión escrita.

2.5.3. CALL COMUNICATIVO

A finales de la década de 1960, tanto el método audio lingual en Estados Unidos como el método oral o situacional británico habían sido rechazados por numerosos lingüistas. Un grupo de lingüistas británicos abanderaron el cambio hacia una metodología cuyo objetivo fuera la competencia comunicativa en lugar del dominio de las estructuras lingüísticas. Al mismo tiempo, el Consejo de Europa se propuso impulsar la enseñanza de idiomas en el continente, por lo

que surgió la necesidad de desarrollar métodos alternativos (Richards y Rodgers, 1986). Wilkins (1973) propuso una definición comunicativa de los idiomas basada en dos tipos de significado: categorías nocionales y categorías comunicativas. El Consejo de Europa incorporó los conceptos desarrollados por Wilkins en sus especificaciones para la enseñanza de idiomas, lo cual fue uno de los factores determinantes para el desarrollo de lo que se llegó a conocer como el enfoque comunicativo. El objetivo de este enfoque no es tanto alcanzar la corrección en las formas lingüísticas como la adquisición de la competencia comunicativa en las destrezas de comprensión oral y escrita, y expresión oral y escrita, mediante actividades que repliquen situaciones reales y que utilicen procesos comunicativos tales como la negociación, la interacción o el intercambio de información (Richards y Rodgers, 1986).

En la década de 1980, los impulsores del CALL comunicativo desincentivaron los ejercicios de práctica y entrenamiento característicos de la fase de CALL conductista, ya que no promovían comunicación auténtica. Por el contrario, lo que hicieron fue poner el foco en el uso de las formas del lenguaje en lugar de en las propias formas y en la enseñanza de la gramática de forma implícita. De esta forma animaban a los estudiantes a realizar producciones orales originales en la LM (Singh, 2015).

Los primeros intentos de CALL siguiendo el enfoque comunicativo pueden rastrearse, por lo tanto, en la década de los 80, particularmente en el desarrollo de juegos, programas de reconstrucción de texto y de lectura espaciada. Los programas desarrollados durante esta época permitían la práctica de las distintas destrezas de la lengua sin recurrir a los ejercicios de práctica (Singh, 2015). Aunque estos programas suponen una extensión del modelo en el que el ordenador cumple la función de tutor y contiene la respuesta correcta a los problemas planteados, al contrario que en los programas de práctica y entrenamiento, el proceso que el estudiante debe seguir para encontrar dicha respuesta presenta un mayor grado de libertad de elección, control e interacción. En otros casos, el propósito de la actividad planteada por el programa no es tanto el dar con una respuesta correcta, sino estimular las capacidades de conversación, expresión escrita o pensamiento crítico de los estudiantes (Warschauer, 1996). Algunos de los ejemplos más interesantes a este

respecto son los juegos *Spion* y *Juegos Comunicativos*. Estos programas, además del enfoque comunicativo, utilizaban la inteligencia artificial y la gamificación en la enseñanza del alemán (Sanders, 1984; Feustle Jr, 1987). Ambos juegos empleaban análisis semántico y sintáctico, lo cual permitía al usuario introducir texto en lenguaje real y que el programa lo pudiera procesar.

También en esta década se generalizó el uso de herramientas que, si bien no se diseñaron específicamente para el aprendizaje de idiomas, ofrecían a los estudiantes la oportunidad de usar y comprender la LE. Además de los procesadores de texto, que ya aparecieron en la fase anterior, comenzaron a usarse verificadores de gramática y de ortografía, y programas de edición y de concordancia (Warschauer, 1996). Los programas de concordancia permitían realizar búsquedas de conjuntos de palabras en corpus electrónicos de textos y de este modo analizar las palabras dentro de su propio contexto (Otto, 2017). En la década de 1980 también se fundaron las primeras asociaciones CALL: CALICO en los Estados Unidos en 1982 y EUROCALL en Europa en 1986 (Davies et al., 2013).

2.5.4. CALL INTEGRATIVO

En la década de 1990 se produjo la transición de una visión cognitiva a una perspectiva socio-cognitiva en el enfoque comunicativo del aprendizaje y la enseñanza de idiomas, lo cual hizo que se cuestionara la idoneidad de enseñar las distintas destrezas de modo compartimentado, tal como se hacía en el CALL comunicativo (Singh, 2015). Bajo el nuevo enfoque, el objetivo de los educadores era integrar en el aprendizaje de idiomas las diferentes destrezas lingüísticas (comprensión oral y escrita, y expresión oral y escrita) mediante la tecnología, recurriendo a contextos y situaciones extraídos del mundo real. Este cambio de paradigma coincidió en el tiempo con dos avances tecnológicos: los ordenadores multimedia y, más particularmente, el desarrollo de internet, lo cual propició los objetivos del nuevo enfoque integrativo (Tafazoli y Golshan, 2014).

Dentro de este cambio de paradigma se encuadran el enfoque por tareas o, en inglés, *task-based language teaching* (TBLT en adelante) y el aprendizaje basado en proyectos (ABP en adelante), metodologías que

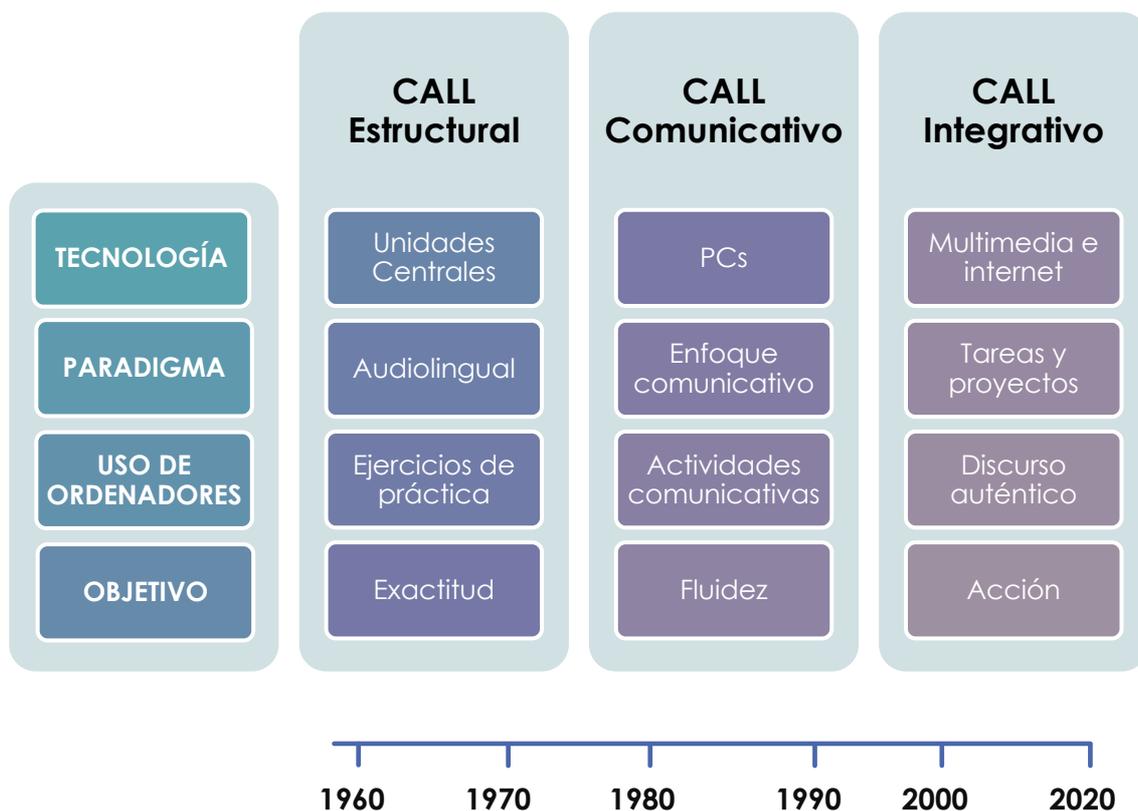
proporcionan a los estudiantes una mayor oportunidad de interacción social real y auténtica, ya que replican el tipo de situaciones comunicativas que se producen en el mundo real (Singh, 2015).

Según Martín Peris (2008), en torno a 1990 el TBLT surgió como evolución de los enfoques comunicativos en el mundo anglosajón. El TBLT es un programa de aprendizaje de lenguas cuyo objetivo es fomentar el aprendizaje mediante el uso real de la lengua en el aula, para lo cual los procesos de aprendizaje deben incluir necesariamente procesos de comunicación. Una de las bases del TBLT consiste en la distinción entre contenidos necesarios para la comunicación y procesos de comunicación. Ello está en consonancia con los postulados del análisis del discurso, según los cuales la comunicación requiere de la adecuada interpretación del sentido de los mensajes, para lo cual es imprescindible el conocimiento del contexto. La otra base del modelo procede de la psicología del aprendizaje y de los estudios de adquisición de lenguas, los cuales defienden que los mecanismos conducentes a la capacidad de uso de la lengua consisten necesariamente en el ejercicio de ese uso. Por este motivo, las unidades del TBLT consisten en actividades de uso de la lengua y no en estructuras sintácticas, como hacían los métodos audiolinguales, ni en nociones y funciones, como hacían los programas nociofuncionales.

Según Martín Peris (2008), el ABP de lenguas extranjeras se aproxima al TBLT, ya que el producto final se constituye a la vez en el objetivo de un conjunto de actividades y tareas, secuenciadas y orientadas a su obtención, y en el medio para el aprendizaje de la lengua. En consecuencia, en la realización de todas las actividades del proyecto es necesario utilizar la lengua que se aprende, activando todas las destrezas lingüísticas y aprendiendo gramática y vocabulario a medida que las necesidades prácticas de desarrollo del proyecto lo vayan exigiendo. El objetivo del aprendizaje por proyectos es potenciar el aprendizaje participativo, la motivación, la relación con el contexto social en que tiene lugar el aprendizaje, la globalidad del aprendizaje y la transversalidad de los contenidos.

En la figura 6 se muestra un resumen de las principales características de cada una de las tres etapas de CALL, junto con una línea temporal que indica de forma aproximada su momento de aparición.

FIGURA 6. ETAPAS DE CALL Y SUS CARACTERÍSTICAS



Nota. Adaptado de *Technological change and the future of CALL*, por M. Warschauer, 2004. En S. Fotos y C. Brown (Eds.). *New perspectives on CALL for second and foreign language classrooms*, (pp. 15-25). Copyright 2004 by Copyright Holder.

En la década de 1990, la capacidad de los ordenadores y de los sistemas de almacenamiento experimentaron una enorme evolución y el disquete dio paso al CD-ROM. Esto permitió el desarrollo de programas multimedia que permitían la reproducción de sonido y de vídeo de calidad creciente. La principal ventaja de la tecnología multimedia consistía en la posibilidad de acceder desde una sola máquina a una gran variedad de medios tales como texto, gráficos, sonido, animación o vídeo. De manera adicional, mediante el uso de hipertexto fue posible enlazar varios tipos de contenido multimedia, lo cual proporcionó ventajas tales como la creación de entornos auténticos que

combinaban lo auditivo y lo visual, la integración de destrezas en una sola actividad o la atención al contenido sin sacrificar la forma ni las estrategias de aprendizaje. Otro de los beneficios del desarrollo de hipermedia fue el fomento de la autorregulación del proceso de aprendizaje de los estudiantes, ya que estos comenzaron a tener acceso a una gran variedad de enlaces con explicaciones gramaticales, glosarios de vocabulario, transcripciones fonéticas y otros elementos que les animaban a adaptar su estrategia de aprendizaje (Warschauer, 1996). Estos programas, que se hicieron enormemente populares en la década de 1990, se conocieron comúnmente como programas tutoriales, ya que el ordenador reemplazaba al profesor o al libro de texto y proporcionaba instrucciones y ejercicios para que los estudiantes pudieran practicar las distintas destrezas de forma autónoma (Gimeno-Sanz, 2016).

Fueron numerosas las universidades que comenzaron a desarrollar programas en CD-ROM durante la década de 1990. Por ejemplo, en 1992 cinco universidades europeas fundaron el consorcio CAMILLE con el objetivo de desarrollar cursos de idiomas en CD-ROM. Una de las instituciones participantes fue la Universidad Politécnica de Valencia, la cual creó dos programas, uno de nivel inicial llamado *Español Interactivo* y otro de nivel inicial-intermedio llamado *Español en marcha*. Ambos programas combinaban vídeo, audio, texto e ilustraciones, y contenían gran variedad de ejercicios interactivos, además de un diccionario bilingüe con reproducción de sonido; libros de referencia con hipertexto sobre cultura española; funciones de lenguaje y gramática; un informe de evaluación detallado, y soporte para los idiomas inglés, francés y alemán. En 1994, la Universidad de Wolverhampton desarrolló un programa multimedia interactivo de inglés de negocios en seis partes bajo el título de *English for Business*. Entre 1996 y 1997, el consorcio TELL, liderado por la Universidad de Hull, creó la serie *Encounters*, formada por cinco conjuntos de CD-ROM y disponible para el aprendizaje de francés, alemán, español, italiano y portugués. Durante esta época también existieron numerosos ejemplos de programas producidos por empresas comerciales, tales como *Talk to Me* (posteriormente renombrado a *Tell Me More*), publicado por la empresa francesa Auralog en 1992; *English Discoveries/English +*, publicado in 1994 por la empresa americana Edusoft Ltd; *Triple Play Plus!*, un curso de inglés publicado en 1995 por la empresa americana Syracuse Language Systems Inc, o los

distintos programas de inglés de la serie *Talk* producidos por la empresa británica Libra Multimedia: *Telephone Talk* y *Small Talk* (1993), *Travel Talk* (1994), *Business Talk* (1995) y *City Talk* (2000) (Gimeno-Sanz, 2016). A finales de los años 90 surgieron los primeros programas de aprendizaje de idiomas que incorporaban reconocimiento automático del habla (Davies et al., 2013).

Sin embargo, pese al enorme potencial de la tecnología multimedia, el impacto que produjo fue limitado. Ello se debió, entre otras razones, a la falta de inteligencia e interacción auténtica del *software* y a la dificultad por parte de los profesores para producir sus propios programas, lo cual dejaba en manos de desarrolladores comerciales, con carencias en formación pedagógica, la producción de contenidos (Warschauer, 1996). Las carencias que presentaban los programas multimedia para ordenador fueron superadas con la llegada a los hogares de internet. Internet se abrió al público en 1993 con el lanzamiento de *Mosaic*, el primer navegador del mercado, seguido en 1994 de *Netscape*. En un principio, la mayoría de las páginas web ofrecían únicamente texto y solo una pequeña parte de ellas lo combinaban con imágenes. Sin embargo, con la implementación del audio y el vídeo en los sitios web, las posibilidades de interacción se incrementaron drásticamente (Davies et al., 2013).

Al inicio del siglo XXI, se produjeron dos avances en internet que resultaron ser de una enorme importancia en el CALL integrativo. Por un lado, la calidad del audio y el vídeo en internet mejoró considerablemente, por lo que comenzaron a aparecer cursos completos de idiomas por internet, muchas veces ofrecidos por entidades comerciales o como parte de proyectos de gobiernos o universidades. Por otro lado, se desarrolló la Web 2.0., lo cual dotó a internet de grandes capacidades integrativas, así como del potencial de convertirse en una plataforma social para la colaboración, la compartición de conocimiento y la creación de redes (Davies et al., 2013). A su vez, se fueron desarrollando herramientas cada vez más completas que permitían a los profesores crear sus propios materiales. Una de las más destacadas fue *Hot Potatoes*, una web diseñada en 1998 por Stewart Arneil y Martin Holmes, de la Universidad de Victoria en Canadá, que facilitaba la compilación de ejercicios de respuesta múltiple, de completar huecos, ordenar frases, enlazar con la opción correcta e incluso la creación de crucigramas (Arneil y Holmes, 2001;

Gimeno-Sanz, 2016). En 2000, la Universidad Politécnica de Valencia creó *InGenio*, la primera herramienta de creación de contenidos CALL completamente en línea. A partir de plantillas, *InGenio* permitía a los profesores de idiomas crear materiales para las clases, los cuales a su vez podían agruparse en bases de datos compartibles con otros usuarios e incluso usarse para crear cursos de forma automática. Otra herramienta de creación de contenidos basada en plantillas fue I4LL (*Integrated Interactive Independent Internet-based Language Learning*), un programa diseñado por la Universidad de Gante, Bélgica, en 2001, como parte del proyecto WELCOME, financiado con fondos de la Unión Europea. *MALTED7* fue otra herramienta del mismo tipo, lanzada en 2004 y financiada por el Ministerio de Educación de España. La herramienta incluía ejercicios de asociación, escritura libre, respuesta múltiple, grabación de voz, transcripción, diálogos, ordenación, asociación, Verdadero/Falso y completar huecos (Gimeno-Sanz, 2016).

La aparición de internet obligó a las universidades de todo el mundo a adaptarse y a reenfocar lecciones, laboratorios, medios de instrucción y de gestión hacia los VLE, siendo *Moodle*, *Teleduc*, *Sakai*, *ProInfo* y *WebCT* los más utilizados (Gomez, 2014). Estos entornos permiten a su vez la distribución de REA o, en inglés, *Open Educational Resources* (OER en adelante), entre los cuales se encuentran los MOOC y los cuales se explicarán con detalle en el siguiente capítulo.

Una de las instituciones pioneras en el uso de la informática fue la Universidad de Phoenix, que ya en 1989 lanzó un campus en línea que permitía la obtención de titulaciones universitarias y másteres (Kaplan y Haenlein, 2016). Actualmente, dos de las universidades a distancia más importantes del mundo son la *Open University* británica y la Universidad Nacional de Educación a Distancia (UNED en adelante) de España. La *Open University* comenzó a ofrecer cursos de idiomas a partir de 1995 (White, 2003) y experimentó un proceso de digitalización a finales del siglo XX. De este modo, para el año 2000 más del 50% de los estudiantes ya accedía a los contenidos que ofrecía la universidad a través de internet (The Open University, s.f.). La UNED, fundada en 1972, comenzó su proceso de virtualización en el curso 2000/2001 y para el año 2005 todos sus cursos eran virtuales. La UNED utiliza como VLE aLF, una plataforma de *e-learning*

de código abierto implementada desde la propia UNED que permite impartir y recibir formación; gestionar y compartir documentos; crear y participar en comunidades temáticas, y realizar proyectos online (García Aretio, 2006). Entre la formación ofertada por la UNED se encuentran los cursos de idiomas ofrecidos por su sección de aprendizaje y enseñanza de lenguas, el Centro Universitario de Idiomas a Distancia, que ofrece distintos niveles de cursos en doce lenguas extranjeras, en las modalidades semipresencial y a distancia.

En las dos últimas décadas se ha generalizado el uso de internet por parte de la población mundial, fenómeno que se ha acentuado con la introducción de las *tablets* y los *smartphones*. Existen innumerables herramientas basadas en las nuevas tecnologías utilizadas tanto en las aulas presenciales como en las virtuales, algunas de ellas específicamente dirigidas a la enseñanza de idiomas. Algunos de los ejemplos más ilustrativos son los siguientes:

- Aplicaciones de videoconferencia generales como *Skype*, *FaceTime* o *Zoom*, o específicas para clases de idiomas como *Lingueo*.
- Plataformas comerciales para el aprendizaje de idiomas, en muchos casos con interfaz web y aplicaciones para móviles, como *Busuu*, *Duolingo* o *Memrise*.
- Comunidades para el aprendizaje de idiomas, siendo *Livemocha* una de las más usadas.
- Plataformas educativas para la comunicación de profesores y alumnos, como *Edmodo* o *Google Classroom*.
- Aplicaciones para crear concursos de preguntas en línea, como *Kahoot!*, *Plickers* o *ClassDojo*.
- Aplicaciones de *flashcards* para aprender vocabulario, como *Quizlet* o *Anki*.
- Herramientas para la creación de avatares, con la opción de grabación de audio, como *Voki*.
- Herramientas de colaboración, entre las que se encuentran blogs y wikis.
- Juegos para ordenador y teléfonos móviles, como *Aworded* (en español apalabrados), *Words with Friends*, *Phrasalstein* o *English Monstruo*.

- Aplicaciones y sitios web para crear cómics y animación de historias, como *Bubblr*, <https://www.makebeliefscomix.com/>, *Storybird* o *Puppet Pals*.
- Aplicaciones para crear pósteres virtuales, como *Glogster* o *Canva*.
- Aplicaciones y webs para ver vídeos o escuchar podcasts y audiolibros, como *Youtube*, *LibriVox* o *iVoox*.
- Recursos en línea con contenido para la programación de las clases. Entre ellos encontramos:
 - Webs colaborativas con bancos de fichas de trabajo con ejercicios, letras de canciones, textos, etc., tales como www.islcollective.com o www.eslprintables.com
 - Webs para aprender idiomas a través de canciones, como www.lyricstraining.com o www.lyricsgaps.com.
 - Webquests.
 - Webs de instituciones con materiales para el aprendizaje de idiomas, como la web de la *BBC Learning English* o la de la *NHK* para el aprendizaje del japonés.

CAPÍTULO 3: LOS MOOC: EVOLUCIÓN DE LOS RECURSOS EDUCATIVOS EN ABIERTO (REA)

3.1. INTRODUCCIÓN

Uno de los mecanismos de generación de contenidos educativos más populares y extendidos en los últimos años, originado a partir de la evolución de la educación a distancia, es el de los OER o REA. Por una parte, la irrupción de los REA se ha visto impulsada por la expansión vertiginosa de las TIC y de internet, lo cual ha posibilitado el intercambio de información a nivel mundial en una escala y velocidad prácticamente inimaginables antes de la década de los 90 (D'Antoni, 2009). Por otro lado, el potencial pedagógico de los REA está profundamente ligado al concepto del aprendizaje basado en recursos, en auge a partir de la segunda mitad del siglo XX, cuando un número creciente de instituciones superiores comenzaron a ofrecer programas educativos presenciales y a distancia. El término «aprendizaje basado en recursos» se basa en el principio de que los educadores deben seleccionar los recursos y métodos más apropiados al contexto en el que imparten la educación. De este modo, se rompen las barreras entre la educación por contacto directo y la educación a distancia, ya que una proporción significativa, pero variable, de la comunicación entre estudiantes y educadores puede llevarse a cabo mediante el uso de diversos medios digitales (Butcher, 2015).

El presente capítulo está dedicado a un tipo específico de REA, los MOOC, y más particularmente los MOOC de lenguas o LMOOC. Por ello, el siguiente apartado se centra en la descripción del término general REA, para posteriormente pasar a la definición de los MOOC y al análisis de algunas de sus características fundamentales, su evolución histórica y su tipología. La parte final del capítulo estará centrada en la definición y descripción de los LMOOC y en las características particulares de su diseño.

3.2. RECURSOS EDUCATIVOS EN ABIERTO (REA)

Las primeras iniciativas para la difusión gratuita de recursos educativos se produjeron en los años 90 del siglo pasado, cuando comenzaron a compartirse

recursos tales como de planes de lecciones, librerías completas o documentación de *software* a través de medios como *Gopher* o páginas web primitivas. Al final de la década, iniciativas surgidas dentro de las universidades como MERLOT, creada por la Universidad de California, o *Connexions*, originada en la Universidad de Rice, comenzaron a ofrecer acceso libre a material curricular universitario. Sin embargo, se considera que el origen de los REA data de 2001, cuando el Instituto Tecnológico de Massachusetts anunció la creación del programa MIT OpenCourseWare, mediante el cual llegaron a publicarse hasta 50 cursos distintos durante el año 2002 (Bliss y Smith, 2017; Stracke et al., 2019). Fue en ese mismo año, 2002, cuando la UNESCO acuñó el término original REA para describir el por entonces nuevo fenómeno global de compartir recursos educativos. En años subsiguientes, se produjeron declaraciones y se confeccionaron guías para fomentar la difusión del movimiento REA. Algunos de ellos fueron la Declaración sobre Educación Abierta de Ciudad del Cabo en 2007, la Declaración sobre REA de Dakar en 2009, la Guía sobre REA en Educación Secundaria, publicada por la Commonwealth de Aprendizaje y la UNESCO en 2011, el Congreso Mundial sobre REA de París organizado en 2012 por la UNESCO, o el II Congreso Mundial sobre REA de Liubliana organizado en 2017 por la UNESCO y que produjo como resultado la redacción de un plan de acción (Bliss y Smith, 2017; Stracke et al., 2019).

La UNESCO describe los REA como «materiales de enseñanza, aprendizaje e investigación en cualquier soporte, digital o de otro tipo, que sean de dominio público o que hayan sido publicados con una licencia abierta que permita el acceso gratuito a esos materiales, así como su uso, adaptación y redistribución por otros sin ninguna restricción o con restricciones limitadas» (s.f.). También se definen como «materiales digitalizados ofrecidos de manera libre y abierta para educadores, estudiantes y autodidactas, para su uso y reutilización en la enseñanza, el aprendizaje y la investigación» (Hylén, 2015, p. 1). Algunos ejemplos de REA serían cursos completos, materiales de cursos, mapas curriculares, aplicaciones multimedia, podcasts, libros de texto, vídeos en *streaming*, pruebas, *software* y cualquier otra herramienta, material o técnica que se use para proporcionar apoyo al conocimiento o que tenga un impacto en el proceso de enseñanza-aprendizaje o en la investigación (Smith y Casserly, 2006; Butcher, 2015).

Hylén (2015) divide los REA en tres tipos de materiales:

- **Contenidos de aprendizaje:** esto incluye cursos completos, *courseware*, módulos de contenido, objetos de aprendizaje, colecciones y revistas científicas.
- **Herramientas:** se refiere al *software* que proporciona apoyo al desarrollo, uso, reutilización e impartición de contenido de aprendizaje, lo cual incluye la búsqueda y organización de contenido, los CMS y LMS, las herramientas de desarrollo de contenidos y las comunidades de aprendizaje en línea.
- **Recursos de implementación:** estos recursos incluyen licencias de propiedad intelectual que promueven la publicación abierta de materiales, además de principios de diseño de buenas prácticas y localización de contenidos.

Johnstone (2005) también divide los REA en tres grupos, de acuerdo con su función en el proceso de aprendizaje:

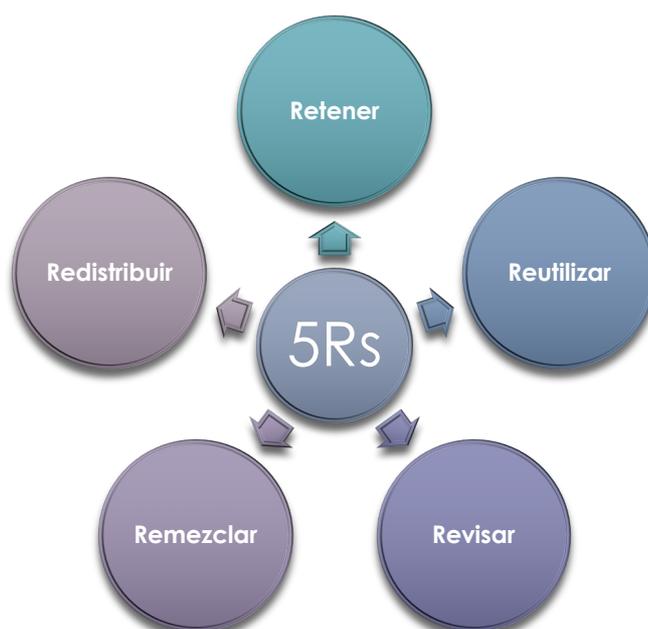
- **Recursos de aprendizaje:** en esta categoría están comprendidos el *courseware*, los módulos de contenido, los objetos de aprendizaje, las herramientas de apoyo al estudiante, las herramientas de autoevaluación y las comunidades de aprendizaje en línea.
- **Recursos para el apoyo de profesores:** son herramientas para profesores y materiales de apoyo que permiten a los profesores crear, adaptar y usar REA, además de materiales de formación para profesores.
- **Recursos que permitan asegurar la calidad** de la educación y las prácticas educativas.

La principal diferencia clave entre los REA y cualquier otro tipo de recurso educativo se encuentra en el hecho de que los REA incorporan una licencia que facilita su reutilización y potencial adaptación, sin tener que solicitar autorización previa al titular de los derechos de autor (Butcher, 2015). Tuomi (2013) define cuatro jerarquías de REA, en función de los derechos contenidos en la licencia. De este modo, REA I garantiza el acceso, REA II incluye derechos de uso, REA III añade derechos de adaptación y REA IV permite derechos de redistribución.

Según Wiley (2014), la apertura de los REA se basa en cinco acciones, que él denomina «las 5R»:

1. **Retener:** es el derecho a hacer, poseer y controlar copias del contenido.
2. **Reutilizar:** es el derecho a usar ampliamente el contenido, como por ejemplo en una clase, un grupo de estudio, una página web o un vídeo.
3. **Revisar:** es el derecho a adaptar, ajustar, modificar o alterar el contenido, lo cual incluye su traducción a otra lengua.
4. **Remezclar:** es el derecho a combinar el contenido original o revisado con otro contenido abierto para crear algo nuevo.
5. **Redistribuir:** es el derecho a compartir con terceros, bien copias del contenido original, bien de la versión revisada o remezclada.

FIGURA 7. LAS 5R DE LOS REA.



Nota. Adaptado de *The Access Compromise and the 5th R*, por D. Wiley, 2014, *Iterating toward openness*. Copyright 2014 by David Wiley.

En el corazón del movimiento de los REA reside la idea de que el mundo del conocimiento es un bien público y de que la tecnología en general y la web en particular ofrecen una oportunidad extraordinaria para que todos puedan compartir, usar y reutilizar el conocimiento (Smith y Casserly, 2006). De este modo, el gran poder de transformación de los REA reside en la facilidad con la

que, una vez digitalizados, pueden compartirse a través de Internet. Según Butcher (2015), la utilización de los REA reporta los siguientes beneficios:

1. Al eliminar restricciones a las copias de recursos, lo cual puede reducir el costo del acceso a materiales educativos, los REA favorecen una mayor disponibilidad de materiales de aprendizaje relevantes y de alta calidad que pueden contribuir a preparar estudiantes y educadores más productivos.
2. El principio de los REA que permite la adaptación y reutilización de los materiales potencia el papel activo del alumno en el proceso educativo y fomenta su creatividad. En consecuencia, ya que los estudiantes aprenden mejor haciendo y creando, los REA contribuyen a la creación de entornos de aprendizaje más eficaces.
3. Los REA aumentan la capacitación, ya que permiten que instituciones y educadores accedan a bajo costo o gratuitamente a los medios de producción de materiales educativos, lo que permite su integración en programas de aprendizaje de alta calidad.
4. Ante las reticencias de educadores y administradores de instituciones con la idea de permitir la distribución libre y gratuita de su propiedad intelectual, debe tenerse en cuenta que cuantas más instituciones utilicen materiales de otras instituciones más se fortalecerán la reputación institucional y la capacidad de atraer a nuevos alumnos. El éxito de las instituciones educativas no reside en los propios contenidos, cuyo valor de venta comercial, que ya de por sí es mínimo, disminuye por el hecho de que una cantidad creciente de material educativo se pone a disposición de forma libre en internet. Al contrario, el prestigio de las instituciones educativas se deriva de la capacidad de guiar a los alumnos a través de recursos educativos mediante un correcto diseño de las vías de aprendizaje y enseñanza, proporcionando a los estudiantes tanto evaluación inteligente y comentarios críticos sobre su desempeño como apoyo efectivo a través de clases prácticas, tutorías, sesiones de orientación individual o por Internet.
5. Compartir contenido públicamente bajo una licencia abierta es una forma segura de proteger los derechos de propiedad intelectual del autor. Esto se debe a que, como el contenido digitalizado puede ser

fácilmente compartido entre alumnos e instituciones, el hacerlo bajo una licencia puede asegurar la atribución del reconocimiento al autor original y facilitar la detección de casos de plagio.

6. Los educadores e investigadores que comparten de forma abierta los contenidos que ellos mismos generan pueden obtener beneficios en sus carreras profesionales. El motivo es que el intercambio y la colaboración con redes de educadores de todo el mundo incrementan las oportunidades de mejora de la práctica docente y del conocimiento sobre áreas específicas, a la vez que producen un aumento de la reputación individual a través de vehículos en línea.

El proyecto OpenEdOz (2016) identifica a su vez seis beneficios clave del empleo de los REA:

1. El desarrollo de economías de escala mediante la coproducción colaborativa de recursos de aprendizaje.
2. El aumento de la calidad del aprendizaje con tiempos y costes menores.
3. La mejor adaptación de los REA a la creciente diversidad de contextos y estilos de aprendizajes actuales.
4. Mayores oportunidades de aprendizaje para comunidades desfavorecidas y para áreas aisladas geográficamente.
5. Mayor colaboración entre proveedores de aprendizaje a través de la revisión por pares y el desarrollo colegial de materiales de aprendizaje.
6. Mayores niveles de transparencia en los procesos educativos.

En el empleo de los REA en educación se encuentra la raíz del concepto de Metodología Educativa Abierta o, en inglés, *Open Educational Practices* (OEP en adelante). La iniciativa europea OPAL, cuyo objetivo es producir un marco de prácticas REA que contribuyan a la mejora de la calidad y la innovación en educación, define OEP como «el uso de REA con el objetivo de mejorar la calidad de los procesos educativos e innovar en los entornos educativos» (Conole y Ehlers, 2010, p. 3) y como «prácticas que apoyan la reutilización y producción de REA a través de políticas institucionales que promueven modelos pedagógicos innovadores, y que respetan y empoderan a los estudiantes como coproductores en su camino de aprendizaje durante

toda la vida» (Andrade et al., 2011, p. 12). Una de las primeras definiciones de OEP fue acuñada por el proyecto OLCOS, perteneciente al programa de *e-learning* de la Comisión Europea, donde se definió OEP como el «conjunto de prácticas que tienen como objetivo involucrar a los estudiantes en un compromiso activo y constructivo con el contenido, las herramientas y los servicios del proceso de aprendizaje, además de promover la autogestión, la creatividad y el trabajo en equipo de los estudiantes» (Geser, 2007, p. 37). De este modo y según argumentan Conole y Ehlers (2010), las OEP implican más que el uso exclusivo de recursos educativos abiertos en un entorno de aprendizaje tradicional cerrado, centrado en la transmisión de conocimiento y orientado hacia un examen final. El empleo de los REA dentro de prácticas educativas auténticamente abiertas debe cumplir los siguientes requisitos que permitan mejorar la experiencia educativa del alumno:

1. Los recursos creados deben diseñarse de tal forma que estén centrados en el alumno.
2. Los alumnos deben estar involucrados en la creación de contenido.
3. Los docentes deben cambiar el enfoque de la enseñanza, desde uno basado en el contenido a otro orientado a los estudiantes.
4. El proceso de aprendizaje debe verse como una parte importante y productiva de la experiencia educativa, de tal modo que el foco no se centre únicamente en los resultados o productos de aprendizaje.
5. Los resultados del aprendizaje deberían ser valorados en su condición de artefactos que compartir, debatir, mejorar y reutilizar. De este modo, las prácticas educativas abiertas se convierten en escenarios educativos en los cuales el aprendizaje se produce mediante interacciones reflexivas dentro de la práctica social.

Según Beetham et al. (2012) hay seis tipos de prácticas que pueden englobarse dentro de los OEP:

1. Producción, gestión, uso y reutilización de REA.
2. Pedagogías abiertas y públicas.
3. Aprendizaje abierto, incluido el aprendizaje entre pares y la acreditación abierta.

4. Práctica académica abierta, la cual incluye investigación abierta, datos abiertos y publicaciones de acceso abierto.
5. Intercambio abierto de ideas de enseñanza.
6. Uso de tecnologías abiertas, incluidas las redes sociales y las herramientas digitales abiertas.

Este conjunto de prácticas se produce a través de tres dimensiones fundamentales dentro de las OEP que engloban todo el proceso de creación, empleo, reformulación y apoyo de los REA (Conole y Ehlers, 2010):

- Los **agentes**, quienes se dedican a crear, usar o apoyar el uso de los REA. Existen dos tipos de agentes: los involucrados en la creación y uso de REA y los involucrados en los aspectos de política y gestión. Los principales agentes son:
 - **Creadores** de los REA.
 - **Usuarios** de los REA.
 - **Administradores**, quienes proporcionan tanto los medios técnicos y organizativos para apoyar los REA como las herramientas y el soporte para la creación y el uso de los REA.
 - **Responsables de políticas**, quienes introducen el fomento de los REA en las agendas políticas.
- La amplia gama de **artefactos mediadores** que permiten la creación y fomento del uso de los REA. Estos incluyen herramientas y recursos para orientar el proceso de creación y el uso de REA, así como las tecnologías que apoyan su alojamiento y gestión.
- Los **factores contextuales** involucrados en la creación, uso o soporte de los REA.

De todo lo anterior se desprende que la clave dentro de las prácticas englobadas por las OEP es el concepto de apertura. Hodgkinson-Williams y Gray (2009) formularon un marco que permite analizar la apertura de las OEP mediante cinco parámetros que se producen a modo de continuo:

- **Técnico**: incluye la interoperabilidad y formatos abiertos; la conectividad; las habilidades y equipos técnicos; la disponibilidad, y la capacidad de descubrimiento de recursos.

- **Legal:** hace referencia tanto a los parámetros de las licencias abiertas como al conocimiento y asesoramiento sobre las licencias abiertas.
- **Cultural:** comprende las concepciones del conocimiento dado o construido y los planes de estudio.
- **Pedagógico:** se refiere a los parámetros demográficos de los estudiantes y sus distintos tipos de compromiso, además de a las estrategias pedagógicas; de aprendizaje y de evaluación, y a los mecanismos de acreditación y certificación.
- **Financiero:** hace referencia a los costes, que varían desde los gratuitos hasta los que cobran una tarifa, e incluye la concepción de modelos comerciales sostenibles.

Son diversos los beneficios de cambiar el foco de atención desde los REA hacia las OEP. En primer lugar, mientras que tradicionalmente el trabajo en el campo de los REA se ha centrado en la disponibilidad y accesibilidad de los contenidos y recursos, el objetivo de las OEP es la creación del ambiente educativo en el que se crean o usan los REA, lo cual redundaría en beneficios en el proceso de enseñanza-aprendizaje. En segundo lugar, las OEP están orientadas a la utilización de los REA en el contexto educativo, creando actividades; comentarios, e interacción en torno a un material de aprendizaje abierto, de tal modo que se permita mejorar la calidad de las experiencias de aprendizaje. En tercer lugar, las prácticas educativas abiertas son prácticas cuyo objetivo es ampliar el paradigma de los recursos y la educación basada en el contenido. Su misión es enfocar la educación hacia las prácticas dentro de un dominio y no solo hacia los recursos, de tal modo que no solo el conocimiento, sino la responsabilidad, sean el objetivo del proceso educativo. En cuarto lugar, las OEP pueden servir de puente entre las experiencias de aprendizaje formales e informales, ya que es posible transferir las OEP desarrolladas a partir de REA en un contexto formal a uno informal, generando así un nuevo conjunto de prácticas asociadas en torno a los REA en este nuevo contexto. Por último, un enfoque hacia las OEP podría actuar como catalizador en la adopción de prácticas educativas más abiertas, al permitir la reconsideración de los roles y límites tradicionales dentro de un enfoque más reflexivo en el que se posibilita la exploración de los mecanismos mediante los cuales tanto los alumnos como los docentes participan activamente en todo el

ciclo REA. Dentro de este enfoque, los alumnos y los profesores se convierten en pares en la validación de los procesos de aprendizaje a través del diálogo crítico, alejándose así de las pruebas orales o escritas que requieren la reproducción de un conjunto predefinido de activos de conocimiento. Esto implica un cambio en el papel de los profesores, quienes pasan de ser transmisores de conocimiento a facilitadores, ayudando así a los estudiantes a validar sus propias experiencias de aprendizaje (Conole y Ehlers, 2010).

Los REA y las OEP se enmarcan dentro del movimiento de educación abierta, término que describe una concepción de la educación que tiene como objetivo eliminar todas las barreras superfluas al aprendizaje y proporcionar a los estudiantes una posibilidad razonable de éxito dentro de un sistema de educación centrado en sus necesidades específicas y localizado en múltiples áreas de conocimiento. El aprendizaje abierto incorpora principios clave tales como el aprendizaje a lo largo de la vida, el proceso de aprendizaje basado en el estudiante, la enseñanza flexible o el reconocimiento de los conocimientos y la experiencia previos. El aprendizaje abierto es un concepto anterior al auge de los REA y supone implicaciones significativamente mayores que el uso de REA en programas educativos, ya que requiere de una serie de elementos que no están relacionados directamente con los REA, como son el análisis sistemático de los sistemas de evaluación y acreditación; el apoyo al estudiante; los marcos de referencia para planes de estudios, o los mecanismos para reconocer conocimientos previos. Los REA, por lo tanto, pueden considerarse como un aspecto de la educación abierta (Salinas y Sureda, 1992; Area y Adell, 2009; Butcher, 2015; Stracke et al., 2019).

Existen una serie de conceptos, como el de *OpenCourseWare* (OCW en adelante) o el *e-learning*, que suelen confundirse habitualmente con los REA. Suele considerarse REA como sinónimo de materiales de cursos abiertos o OCW, los cuales son materiales de nivel universitario de alta calidad publicados de forma libre y gratuita, y organizados como cursos digitales. Los OCW incluyen frecuentemente materiales para la planificación de cursos, herramientas de evaluación y contenido temático. Sin embargo, aunque REA y los materiales OCW puedan tener un significado próximo, la definición de OCW se

corresponde de forma más apropiada con un subconjunto específico y más estructurado de los REA (Guzmán Parra y Vila Oblitas, 2011; Butcher 2015).

El término REA también se confunde habitualmente con *e-learning*. Sin embargo, al contrario que con los OCW, en ningún caso aprendizaje abierto y *e-learning* pueden considerarse sinónimos de los REA. Respecto al aprendizaje electrónico, en línea o *e-learning*, si bien es cierto que muchos cursos de aprendizaje electrónico se valen de los REA, los contenidos REA pueden producirse en medios distintos al electrónico, tales como textos en papel impreso. Incluso cuando pueden ser compartidos en formato digital, es frecuente el caso de países en vías de desarrollo con deficiencia de medios en los que los REA se imprimen y distribuyen como recursos impresos. Esto imposibilita que REA y *e-learning* puedan ser considerados sinónimos (Butcher, 2015).

En los últimos años, dentro del movimiento de la educación abierta, los MOOC han surgido como una de las innovaciones con mayor crecimiento mundial. No obstante, si se analizan los MOOC como recursos basados en contenido a través del parámetro de apertura, la mayoría de los MOOC no pueden ser considerados como REA debido, en primer lugar, a que su tamaño y complejidad dificultan su reutilización y redistribución, y, en segundo lugar, a sus restricciones de apertura, tanto en lo que respecta a la ausencia de costes (muchos MOOC requieren una tasa de expedición de certificados) como a las licencias para su modificación y redistribución (Ozturk, 2015; Wiley et al., 2015; Stracke et al., 2019). Los casos excepcionales de MOOC con licencia de reutilización y adaptación podrían clasificarse bien como un REA único, lo cual, debido a su gran tamaño, crearía un subtipo específico de REA, o bien como colecciones de múltiples REA. Sin embargo, si se analizan desde el punto de vista de la innovación en el aprendizaje, los MOOC son potencialmente mucho más que simples REA, ya que, además de recursos, incluyen métodos y rutas pedagógicas. Como se analizará posteriormente, los MOOC ofrecen una experiencia de aprendizaje completa basada en un diseño pedagógico determinado por los objetivos del curso y hacen libres, no solo los materiales, sino también los procesos de interacción, los cuales se convierten en el centro del aprendizaje. De este modo, los MOOC irían más allá de los REA, ya que se

postularían como un instrumento potente para la transformación y mejora de la calidad educativa a través de un enfoque articulado en torno al aprendizaje entre pares y las comunidades en línea (Durall et al., 2012; Patru y Balaji, 2016; Bliss y Smith, 2017; Stracke et al., 2019).

3.3. MASSIVE OPEN ONLINE COURSES (MOOC)

3.3.1. DEFINICIÓN

Como se ha especificado anteriormente, MOOC es el acrónimo en inglés de *Massive Open Online Course* o, en español, Curso en Línea Masivo y Abierto. Consecuentemente, se suele definir el término MOOC a través de los cuatro conceptos que forman su acrónimo: son cursos, porque plantean una estructura enfocada a la enseñanza y a la superación de pruebas; son masivos, porque están enfocados a una demanda de millones de personas en todo el mundo; son abiertos, porque sus contenidos están generalmente a libre disposición del estudiante, que puede compartirlos e incluso modificarlos, y son en línea, porque se accede a ellos a través de Internet (Pedreño, et al., 2013). Los proyectos HOME y ECO junto con OpenUpEd, uno de los mayores proveedores de MOOC, los definen como «cursos en línea diseñados para un gran número de participantes, a los que cualquier persona puede acceder desde cualquier sitio mediante una conexión a internet, que están abiertos a todo el mundo sin requisitos de acceso y que ofrecen la experiencia completa de un curso en línea de manera gratuita» (Jansen y Schuwer, 2015, p. 11).

Según Haggard (2013), un MOOC típico tiene una duración de cuatro a diez semanas que se dedican en su mayor parte al aprendizaje, las últimas de las cuales suelen dedicarse a la producción de un trabajo. Los participantes dedican una media de dos a seis horas semanales al curso, aunque existe un porcentaje pequeño de alumnos más comprometidos que pueden dedicar más tiempo. El número de inscritos en un MOOC puede alcanzar las decenas de miles, mientras que el número de participantes que completan el curso y obtienen un certificado como mucho alcanza cifras de cuatro dígitos. En consecuencia, normalmente el volumen de materiales consumidos durante el curso va decreciendo conforme lo hace el compromiso de los estudiantes,

aunque los recursos publicados en el MOOC permanecen accesibles después de su finalización.

Las cuatro dimensiones que conforman el concepto de los MOOC son, por consiguiente, las que conforman su acrónimo, las cuales se pasan a describir en mayor detalle a continuación:

- **Masivo:** según Castrillo (2014), es el gran número de participantes, potencialmente ilimitado, lo que diferencia los MOOC de otros medios de enseñanza en línea. Esta dimensión masiva de los MOOC define el diseño del curso, que es específico para poder dar cabida a un gran número de participantes, de tal forma que el esfuerzo que se requiere para proporcionar todos los servicios asociados al curso no se incrementa de forma significativa al aumentar el número de participantes (Jansen y Schuwer, 2015; Patru y Balaji, 2016). Para Downes (2013), un elemento fundamental de esta masividad es que la tecnología con la que se implementa el curso tiene la capacidad de posibilitar conversaciones y actividades a través de múltiples plataformas e involucrar a los participantes en ellas. Downes propone el número de Dunbar, 150, como el número de participantes activos (que no debe confundirse con el número de personas inscritas al MOOC) a partir del cual un curso puede considerarse verdaderamente masivo, ya que este número representa el número teórico máximo de personas con las que alguien puede interactuar. Esta limitación obliga a los participantes a seleccionar las relaciones y los recursos con los que se interacciona dentro del MOOC.
- **Abierto:** como indican Jansen y Schuwer (2015), la apertura de un MOOC abarca varios aspectos que conforman algunos de los rasgos característicos de este tipo de cursos:
 - Los cursos están disponibles para casi todo el mundo, sin limitaciones. El único requisito, al ser cursos en línea, es disponer de una conexión a internet.
 - El concepto de apertura también implica la ausencia de requisitos de entrada, por lo que no es necesario acreditar ningún tipo de titulación para acceder a los MOOC. Esto no implica que el nivel de conocimientos del MOOC sea un nivel básico o inicial. De

hecho, como indican Patru y Balaji (2016), los participantes con experiencia académica limitada suelen experimentar dificultades para completar los MOOC.

- Los MOOC pueden completarse de forma gratuita, ya que los participantes tienen acceso a la experiencia completa del curso sin ningún coste. Sin embargo, la transcripción del trabajo realizado a créditos universitarios o la emisión de certificados sí conllevan normalmente el pago de una tarifa (Durall et al., 2012; Patru y Balaji, 2016).
- **En línea:** el curso completo está disponible en internet y se puede acceder a él utilizando un ordenador portátil, de sobremesa, una *tablet* o un *smartphone*. Esto implica que el 100% del MOOC debe ofrecerse en línea, de lo contrario se trataría de un curso híbrido (Patru y Balaji, 2016).
- **Curso:** los MOOC son cursos que ofrecen una experiencia de aprendizaje completa. Esto significa que están estructurados en torno a un conjunto de objetivos de aprendizaje, dentro de un área de estudio definida, y que incluyen los materiales del curso y herramientas de evaluación tales como pruebas, valoraciones, un examen final y un certificado de compleción (Patru y Balaji, 2016). Según Jansen y Schuwer (2015), un MOOC debe ofrecer una serie de elementos tales como contenido educativo, que puede ser en forma de vídeo, audio, texto, animaciones, redes sociales, juegos o simulaciones; la posibilidad de interacción con el equipo docente y entre participantes mediante redes sociales, foros, blogs o agregadores de RSS, de tal forma que sea posible construir una comunidad de aprendizaje; actividades, tareas, pruebas y algún tipo de valoración, que puede ser automática, a través de pares o por parte del equipo docente; algún tipo de reconocimiento, como insignias o un certificado de finalización, y un plan de estudios con instrucciones sobre cómo organizar el aprendizaje a partir del material ofrecido y las interacciones diseñadas. La unidad de estudio mínima que debe abarcar un MOOC es un crédito *European Credit Transfer and Accumulation System* (ECTS en adelante), lo cual equivale a entre 25 y 30 horas de estudio, aunque lo normal es que la duración de los MOOC oscile entre uno y cuatro créditos ECTS.

FIGURA 8. LAS CUATRO DIMENSIONES DE LOS MOOC. FUENTE: ELABORACIÓN PROPIA.



Al suponer un cambio en los esquemas instructivos, los MOOC requieren un cambio en el diseño de la formación, ya que la base de cada MOOC es la agregación y redistribución de un conjunto amplio y diverso de contenidos, aportados por una variedad de expertos, educadores e instructores de un campo específico. Al adoptar una concepción abierta del aprendizaje y ofrecer oportunidades de formación, con independencia de la afiliación a una institución en particular, los MOOC también requieren un cambio respecto a las

expectativas de aprendizaje de los alumnos, ya que son los participantes los que controlan cómo, dónde y cuándo aprenden (Durall et al., 2012).

3.3.2.HISTORIA DE LOS MOOC

Algunos autores consideran que el primer curso que reúne las características atribuidas a los MOOC fue un curso sobre educación abierta organizado en 2007 por David Wiley, profesor de la Universidad Estatal de Utah (USA) (López Meneses et al., 2015). Sin embargo, el primer curso en utilizar un *software* que permitiera su distribución fue *Connectivism and Connective Knowledge (CCK08)*, diseñado en 2008 por George Siemens y Stephen Downes para la Universidad de Manitoba. 25 alumnos pagaron la matrícula de este curso y obtuvieron el correspondiente título, pero, de manera adicional, se ofreció al público general la posibilidad de seguir el curso a través de internet de forma gratuita y sin acreditación, oportunidad que fue aprovechada por 2200 estudiantes. Ante el gran número de participantes inscritos, Dave Cormier y Bryan Alexander acuñaron el término *Massive Open Online Course* o MOOC. Debido al enorme éxito del curso, Siemens y Downes continuaron ofreciendo nuevas versiones de este, además de organizar cursos distintos siguiendo el mismo modelo (Downes, 2012; Castrillo et al., 2018).

A pesar del éxito de los cursos, el fenómeno de los MOOC tuvo que esperar hasta el año 2012 para popularizarse. Uno de los motivos fue el curso en línea *Inteligencia artificial* dirigido por Sebastian Thrun y Peter Norvig, que la Universidad de Stanford ofreció en abierto en otoño de 2011 a través de una compañía *startup* llamada Know Labs. En este curso, en el que inicialmente se matricularon 58.000 personas, se llegó a alcanzar la cifra de 160.000 estudiantes de más de 190 países. El MOOC estaba basado en una serie de cursos tradicionales que se habían impartido con anterioridad en la Universidad de Stanford, los cuales se adaptaron recurriendo a vídeos en *streaming* y a tecnología interactiva para las pruebas que servían de medio de evaluación del curso. A raíz del éxito del curso, en 2012 Sebastian Thrun fundó junto a Michael Sokolsky la plataforma MOOC Udacity (Markoff, 2011; Pappano, 2012; López Meneses et al., 2015; Castrillo et al., 2018; Udacity, s.f.). De manera similar, en mayo de 2012, el Instituto Tecnológico de Massachussets y la Universidad de

Harvard lanzaron edX³, una plataforma de MOOC sin ánimo de lucro que inauguró sus cursos en otoño de 2012 con 370.000 alumnos inscritos (Pappano, 2012). La creación de Coursera data de abril de 2012, cuando los profesores de informática de la Universidad de Stanford Daphne Koller y Andrew Ng ofertaron sus cursos en forma de MOOC. A estos cursos se fueron uniendo otros de las universidades de Princeton y la Universidad de Pennsylvania, hasta llegar a las más de 200 instituciones de todo el mundo que actualmente ofrecen sus cursos en la plataforma. El número de estudiantes que ha seguido alguno de los MOOC de Coursera es de 97 millones, lo que convierte a Coursera en el mayor proveedor de MOOC del mundo (Lewin, 2012; Castrillo et al., 2018; Shah, 2021; Coursera, s.f.).

Desde finales del año 2012 y durante 2013, se empezaron a crear plataformas MOOC fuera de los Estados Unidos. A finales de 2012 la *Open University* creó *FutureLearn*, un proveedor de MOOC de Reino Unido que comenzó a ofertar cursos en septiembre de 2013 y que actualmente se ha convertido en la mayor plataforma de MOOC de Reino Unido y la cuarta del mundo, con cerca de 17 millones de alumnos. En colaboración con instituciones universitarias de todo el mundo, la mayoría de ellas europeas, *FutureLearn* cuenta con su propio programa de credenciales y ofrece 24 programas distintos de grado (Shah et al., 2022; Shah, 2021; FutureLearn, s.f.). La plataforma de MOOC OpenHPI, creada por el Hasso Plattner Institute en Potsdam, Alemania, inauguró su primer MOOC en septiembre de 2012 y está especializada en el campo de las TIC (Meinel y Willems, 2013; Shah et al., 2022). La plataforma francesa FUN fue creada en octubre de 2013 por el ministerio de educación francés y desde entonces 2 millones de estudiantes han participado en algunos de sus 800 MOOC (Shah et al., 2022; FunMOOC, s.f.). También en 2013 se crearon las plataformas de MOOC asiáticas eWant, fundada por la Universidad Nacional Chiao Tung en Taiwan y que actualmente cuenta con 1.200 MOOC de 91 universidades distintas, y XuetangX, fundada por la universidad Tsinghua y supervisada por el ministerio de educación chino, la cual se ha convertido en la principal plataforma de MOOC de China y una de las de mayor crecimiento en todo el mundo, con más 14 millones de usuarios. Aunque creada en 2017,

³ <https://www.edx.org>

merece la pena mencionar la plataforma nacional de MOOC india SWAYAM, que en tan solo cinco años ha logrado atraer a más de 21 millones estudiantes a alguno de sus cursos y la cual, si sus tasas de crecimiento permanecen constantes, se podría convertir en la mayor plataforma de MOOC en el mundo. SWAYAM ofrece 2.253 MOOC en colaboración con 203 universidades indias (Shah et al., 2022).

El primer MOOC en España fue un curso de criptografía llamado *El algoritmo RSA*, lanzado el 15 de marzo de 2012 como parte del proyecto *Crypt4you*, de la Universidad Politécnica de Madrid (Álvarez, 2012). Sin embargo, la primera institución de educación superior que ofreció MOOC en una plataforma organizada fue la UNED, en octubre de 2012, con UNED COMA (Polo, 2013). La oferta de MOOC de la UNED se proporciona actualmente a través de la plataforma UNED Abierta⁴, iniciada en octubre de 2012 para reunir diferentes iniciativas formativas dentro de la universidad. En enero de 2016 se creó el portal IEDRA⁵, que integra todos los recursos en abierto de la UNED (Castrillo y Mañana, 2017). En enero de 2013 se creó Miríadax, participada originalmente en un 50% por Telefónica y la Fundación Universia del Banco Santander, y UNED Abierta. Desde entonces, Miríadax se ha convertido en la primera plataforma de MOOC iberoamericana, con más de 100 instituciones colaboradoras y más de 6 millones de alumnos matriculados en sus cerca de 600 MOOC en español y portugués (Castrillo et al., 2018; Shah et al., 2022).

Durante la pandemia causada por el coronavirus en el año 2020, el fenómeno de los MOOC experimentó un agudo crecimiento, ya que los confinamientos impuestos por todo el mundo propiciaron el que muchas personas se sintieran atraídas a la educación en línea. De hecho, una tercera parte de todos los estudiantes que alguna vez se han registrado en una plataforma MOOC lo hicieron durante el año 2020. Se pasó de la cifra de 110 millones de participantes en alguno de los 13.500 MOOC ofrecidos por más de 900 universidades de todo el mundo⁶ en 2019, lo cual ya representaba un

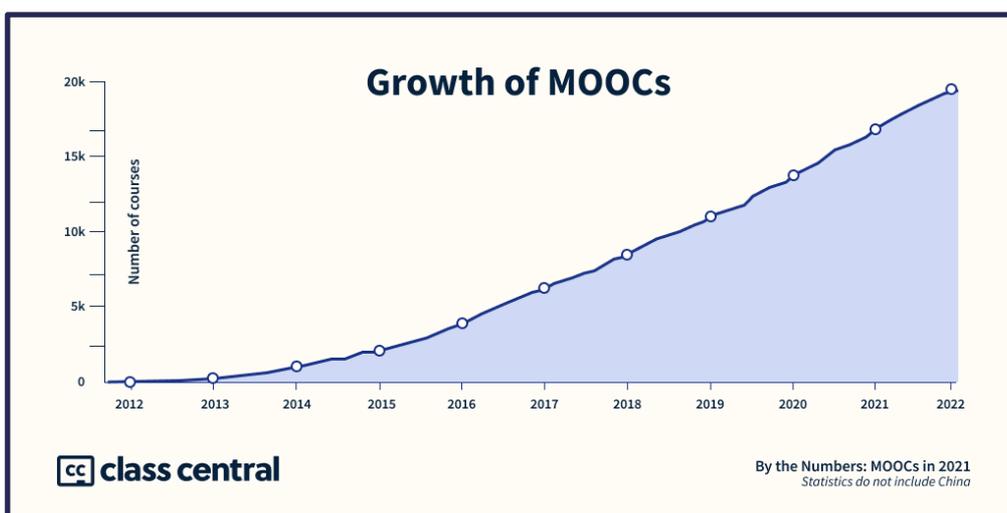
⁴ <https://unedabierta.uned.es/>

⁵ <http://iedra.uned.es/>

⁶ Las estadísticas referentes a los MOOC ofrecidas en esta sección no incluyen datos relativos a China, ya que las fuentes citadas consideran que tanto la inexistencia de ciertos datos, como la

aumento del 10,9% respecto a 2018, a 180 millones de participantes en 16.300 MOOC ofrecidos por 950 universidades en 2020, lo cual supuso un incremento del 16,3%. Aunque a un menor ritmo, durante el año 2021 el interés en los MOOC siguió experimentando un auge abrumador, al lanzarse alrededor de 3.100 nuevos cursos y añadirse 40 millones de nuevos usuarios, un incremento del 12,2% respecto al año anterior. De este modo, en tan solo una década se ha pasado de los 160.000 participantes en el MOOC *Inteligencia artificial* de 2011 a 220 millones de participantes en alguno de los MOOC ofrecidos por más de 950 universidades de todo el mundo, lo que significa que el 2,8% de la población mundial se ha registrado alguna vez en un MOOC (Shah, 2018b; Shah, 2019; Shah, 2020; Shah, 2021).

FIGURA 9. CRECIMIENTO DEL NÚMERO DE MOOC EN EL PERIODO 2012-2022.



Nota. Tomado de *By the Numbers: MOOCs in 2021*, por D. Shah, 2021. Copyright 2011-2022 by Class Central.

En los últimos años, el número de grados en línea y de micro credenciales ha experimentado un gran crecimiento. Como puede observarse en la tabla 2, la cantidad de grados se ha septuplicado desde 2017, pasándose en tan solo cuatro años de 10 grados a 70, los cuales son ofrecidos en su práctica totalidad por las plataformas Coursera, edX y FutureLearn (Shah, 2019; Shah, 2021).

imposibilidad de validación de los datos existentes impiden reflejar de forma fehaciente la situación real de los MOOC en China.

TABLA 2. CRECIMIENTO DEL NÚMERO DE GRADOS EN LÍNEA EN EL PERIODO 2017-2021.

	2017	2018	2019	2020	2021
Coursera	4	11	16	25	34
edX	1	9	10	13	13
FutureLearn	4	18	23	28	22
Udacity	1	1	1	1	1

Nota. Adaptado de *By the Numbers: MOOCs in 2021*, por D. Shah, 2021. Copyright 2011-2022 by Class Central.

Otro campo reciente de crecimiento son las micro credenciales. Una micro credencial es una credencial que reconoce una habilidad o logro particular. Las micro credenciales pueden presentarse en variedad de formatos, entre ellos los certificados, los nanogradados, las insignias digitales o las insignias abiertas (Clements, West y Hunsaker, 2020). En 2021, se crearon 500 nuevas micro credenciales, alcanzándose el total de 1670, lo cual supone un aumento del 334% en tan solo cuatro años, como puede observarse en la tabla 3 (Shah, 2018a; Shah, 2021).

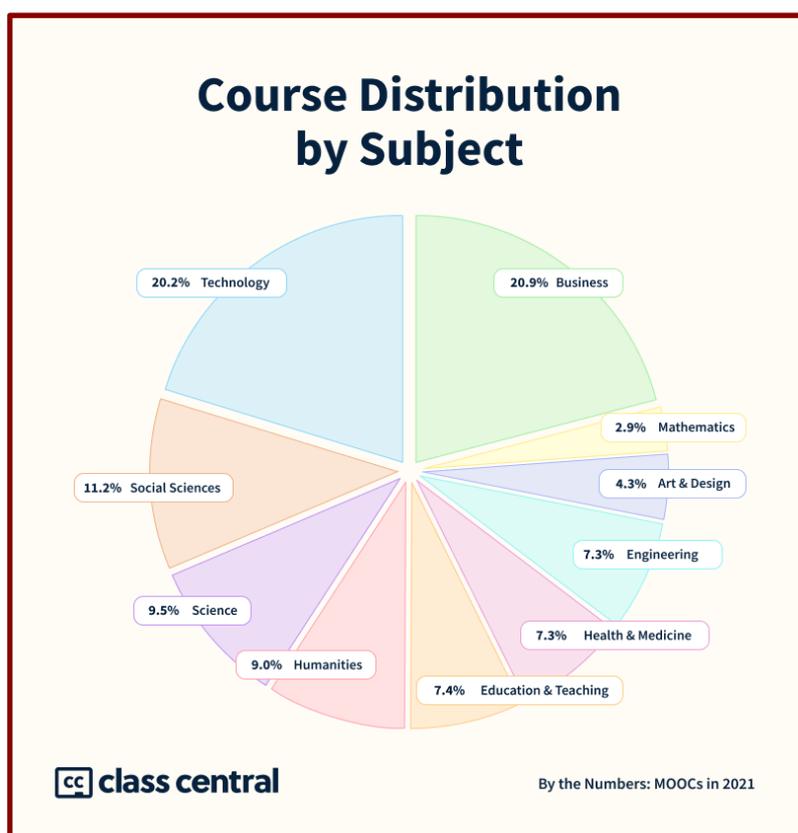
TABLA 3. CRECIMIENTO DEL NÚMERO DE MICRO CREDENCIALES EN EL PERIODO 2017-2021.

2017	2018	2019	2020	2021
500	630	820	1180	1670

Nota. Adaptado de *By the Numbers: MOOCs in 2021*, por D. Shah, 2021. Copyright 2011-2022 by Class Central.

En cuanto al campo de especialización de los MOOC, como puede observarse en la figura 10, tanto los MOOC pertenecientes al área del comercio como los dedicados a la tecnología representan 2 de cada 10 MOOC existentes, cifras extrapolables a las de años previos. El total de MOOC de humanidades, entre los que se categorizan los LMOOC, representa un 9% de los MOOC (Shah, 2021).

FIGURA 10. DISTRIBUCIÓN DE LOS MOOC EN 2021 POR TEMÁTICA



Nota. Tomado de *By the Numbers: MOOCs in 2021*, por D. Shah, 2021. Copyright 2011-2022 by Class Central.

3.3.3. TAXONOMÍAS DE MOOC

De las numerosas taxonomías utilizadas para categorizar los MOOC, existe una clasificación fundamental basada en la teoría de aprendizaje subyacente a su diseño que los divide en dos grupos: cMOOC y xMOOC. Los cMOOC, o MOOC conectivistas, son el grupo al que pertenece el primer MOOC diseñado en 2008 por Siemens y Downes, mientras que los xMOOC son el grupo al que pertenece el MOOC *Inteligencia artificial* de 2012, creado por Thrun y Norvig y que sirvió como modelo para un gran número de los MOOC de plataformas como Udacity, Coursera o EdX (Jacoby, 2014)

Los cMOOC están basados en el conectivismo, un enfoque pedagógico de aprendizaje basado en el constructivismo social cuyo objetivo es lograr la apertura y la enseñanza participativa (Jacoby, 2014). El conectivismo se plantea como un avance del conductismo, el cual postula que el aprendizaje es

consecuencia de la repetición de respuestas a estímulos; del cognitivismo, cuyo objeto de estudio es la forma en la que la mente interpreta, procesa y almacena información en la memoria, y del constructivismo, para el que el aprendizaje implica una constante construcción de nuevos conocimientos y reestructuración de los previos (Martín Peris, 2008). El objetivo del conectivismo es mejorar la calidad de la educación a través del fortalecimiento de los vínculos horizontales y la estimulación de la cooperación conjunta en grupos de estudiantes, a través de la recopilación y difusión del conocimiento contribuido por los participantes. El enfoque pedagógico que se usa en los cMOOC es flexible y se adapta a las necesidades específicas de los participantes, ya que ayuda a encontrar personas con los mismos patrones de pensamiento y ofrece la oportunidad de ampliar la red de contactos. De hecho, la interacción principal en los cMOOC se produce entre los propios participantes, de tal modo que, en lugar de seguir una ruta lineal, cada participante es el encargado de crear su propia trayectoria. De esta forma, el desarrollo del conocimiento, la creatividad y la comunicación de los participantes se efectúa a través de las herramientas de la Web 2.0, lo cual implica que la mayoría de la actividad de los estudiantes sucede fuera de la plataforma que aloja el MOOC, en elementos de internet tales como blogs y portafolios personales, redes sociales, wikis, mundos virtuales o webs de alojamiento de vídeos como Youtube. Los participantes del MOOC transmiten su información usando, en muchos de los casos, un agregador de contenido, el cual lo añade a otras fuentes de información como los foros del MOOC. Toda esta información se combina en una serie de enlaces y descripciones que se envía de manera automática a todos los participantes del curso, normalmente mediante correo electrónico, en forma de boletín de noticias o *newsletter*. Los estudiantes acceden a la información, reflexionan sobre ella, entregan sus actividades y con ello se retroalimenta el proceso (Masters, 2011; Clark, 2013; Seidametova, 2018).

El modelo de los xMOOC, por el contrario, es de enfoque conductista y cognoscitivista, con un diseño de curso individualista (Conole, 2016; Kop, 2011). Los xMOOC tienden a ser versiones adaptadas a las características de las plataformas de los MOOC de los formatos de aprendizaje desarrollados tradicionalmente por las universidades en el marco de sus acciones de *e-learning*. De este modo, un profesor o grupo de profesores, en su papel de

expertos, se encargan de seleccionar los contenidos específicos del xMOOC, los cuales son posteriormente transmitidos a los estudiantes para su absorción a través de la plataforma MOOC (Vázquez et al., 2013; Cabero et al., 2014). En consecuencia, los xMOOC son de naturaleza más tecno céntrica, tanto en su diseño y estructura como en su filosofía (Jacoby, 2014). En ellos, las vídeo clases, y más particularmente las exposiciones apoyadas en presentaciones, desempeñan un papel muy significativo como elemento de presentación de los contenidos. El modelo de evaluación que se utiliza en los xMOOC es muy parecido al seguido en las clases tradicionales de formación virtual, con herramientas de evaluación estandarizadas y automatizadas. En numerosas ocasiones, los exámenes que el estudiante debe superar para adquirir la certificación del curso se construyen de forma aleatoria a partir de una base de datos de preguntas (Cabero et al., 2014).

Lane (2012) amplía estas dos categorías básicas de MOOC al añadir el MOOC basado en tareas o tMOOC. Este tipo de MOOC, cuyo modelo es el curso ds106, un curso certificable sobre narrativa digital iniciado por Jim Groom en la primavera de 2010 (ds106, s.f.), presenta un enfoque mixto entre el instructivismo y el constructivismo. En los tMOOC, el aprendizaje se halla distribuido y en diferentes formatos, y se adquiere mediante una serie de trabajos, tareas o proyectos que presentan múltiples formas de resolución, pero cuyo carácter obligatorio hace que sea imposible pasar a nuevos aprendizajes hasta haber adquirido las habilidades previas. Por consiguiente, los tMOOC hacen hincapié en el desarrollo por parte de los alumnos de las habilidades necesarias para la resolución de determinados tipos de tareas. Algunos ejemplos de las tareas que los participantes deben realizar en los tMOOC son las siguientes: resolución de casos; lectura y análisis de documentos; construcción de recursos en diferentes formatos; análisis de sitios web; elaboración de blog y wikis; elaboración de mapas conceptuales; resolución de problemas, y configuración de crucigramas sobre términos científicos. Al contrario que en los cMOOC, la comunidad en los tMOOC representa un objetivo secundario, aunque su papel sigue siendo crucial a la hora de comentar y resolver dudas puntuales. Los métodos tradicionales de evaluación que implementan los xMOOC presentan problemas de compatibilidad con los tMOOC, por lo que el análisis de la calidad de las tareas elaboradas por los participantes es

efectuado bien por un ayudante en el programa, bien mediante evaluación por pares (Lane, 2012; Martí, 2012; Cabero et al., 2014).

Seidametova (2018) añade a los xMOOC y los cMOOC otros dos tipos: los quasi MOOC y los MOOC híbridos. Los quasi MOOC son cursos que pueden no estar desarrollados por un profesor certificado, cuyo objetivo es proporcionar a estudiantes de distintos grupos de edad acceso a colecciones de recursos de aprendizaje gratuitos de disciplinas variadas en modo de mini lecciones. Los MOOC híbridos, hMOOC o MOOC 3.0 integran y combinan clases en línea con el aprendizaje presencial y pueden incluir conceptos como las *flipped-classrooms*.

Una de las primeras y más amplias clasificaciones de MOOC la estableció Clark (2013) al dividir funcionalmente los MOOC en ocho categorías consideradas mutuamente no excluyentes. Además de los cMOOC, la clasificación de Clark incluye los siguientes tipos de MOOC:

1. Los **transferMOOC** son adaptaciones literales de cursos ya existentes a plataformas de MOOC, con el objetivo de transferir contenido al alumno. El contenido se transmite mediante clases magistrales en vídeo y la lectura de los materiales del curso, y la evaluación se realiza mediante exámenes en línea. Los transferMOOC se hallarían por tanto en la parte más tradicional del espectro de los MOOC (Clark, 2013; Pilli y Admiraal, 2016).
2. Los **madeMOOC** tienden a ser más innovadores en el uso de los vídeos, ya que evitan las clases magistrales y favorecen las explicaciones de pizarra, las tareas prácticas y los procedimientos tomados del mundo real. De manera adicional, los madeMOOC ofrecen materiales de mayor calidad formal y tareas más exigentes y elaboradas. El enfoque de este tipo de MOOC es interactivo y en estos cursos son frecuentes el trabajo colaborativo, la evaluación por pares y la interacción con *software* para resolver problemas, lo cual permite recibir valoraciones instantáneas (Clark, 2013; Pilli y Admiraal, 2016).
3. Los **Synchronous Massive Open Online Courses** (SMOOC en adelante) o **synchMOOC**, con el objetivo de incrementar la motivación de los

participantes y alinear la disponibilidad de los docentes con el trabajo de los alumnos, establecen fechas fijas de comienzo, finalización y entrega de tareas, las cuales suelen coincidir con las fechas habituales de un curso académico. En los SMOOC, los profesores retransmiten las clases en directo para que los alumnos las visualicen durante las horas de clase designadas, de tal forma que todos los alumnos sigan los mismos horarios (Clark, 2013).

4. Los **asynchronous MOOC** o **asynchMOOC** establecen una mayor libertad respecto a fechas y plazos: no suelen tener fecha de finalización; las fechas de comienzo, en caso de existir, son muy frecuentes, y las fechas de entrega de tareas son flexibles. Esto es ideal para personas con poca disponibilidad de horarios, ya que permite a los estudiantes seguir su propio ritmo. En consecuencia, algunas de las ventajas pedagógicas de los asynchMOOC son la posibilidad de ser seguidos en cualquier momento y lugar; su mejor adaptación a los diferentes husos horarios, y una mayor flexibilidad, lo cual contribuye a reducir las tasas de abandono (Clark, 2013; Pilli y Admiraal, 2016).
5. Los **adaptiveMOOC** usan algoritmos adaptativos para la creación de experiencias de aprendizaje personalizadas, basadas en la historia personal, las necesidades y las preferencias de los alumnos. Dichas rutas personalizadas están basadas en redes de prerrequisitos, métodos dinámicos de evaluación y LA, y adicionalmente se utilizan para la mejora continua del curso (Clark, 2013).
6. Los **groupMOOC** (GROOC en adelante) son cursos colaborativos de estructura cerrada y orientados a la institución que los proporciona, cuyo objetivo es la reducción de las tasas de abandono. Al inicio del GROOC, un programa informático se encarga de crear grupos reducidos a partir de las capacidades, el perfil y la procedencia geográfica de los participantes, pero estos grupos pueden ir reconfigurándose a lo largo del curso. Como el currículum del GROOC se desarrolla a partir de las diferentes aportaciones de los participantes, estos trabajan colaborativamente para compartir su conocimiento y experiencias con los demás, adquiriendo de ese modo el papel de co-creadores. Los GROOC suelen contar con un mentor que dirige y organiza el trabajo,

refuerza el compromiso y supervisa el progreso (Clark, 2014; Pilli y Admiraal, 2016; Castrillo et al., 2018).

7. Los **miniMOOC** presentan temporalizaciones más reducidas, de días e incluso de horas, lo cual se extiende a sus contenidos y al desarrollo de habilidades. Los miniMOOC son característicos de la formación *e-learning* comercial y resultan más apropiados para campos y tareas específicos que posean objetivos de aprendizaje claros y precisos. El movimiento *Open Badge* está asociado con este tipo de MOOC (Clark, 2014).

Conole (2014) propone una clasificación de los MOOC en función a un conjunto de doce dimensiones: grado de apertura; escala de participación o masificación; la cantidad de uso de recursos multimedia; la cantidad de comunicación; la extensión en que se promueve la colaboración; el modo de encaminar a los participantes, lo cual iría desde una perspectiva centrada en el estudiante a una altamente estructurada y centrada en el profesor; el nivel de control de la calidad; el grado al que se fomenta la reflexión; el nivel de evaluación; el grado de formalidad; autonomía, y diversidad.

Adicionalmente a los xMOOC, los cMOOC y los GROOC, Castrillo et al. (2018) identifican los siguientes siete tipos de MOOC:

Small Private Online Courses (SPOC en adelante) son cursos dirigidos a grupos reducidos de estudiantes de un perfil bien definido que tienen como objetivo la personalización del aprendizaje. En este tipo de MOOC, el profesor adapta el contenido a las necesidades de los participantes, para así optimizar su fidelidad y permanencia. Para poder lograrlo, se fijan unas ratios reducidas, normalmente de menos de 30 alumnos, y en ocasiones se seleccionan los estudiantes en base a experiencias personales comunes. Los SPOC pueden tener costes de matrícula asociados y, al final de ellos, los participantes tienen la posibilidad de obtener un diploma.

Community/Corporate Open Online Courses (COOC en adelante) son cursos creados *ad hoc* dentro de una empresa u organización, frecuentemente en colaboración con asesores externos, para la formación o el perfeccionamiento en un área prioritaria de las competencias y habilidades de

los propios empleados de la compañía o de los clientes y proveedores (Pilli y Admiraal, 2016; Castrillo et al., 2018).

Los **Distributed Open Collaborative Course** (DOCC en adelante) son cursos orientados hacia actividades institucionales y colaborativas que otorgan relevancia al perfil de las personas con las que se comparte el aprendizaje. Estos MOOC no imparten una programación centralizada y única, sino que están organizados alrededor de una serie de recursos de aprendizaje principales cuyo tema se diversifica en función de los nodos participantes. Al considerarse que el grado de experiencia está distribuido por toda la red de participantes, en los DOCC existe una red mundial de expertos de diferentes contextos institucionales con perfiles tan diversos como los de desarrolladores, activistas o miembros de diferentes grupos sociales o profesionales (Pilli y Admiraal, 2016; Castrillo et al., 2018).

Los **Nano Open Online Courses** (NOOC en adelante) son cursos que se centran, bien en un área de conocimiento específica, bien en un elemento clave de una competencia o bien en una destreza, cubriendo de esta forma la formación altamente especializada que de manera cada vez más frecuente reclaman las empresas. Una característica de estos cursos es la inclusión de actividades que permitan evidenciar el aprendizaje adquirido, el objetivo alcanzado y la competencia desarrollada. Los NOOC son cursos con una duración muy reducida, que puede variar entre 1 y 20 horas, por lo que se consideran pequeñas píldoras formativas especialmente indicadas para participantes que necesiten formación específica dentro de un área. Además de ofrecerse de forma individual, los NOOC pueden integrarse dentro de un programa de aprendizaje más amplio organizado modularmente en forma de cursos escalables y editables (Castrillo et al., 2018).

Los **Self-Paced Open Online Course** (SPOOC en adelante) son cursos masivos, abiertos y gratuitos que operan como una comunidad de conocimiento distribuido, ya que están basados en el aprendizaje social y abierto. Estos cursos no tienen fecha de finalización, por lo que los estudiantes pueden seguirlos a su propio ritmo e ir seleccionando su propia ruta de aprendizaje a través de actividades que generan interacción, producción

agregada, conocimiento compartido y redes profesionales. El objetivo de los SPOOC es el desarrollo de competencias, especialmente las relacionadas con la colaboración en red; la gestión autónoma del aprendizaje, y la participación en comunidades educativas (Pilli y Admiraal, 2016; Castrillo et al., 2018).

Los **Rhizomatic Massive Open Online Courses** (rMOOC en adelante) son cursos que se basan en una filosofía de aprendizaje rizomático, un modelo de aprendizaje en el que el currículo es desarrollado principalmente por la comunidad de estudiantes y en el que se hace un uso amplio y creativo de las redes sociales para propiciar el descubrimiento del conocimiento (Castrillo et al., 2018).

Los **MicroMáster_MOOC** representan una configuración modular de varios MOOC, normalmente cinco, que conforman un Máster de duración reducida que facilita una mayor adaptabilidad a las necesidades formativas de los estudiantes, a la vez que reduce el coste de matrícula a la mitad. Es una de las tendencias surgidas en los últimos años en el MIT⁷, institución que permite a los estudiantes iniciar la formación en el primer semestre con MOOC gratuitos. Si los alumnos quieren continuar con el MicroMáster, deben superar un examen previo de contenidos, abonar una tasa de 150 dólares en concepto de expedición del certificado de los MOOC y seguir el segundo semestre en el campus de forma presencial (Castrillo et al., 2018).

Pilli y Admiraal (2016) dividen los MOOC en función de dos parámetros: apertura o grado de accesibilidad y flexibilidad, y masividad o número de participantes. De este modo, se configuran cuatro grandes categorías MOOC: (I) de pequeña escala y poca apertura, (II) de pequeña escala y mayor apertura, (III) de gran escala y menor apertura y (IV) de gran escala y mayor apertura. A continuación, se describen los tipos de MOOC que los autores incluyen en cada categoría, algunos de los cuales ya han sido analizados anteriormente.

⁷ <https://micromasters.mit.edu/>

PEQUEÑA ESCALA Y POCA APERTURA

Estos MOOC se corresponden con los cursos en línea tradicionales. En ocasiones, un mayor grado de apertura implica que el MOOC se abra a un número limitado de alumnos fuera del campus o que algunas partes del contenido del curso estén disponibles de manera libre. Dentro de esta categoría encontramos los task-based MOOC, los SPOC y los GROOC.

PEQUEÑA ESCALA Y MAYOR APERTURA

Los MOOC de esta categoría cuentan con herramientas de código abierto y en ellos se anima a los participantes a que compartan los resultados de su aprendizaje mediante licencias abiertas. Los MOOC de mayor apertura tienden a emplear redes sociales y de aprendizaje que permiten a los estudiantes desarrollar su autonomía de aprendizaje, pudiendo recurrir a las orientaciones de un instructor. Además de los cMOOC, los COOC, los DOCC y los adaptiveMOOC, ya analizados anteriormente, los siguientes MOOC pertenecen a esta categoría:

Los **Big Open Online Courses** (BOOC en adelante) son cursos abiertos a un número limitado de estudiantes, normalmente 500, pero con ventajas similares a las del resto de MOOC. El objetivo de estos cursos es incrementar la interacción de los participantes, para lo cual se recurre a WikiFolios creados por los estudiantes donde se recogen reflexiones informales, tareas formales bajo demanda e insignias digitales (Pilli y Admiraal, 2016).

El objetivo de los **Participatory Open Online Courses** (POOC en adelante) no es crear cursos masivos, sino cursos participativos que permitan a los estudiantes involucrarse con gente de dentro y fuera del programa. Por este motivo, se espera que los participantes compartan, colaboren y creen conocimiento de forma activa y usando la plataforma del curso (Pilli y Admiraal, 2016).

Los **Little Open Online Courses** (LOOC en adelante) están dirigidos principalmente a estudiantes registrados en programas de la institución que los

ofrece, aunque pueden llegar a admitir hasta cinco alumnos que no pertenezcan a la universidad. Los participantes de los LOOC entregan trabajos, participan en talleres y reciben valoraciones individualizadas y rápidas por parte de los docentes (Pilli y Admiraal, 2016).

Los **Game-Based Learning MOOC** (gMOOC en adelante), propuestos por Sherry Jones y Jack Park, conjugan los principios conectivistas y descentralizados de los cMOOC con juegos digitales, con el objetivo de lograr un aprendizaje basado en la inmersión. Los gMOOC incluyen vídeo conferencias, solución de problemas en grupo, escritos colaborativos, juegos y ejercicios digitales semanales y chats, tanto síncronos como asíncronos (Pilli y Admiraal, 2016).

Los **Project-Based MOOC** (pMOOC en adelante) están basados en el desarrollo colectivo de pequeños proyectos innovadores. Este tipo de MOOC presentan un gran potencial en el campo de los medios y redes de comunicación, ya que su gran número de participantes hace posible abordar nuevos temas, afrontar los problemas educativos o incrementar la investigación en redes sociales (Pilli y Admiraal, 2016).

MAYOR ESCALA Y MENOR APERTURA

Esta categoría de MOOC permite un número ilimitado de participantes, pero se reduce la apertura en varios ámbitos, como las tasas de registro, la duración de los cursos, la personalización de los contenidos, el perfil de los estudiantes, el diseño del curso o los mecanismos de evaluación. Además de los MOOC híbridos, los synchMOOC y los miniMOOC, los siguientes tipos de MOOC pertenecen a esta categoría:

Los **Vocational Open Online Courses** (VOOC en adelante) son cursos de duración muy reducida, típicamente de una hora, y de naturaleza más práctica, basados en la pedagogía empleada en los programas de formación profesional. Como el objetivo de los VOOC es el desarrollo de habilidades, además de a vídeos de expertos explicando los contenidos, estos cursos recurren a tareas prácticas y procedimientos que muestran ejemplos tomados del mundo real (Clark, 2014; Pilli y Admiraal, 2016).

Los **Personalized Open Online Courses** (POOC en adelante) proporcionan educación no estandarizada masiva personalizando los contenidos a través de la tecnología, de forma que el aprendizaje se adapte a las características, preferencias y actividades de los estudiantes. En consecuencia, los materiales, contenidos, valoraciones y métodos de evaluación se formulan en paralelo a las características del andamiaje cognitivo de los participantes con la ayuda de rutas individualizadas de aprendizaje (Pilli y Admiraal, 2016).

MAYOR ESCALA Y APERTURA

Esta categoría de MOOC suele estar altamente centralizada, con poca flexibilidad en los objetivos del curso, los contenidos, la secuenciación y los mecanismos de evaluación, elementos que son determinados de antemano. Además de los xMOOC, los transferMOOC, los madeMOOC, los asynchMOOC y los SPOC, dentro de esta categoría se incluyen los siguientes tipos de MOOC:

Los **flex-MOOC** son cursos centrados en el estudiante que permiten la creación de rutas de aprendizaje propias en función de las fortalezas y necesidades individuales. Para ello, el participante puede elegir entre distintos módulos de contenido y configurar su propia secuencia de aprendizaje y temporalización.

Los **iMOOC** combinan el aprendizaje autónomo con una fuerte dimensión social y están basados en la flexibilidad del estudio, la interacción entre participantes, la inclusión digital y el papel central del alumno. El aprendizaje en los iMOOC se basa en la realización de actividades que impliquen la creación de un producto final, como por ejemplo un texto; un vídeo; una presentación, o un mapa conceptual, que al finalizar el curso es publicado en línea para demostrar el conocimiento y las habilidades de los participantes (Gilliot et al., 2013; Moreira Teixeira y Mota, 2014; Pilli y Admiraal, 2016).

Los **Massive Open Online Courses for Educators** (MOOC-Ed en adelante) son experiencias de desarrollo profesional que ponen a los educadores en contacto con nuevas oportunidades de aprendizaje tecnológicas alrededor de un tema específico. Aunque estos MOOC incluyen recursos principales y materiales adicionales, a la vez permiten un grado elevado de personalización

y flexibilidad, ya que no hay una ruta fija, sino que el estudiante puede modular el ritmo del aprendizaje en función a sus objetivos particulares y las necesidades específicas de su distrito escolar. El aprendizaje se produce a través de actividades por pares, y estudios de caso o proyectos, para los que se cuenta con la guía de facilitadores y de expertos en educación, además de la colaboración de los propios compañeros (Pilli y Admiraal, 2016).

En los **Massive Open Online Research** (MOOR en adelante) participantes de todo el mundo con diferentes historiales de aprendizaje y habilidades colaboran de manera práctica en tareas de investigación. De este modo, los MOOR se postulan como una forma prometedora de combinar el aprendizaje en el aula con la investigación en el mundo real (Pilli y Admiraal, 2016).

TABLA 4. CLASIFICACIÓN DE LOS MOOC SEGÚN LAS TAXONOMÍAS DE LANE (2012), CONOLE (2014), PILLI Y ADMIRAAL (2016), CASTRILLO ET AL. (2018) Y SEIDAMETOVA (2018). ELABORACIÓN PROPIA.

PEQUEÑA ESCALA Y POCA APERTURA					
Lane	Conole	Castrillo et al.			
tMOOC	GROOC	GROOC	SPOC	MicroMáster_MOOC	
Pilli y Admiraal	GROOC	tMOOC	SPOC		
PEQUEÑA ESCALA Y MAYOR APERTURA					
Lane	Seidametova	Conole	Castrillo et al.		
cMOOC	cMOOC	cMOOC	cMOOC	COOC	
Pilli y Admiraal		adaptive	DOCC	SPOOC	
cMOOC	COOC	adaptive	LOOC	gMOOC	
DOCC	BOOC	POOC (Participatory)		pMOOC	
MAYOR ESCALA Y MENOR APERTURA					
Conole	Seidametova	Pilli y Admiraal	SMOOC	POOC (Personalized)	
SMOOC	híbridos	híbridos	VOOC	miniMOOC	
miniMOOC	Castrillo et al.			NOOC	
MAYOR ESCALA Y APERTURA					
Conole		Castrillo et al.			xMOOC
xMOOC	transferMOOC	Pilli y Admiraal	xMOOC	flex-MOOC	SPOC
madeMOOC	asynchMOOC	iMOOC	transferMOOC	MOOC-Ed	
Lane	xMOOC	MOOR	asynchMOOC	madeMOOC	
Seidametova		xMOOC		quasi MOOC	

Tras haber analizado las distintas taxonomías de MOOC, el resto del capítulo profundizará en los MOOC dedicados a la enseñanza de lenguas extranjeras, denominados LMOOC o *Language MOOC*. En principio, los LMOOC podrían diseñarse siguiendo cualquiera de las tipologías de MOOC descritas anteriormente. Sin embargo, la temática particular de los LMOOC los diferencia del resto y dota a estos cursos de unas características especiales que deben tenerse muy presentes a la hora de configurar su diseño.

3.4. LANGUAGE MOOC (LMOOC)

3.4.1. DEFINICIÓN Y CARACTERÍSTICAS

Los LMOOC son cursos dedicados y en línea para segundos idiomas, con acceso sin restricciones y una participación potencialmente ilimitada. Estos cursos constituyen una combinación de prácticas y herramientas que fomentan el uso auténtico y significativo del idioma por parte de los estudiantes, además del desarrollo autónomo de las habilidades en el idioma extranjero. Los LMOOC pueden diseñarse de manera efectiva para facilitar el desarrollo de las competencias lingüísticas comunicativas en grupos potencialmente masivos y altamente heterogéneos cuyo principal interés compartido sea aprender una lengua extranjera. (Martín-Monje y Bárcena, 2014; Sokolik, 2014; «MOOC para la enseñanza de idiomas», 2022).

Read (2014) considera que los LMOOC poseen un gran potencial para desarrollar las competencias lingüísticas, especialmente las productivas e interactivas. Esto se debe a que combinan lo mejor del aprendizaje formal e informal, ya que aúnan el contenido y las actividades educativas de un curso estructurado, y las tecnologías y herramientas basadas en redes sociales adecuadas. Colpaert (2014) considera que, para alcanzar su potencial, los LMOOC necesitan adaptarse a las teorías constructivistas y de adquisición de SL, lo cual los obliga a ser polimórficos; adaptados a los objetivos de aprendizaje, y a los distintos tipos de estudiantes y a sus contextos, sin por ello prescindir de un diseño metodológico, sistemático y globalmente aceptado.

Colpaert (2014) establece cuatro requisitos esenciales para los LMOOC:

- La **modularidad** implica la concepción del LMOOC, no solo como un curso completo, sino como un conjunto concreto de módulos REA que puedan ser fácilmente implementados en cualquier otro entorno de aprendizaje, tanto virtual como presencial.
- La **especialización** significa que los LMOOC no deben adaptarse al máximo número de usuarios posible, sino a todos aquellos con los mismos objetivos de aprendizaje, ya que cuanto más concretos sean dichos objetivos mayor es la capacidad del LMOOC para dar respuesta a las necesidades de los participantes. Este requisito, en combinación con la modularidad, produce LMOOC de carácter práctico y específico.
- La **adaptación** de los LMOOC se produce en tres niveles distintos: (a) el diseño general del LMOOC debe estar basado en un análisis profundo del grupo de estudiantes al que está dirigido, mediante el empleo de LA; (b) el diseño del LMOOC debe ofrecer, en función del perfil de aprendizaje del grupo de participantes, diferentes rutas de aprendizaje; tipos de tareas; valoraciones; niveles, y tipos de apoyo, y (c) el propio sistema debería ser capaz de adaptarse a diferencias individuales en estilos y ritmos de aprendizaje; preferencias gráficas, o elementos de accesibilidad.
- La **co-construcción** implica la combinación dentro del LMOOC de los conocimientos y habilidades de los propios participantes, y es una de las claves para que los LMOOC sean capaces de sobrevivir.

Algunos de los primeros LMOOC, los cuales comenzaron a aparecer en octubre de 2012, se originaron en España, más particularmente en la UNED. Entre los primeros cursos se encuentran *Iniciando el inglés: aprender las primeras mil palabras*, que contó con unos 45 mil estudiantes; *Inglés Profesional / Professional English*, con más de 50 mil estudiantes, y *Alemán para hablantes de español*, ganador en 2013 del premio al mejor MOOC en la I Edición de los premios Miráda X (Universia España, 2013; Ventura et al., 2014; Conde Gafaro, 2020). Fuera de España, uno de los ejemplos pioneros dentro del campo de los LMOOC fue el curso *SpanishMOOC*, dirigido por Ryan y Scott Rapp, e inaugurado en enero de 2013. Este LMOOC empleaba un sistema de aprendizaje adaptativo para personalizar las valoraciones y la secuenciación del contenido, y recurría a la aplicación *Google Hangouts* para la práctica de las destrezas de

producción oral (Godwin-Jones, 2014; Martín-Monje et al., 2014). Otro ejemplo son los dos LMOOC creados por Todd Bryant en 2013 para su sitio web *The Mixer*⁸, los cuales actualmente están integrados en el campus virtual de la Dickinson College University. Estos LMOOC son *English MOOC: Open Course for Spanish Speakers Learning English*, un LMOOC de inglés para hispanoparlantes, y *MOOC de Español: Curso abierto para hablantes de inglés que deseen mejorar su español* (Blyth, 2013; Martín-Monje et al., 2014).

Los LMOOC se presentan como un campo de estudio incipiente, con un número creciente de investigaciones que abordan distintos aspectos técnicos, pedagógicos y metodológicos de los LMOOC. Algunas de las más relevantes son los trabajos monográficos a cargo de Martín-Monje y Bárcena (2014), y de Dixon y Thomas (2015), los cuales se han convertido en obras de referencia en el estudio de los LMOOC. También es digno de destacar el trabajo sistemático de revisión bibliográfica sobre LMOOC a cargo de Sallam et al. (2020), en el que se identifican los principales tipos de publicaciones relativas a LMOOC; los países, instituciones y autores punteros en investigación sobre LMOOC; los trabajos con mayor impacto, y las distintas categorías temáticas dentro de la investigación en LMOOC.

A partir de los primeros LMOOC, en su mayoría cursos introductorios a las lenguas que han sido tradicionalmente objeto de aprendizaje, la temática de los LMOOC se ha ido diversificando y se han creado LMOOC diseñados para el desarrollo de destrezas o elementos específicos dentro del aprendizaje de una lengua. De este modo, los LMOOC se han utilizado con éxito en el aprendizaje de vocabulario (Ventura y Martín-Monje, 2016; Chacón-Beltrán, 2018), en el desarrollo de la producción oral (García y Appel, 2016; Ramanarayanan et al., 2018), de la pronunciación (Rubio, 2014; Solans García y Estebas Vilaplana, 2020), de la producción escrita (Whitmer et al., 2014; McCorkle et al., 2016; Xinying, 2017; Gilliland et al., 2018; Kandeel, 2019) y de la comprensión escrita (Xinying, 2017).

En los últimos años también se han creado LMOOC orientados a lenguas para fines específicos y al aprendizaje de lenguas minoritarias, los cuales han

⁸ <https://www.language-exchanges.org/es>

sido objeto de diversas investigaciones. De hecho, Vorobyeva (2018) defiende los LMOOC como una herramienta beneficiosa en el estudio de lenguas menos comunes, incluso en nivel principiante, y teoriza sobre la potencialidad de los LMOOC para la preservación de lenguas en riesgo de desaparición. Uno de los estudios llevados a cabo sobre LMOOC de lenguas para fines específicos es el de Shalatska (2018), quien concluye que este tipo de cursos ayuda a los estudiantes a crear una base de lenguaje profesional; a demostrar dominio de un idioma extranjero, y a crear condiciones para el desarrollo activo y el pensamiento independiente. También Chen (2017) analiza cómo la combinación de un LMOOC de inglés para fines específicos con portafolios digitales contribuye a la mejora de la competencia lingüística y la autonomía de los estudiantes. El estudio de Ventura y Martín-Monje (2016) revela que la inclusión de actividades en Facebook dentro de un LMOOC de inglés para fines específicos produjo un impacto positivo en la motivación de los estudiantes para aprender vocabulario especializado y una reducción de las altas tasas de abandono. Otro trabajo pionero es el de Orsini-Jones et al. (2017), quienes adoptan un enfoque de *blended learning* al combinar con éxito un LMOOC y clases presenciales dentro de un programa de formación para profesores de inglés como lengua extranjera.

Como se ha analizado anteriormente, la apertura es uno de los pilares conceptuales dentro del movimiento de los MOOC. Por este motivo, son numerosos los estudios que analizan distintos aspectos e implicaciones del concepto de apertura en los LMOOC. De este modo, Díez-Arcón (2021) analiza los perfiles socioeconómicos de los participantes de LMOOC y corrobora las barreras existentes para que este formato sea accesible a una extensión de la población más amplia y heterogénea. Por su parte, Rodrigo (2014), partiendo del potencial de los MOOC para la inclusión social, aborda la cuestión de la accesibilidad de los LMOOC y propone un modelo de diseño que permite que las personas con necesidades especiales puedan disfrutar de las oportunidades formativas que ofrecen los LMOOC. También están enmarcados dentro del campo de la integración social los trabajos de Castrillo y Sedano (2021), y de Bárcena et al. (2020), publicados a partir de los resultados de las dos ediciones de *Puertas Abiertas*, un LMOOC de español centrado en las necesidades lingüísticas inmediatas de inmigrantes y refugiados que llegan a España, cuya

primera edición será la fuente de datos de esta investigación. La investigación de Castrillo y Sedano (2021) muestra la efectividad de la creación de sinergias y comunidades entre académicos; trabajadores de Organizaciones no Gubernamentales y de Grupos de Apoyo a Refugiados; voluntarios, y refugiados e inmigrantes, ya que se observa una mejora en las tasas de finalización de los cursos y un elevado grado de satisfacción de los participantes respecto a la consecución de los objetivos de integración social perseguidos por ambos LMOOC. Por su parte, el trabajo de Bárcena et al. (2020) desarrolla el concepto de lenguaje inclusivo y demuestra los beneficios de su utilización dentro de los LMOOC. La cuestión de la internacionalización también ha sido analizada por varios autores. Uno de ellos es Godwin-Jones (2014), quien propone soluciones a problemas tales como la falta de infraestructuras en países subdesarrollados, la hegemonía del inglés en este tipo de cursos y la falta de familiaridad de los estudiantes con las metodologías en línea. Otros estudios como los de Troncarelli y Villarini (2017), y McLoughlin y Magnoni (2017) analizan los beneficios de los LMOOC en el caso particular de la movilidad de estudiantes entre universidades. También relativo al concepto de apertura de los LMOOC, Álvarez (2014) analiza las implicaciones éticas del diseño de los LMOOC, las cuales engloban las ideas de justicia social; derechos humanos; libertad, y accesibilidad y teoriza sobre el grado en que los LMOOC pueden promover o desalentar la ética y la estética en la enseñanza de idiomas.

Un gran porcentaje de la investigación en el campo de los LMOOC analiza la relación entre el éxito en la finalización de estos cursos y distintas características de los participantes tales como la motivación (Beaven et al., 2014; Henry y Marrs, 2015; Fuchs, 2016); las interacciones entre los estudiantes en los foros de la plataforma (Santos et al., 2014; Martín-Monje et al., 2017; McMinn, 2017) y en redes sociales (Ventura et al., 2014; Ventura y Martín-Monje, 2016; Nic Giolla Mhichíl et al., 2018); el desarrollo y reconocimiento de destrezas metacognitivas (McLoughlin y Magnoni, 2017); las estrategias de aprendizaje (Bárkányi, 2018; Chacón-Beltrán, 2018); las actitudes y creencias (Bárkányi y Melchor-Couto, 2017; Orsini-Jones et al., 2017; Xinying, 2017; Wang et al., 2018); el compromiso (Ventura y Martín-Monje, 2016), o los estilos de aprendizaje y los estilos cognitivos de los participantes (Hashim et al., 2018).

Existe a su vez un subgrupo de estudios que indaga en los beneficios de la combinación de los LMOOC con distintos tipos de metodologías y tecnologías. Algunos de ellos investigan la integración de los LMOOC dentro de las *flipped classroom* y concluyen que el empleo de LMOOC en dichos entornos metodológicos promueve las habilidades de aprendizaje activo, autónomo y colaborativo de los estudiantes; mejora las actitudes de los alumnos y la comprensión de los entornos digitales de aprendizaje centrados en el estudiante, y contribuye al desarrollo competencial (Xinying, 2017; Wang et al., 2018). La integración de los LMOOC dentro del *blended learning*, lo cual implica combinar un LMOOC con clases presenciales, también ha sido objeto de estudio en los últimos años. Según Bartalesi-Graf (2017), esta metodología atrajo a un mayor número de alumnos y produjo un incremento en la satisfacción de los estudiantes que participaron de la experiencia educativa. Orsini-Jones et al. (2017) concluyen que la inclusión de LMOOC en el currículo de un programa de *blended learning* contribuye a un reajuste por parte de los estudiantes respecto a creencias en aspectos como la gramática o el aprendizaje en línea. Dentro del campo de *Mobile Assisted Language Learning* (MALL en adelante), Ibáñez Moreno y Traxler (2016) proponen los estándares técnicos, pedagógicos y lingüísticos requeridos para el diseño de un LMOOC basado en MALL. Por su parte, Read y Bárcena (2020) proponen un marco teórico para la enseñanza y el aprendizaje de idiomas basado en el Marco Común Europeo de Referencia para las Lenguas (MCERL en adelante), el cual proporciona la estructura necesaria para la elaboración sistemática de LMOOC, programas de estudio y materiales didácticos para aplicaciones MALL. Este marco también posibilita la estructuración del conocimiento y las habilidades requeridas en el aprendizaje de una segunda lengua de manera coherente con el paradigma constructivista, facilitando el aprendizaje en línea de un segundo idioma. Otros casos de metodologías e innovaciones tecnológicas que han sido combinadas con LMOOC por diversos autores son: la gamificación, para la cual Metwally y Yining (2017) proponen un enfoque teórico multimodal aplicado al aprendizaje del idioma y la cultura chinos que evidencia un aumento en la motivación para completar el curso por parte del alumno, y una reducción de las tasas de abandono; las redes sociales, cuya inclusión contribuye a una mejora en la motivación y el compromiso de los estudiantes (Ventura y Martín-Monje, 2016);

los portafolios digitales, los cuales, en combinación con los LMOOC, fomentan la autonomía de los alumnos; permiten reflejar sistemáticamente el grado de participación de los estudiantes en el proceso de aprendizaje, y posibilitan la monitorización del progreso de los participantes mediante un sistema de evaluación formativa que proporciona información valiosa sobre la evolución competencial (Chen, 2017), o el aprendizaje cooperativo, para el cual Teixeira y Mota (2014) proponen un diseño específico de LMOOC.

De manera adicional, en los últimos años las instituciones gubernamentales han mostrado su interés por los LMOOC y han surgido varios proyectos para su estudio y difusión. Uno de ellos es el proyecto *LangMOOC*, en colaboración con *Open Education Europa*⁹, un proyecto europeo que se centra específicamente en los LMOOC. El Proyecto ECO, encuadrado en el programa marco europeo para la competitividad y la innovación, trata de documentar las mejores prácticas de los enfoques de diseño MOOC e incluye varios LMOOC en su catálogo¹⁰.

3.4.2. PRINCIPALES CRÍTICAS A LOS LMOOC

Ciertos aspectos tanto de los MOOC en general como de los LMOOC en particular han sido objeto de crítica por parte de investigadores y especialistas. De hecho, los LMOOC aúnan los problemas intrínsecos comunes a todos los MOOC con las dificultades específicas de la enseñanza a distancia de segundas lenguas. Uno de los principales problemas de los MOOC es su elevada tasa de abandono, la cual oscila entre el 75 y el 95% (Poy y Gonzales-Aguilar, 2014). Cabero (2015) asocia las elevadas tasas de abandono a tres factores principales: la tipología de los participantes; su debilidad en ciertas competencias, como la digital o la autorregulación del aprendizaje, y el diseño de los MOOC.

Algunas de las principales críticas hacia los MOOC se refieren a la dicotomía xMOOC y cMOOC, y a los inconvenientes educativos de cada uno de estos dos principales grupos. Numerosos críticos rechazan el diseño de los

⁹ <http://www.langmooc.com/>

¹⁰ <https://eco-learning.eu/>

xMOOC, argumentando que representan un retroceso pedagógico, ya que intentan replicar el modelo universitario tradicional mediante cursos en línea, lo cual aún carencias metodológicas con los problemas de individualización relativos al componente masivo de los MOOC. Otro de los aspectos más polémicos de los xMOOC es una evaluación basada en pruebas de conocimiento que no valoran la participación ni la interacción del estudiante con sus compañeros y con el docente (Cabero, 2015). Por otro lado, muchos participantes de cMOOC se quejan de la confusión creada por la multitud de rutas de comunicación y de que, en ocasiones, el tiempo real requerido para completar los cursos supera con creces lo programado inicialmente (Martí, 2012; Conole, 2014).

Respecto a la supuesta apertura de los MOOC, se ha cuestionado que los cursos sean completamente abiertos, ya que la mayoría de los participantes de los MOOC son jóvenes (predominantemente varones) de países desarrollados; con educación superior y empleados, y que no se enfrentan a barreras económicas para acceder a la educación superior. Adicionalmente, en los países en vías de desarrollo y muy especialmente en las zonas rurales, la infraestructura existente imposibilita el acceso a los recursos de aprendizaje en línea (Christensen et al., 2013; Liyanagunawardena, Williams et al., 2013).

Respecto a las acreditaciones de los MOOC, uno de los principales inconvenientes es que normalmente no son gratuitas y suelen carecer de reconocimiento en el mercado cultural, académico y económico (Lane y Kinser, 2012; Walton et al., 2014). Por último, el componente masivo de los MOOC también ha sido puesto en tela de juicio al acusarse a estos cursos de homogeneizar la educación, limitando la conciencia global y la diversidad de experiencias educativas (Lane y Kinser, 2012).

Las peculiaridades intrínsecas al aprendizaje de un idioma, como la necesidad de interacción; la práctica de las destrezas comunicativas, o la corrección de los errores de uso, plantean retos añadidos en el diseño de los LMOOC (Amorós-Negre et al., 2018). El hecho de que un LMOOC sea realmente masivo dificulta ciertos aspectos del aprendizaje de una lengua, entre ellos el aprendizaje inferencial de las estructuras gramaticales; la competencia

comunicativa en entornos reales, o las dimensiones cognitivas y afectivas del aprendizaje de segundas lenguas. De manera adicional, el gran número de participantes dificulta la corrección de tareas y la comunicación de valoraciones personalizadas a los alumnos (Stevens, 2013; Godwin-Jones, 2014; Read, 2014). Incluso en los LMOOC verdaderamente comunicativos, donde los estudiantes utilizan el medio de instrucción como medio de comunicación, tanto la carga cognitiva como la complejidad de las tareas se ven incrementadas (Martín-Monje y Bárcena, 2014). Numerosos autores coinciden en que la clave para que los LMOOC superen los inconvenientes señalados por las críticas está en un correcto diseño. Algunos autores proponen recurrir a un modelo híbrido que aproveche las ventajas sociales de los xMOOC y la escalabilidad de los cMOOC (Moreira Teixeira y Mota, 2014; Godwin-Jones, 2014; Read, 2014; Panagiotidis, 2019).

3.4.3. DISEÑO DE UN LMOOC

El diseño adecuado de un LMOOC representa un reto singular, ya que aúna las dificultades de la creación de un MOOC con las particularidades pedagógicas de la enseñanza de segundas lenguas. De este modo, los criterios a tener en cuenta en el diseño de un MOOC pueden dividirse en dos grandes categorías: criterios pedagógicos y criterios técnicos (Yousef et al., 2014).

Dentro de los criterios pedagógicos, dos de los elementos principales son el diseño instruccional y los mecanismos de evaluación. Castrillo et al. (2018) destacan la importancia de una definición clara tanto del diseño instruccional como de los recursos con los que se trabajará cada tema. El diseño instruccional debe partir de unos objetivos de aprendizaje concretos, proporcionados a los participantes del curso de una forma clara y accesible. La masividad de estos cursos hace que la pluralidad de los perfiles de los estudiantes aumente y que sus objetivos a la hora de seguir el MOOC sean distintos. De este modo, el éxito del LMOOC dependerá de la medida en que los objetivos del curso se adapten a los objetivos particulares de los participantes en cuanto al propósito y el grado al que quieren aprender la lengua. Por este motivo, es importante animar a los estudiantes a que fijen sus propias metas y que utilicen el LMOOC para conseguirlas (Martín-Monje y Bárcena, 2014).

Como se ha mencionado previamente, las dos grandes categorías pedagógicas dentro de los MOOC son los xMOOC y los cMOOC, ambas con ventajas e inconvenientes. Los cMOOC ofrecen contenido en red y fomentan, siguiendo los postulados constructivistas, la auto organización de los participantes y el progreso colectivo, mientras que los xMOOC centralizan el contenido en una plataforma. Según Bárcena et al. (2015), aunque los cMOOC comparten algunas características con el enfoque comunicativo en la enseñanza de lenguas extranjeras, ya que ambos ponen énfasis en la interacción y la construcción de una comunidad, existen evidencias que muestran que el paradigma conectivista no encaja bien cuando se usa como modelo único para la enseñanza de idiomas en línea. Por otro lado, los xMOOC poseen la estructura de un curso universitario estándar, al contar con un plan de estudios; una secuencia de actividades, y mecanismos de evaluación, por lo que gran parte de los participantes estarán familiarizados con el funcionamiento de estos cursos.

A este respecto, Bárcena et al. (2015) rechazan la suposición de que los cMOOC sean mejores que los xMOOC y argumentan que lo que requieren los LMOOC es una plataforma que sea lo suficientemente flexible como para integrar lo mejor de ambos tipos. Godwin-Jones (2014) afirma que el enfoque óptimo para estructurar un LMOOC es proporcionar un sistema de aprendizaje adaptativo dentro de un extenso entorno de aprendizaje social personalizable, combinando un formato de estilo xMOOC con un cMOOC. La combinación entre el aprendizaje por ordenador y el aprendizaje social dependerá de los niveles de competencia y de las habilidades que se aprenden o evalúan, con una mayor importancia de la dimensión social a medida que se desarrolla la competencia.

Independientemente de la tipología de LMOOC implementada, la estructura del curso debe ser el primer elemento en diseñarse y debe dividirse en módulos, siendo conveniente incluir un Módulo 0 que permita que la comunidad de aprendizaje comience a crearse. Aunque tradicionalmente la duración de los MOOC oscila entre las 6 y las 8 semanas, Castrillo et al. (2018) recomiendan una duración de 3 a 4 semanas, con una carga semanal de trabajo de entre 3 y 5 horas, aunque ello implique realizar varios cursos de

diferentes niveles en lugar de un único curso más extenso. Estas autoras también recomiendan redactar un programa que explicita los objetivos del LMOOC y los requisitos previos, además de proporcionar a los participantes una guía de estudios y un cronograma con una propuesta de secuenciación de contenidos. Otros elementos aconsejables son: comenzar cada módulo con un vídeo de presentación; dividir los módulos en distintos temas que incluyan vídeos y ejercicios interrelacionados; incluir materiales adicionales, tanto obligatorios como opcionales; introducir distintos tipos de actividades, enfatizando siempre cuáles son obligatorias y cuáles opcionales, y realizar pruebas al final de cada módulo (UTEID, 2014, Narang, 2020).

La elección de la plataforma en la que se implementará el LMOOC implica la selección del portal educativo donde se va a difundir e impartir el MOOC, y de la herramienta de creación de cursos que se va a utilizar, lo cual tendrá consecuencias importantes tanto en el diseño instruccional como en la parte técnica (UTEID, 2014; Castrillo et al., 2018). Existen tres opciones principales en el diseño de la arquitectura *software* de un MOOC. La primera es recurrir a una de las plataformas ofrecidas por los proveedores de MOOC, la segunda es proporcionar acceso abierto a un VLE ya existente y la tercera es implementar un espacio web propio.

A este respecto, Read (2018) plantea tres preguntas iniciales que deben responderse antes de comenzar a desarrollar un LMOOC. La primera de ellas es si el LMOOC necesita una plataforma distinta a la de otros MOOC. La respuesta a esta cuestión depende en gran medida de la tipología del curso y el enfoque pedagógico. Por este motivo, una plataforma de un proveedor de MOOC que permita la implementación de los tipos de técnicas y actividades exigidos por el enfoque adoptado constituiría una elección adecuada, que además contará con el apoyo y soporte técnico que ofrecen este tipo de plataformas (Castrillo et al., 2018). La segunda pregunta concierne a la robustez y escalabilidad de la plataforma seleccionada. Aunque es cierto que la implementación y organización propia de una plataforma confiere mayor libertad a la hora de estructurar el curso, el elevado número de participantes en el curso supone un consumo de recursos y un volumen de intercambio de datos que probablemente requiera un diseño de arquitectura que solo un proveedor de

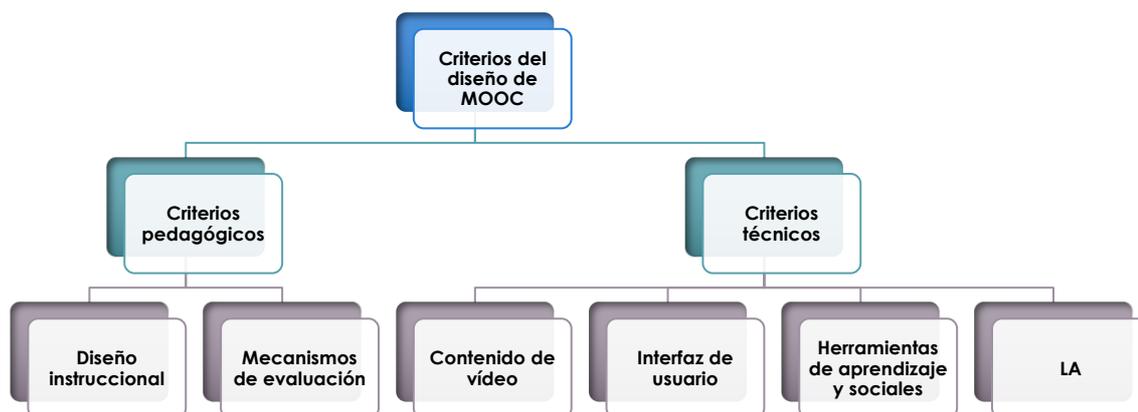
MOOC o un VLE ya existente pueden proporcionar (Read, 2018). La tercera cuestión planteada es la idoneidad de eliminar las restricciones de un VLE para convertirlo en una plataforma de LMOOC. La respuesta a esta pregunta depende de nuevo del enfoque pedagógico, ya que no todos los VLE permiten incorporar las herramientas y la filosofía social de la Web 2.0, lo cual limitaría la efectividad del curso (Read, 2018).

Los mecanismos de evaluación de un MOOC dependen en gran medida de la tipología del curso. De este modo, los xMOOC suelen recurrir a cuestionarios de evaluación tipo test, mientras que en los cMOOC la evaluación tiende a orientarse hacia la realización de proyectos o trabajos en línea (UTEID, 2014). Los mecanismos apropiados de evaluación en los LMOOC suponen un gran reto, ya que es difícil evaluar la producción lingüística de los alumnos mediante preguntas tipo test. Por ese motivo, los dos modos de evaluación más ampliamente utilizados son la autoevaluación, en la que la evaluación se dirige a la comprensión y asimilación individual de los contenidos teóricos del curso, y la evaluación por pares o *peer-to-peer* (P2P en adelante), que incide en el componente social y colaborativo del proceso de aprendizaje, ya que los participantes son los encargados de corregir los trabajos de otros compañeros (Martín-Monje y Bárcena, 2014; Castrillo et al., 2018).

El sistema automatizado de los cuestionarios de autoevaluación resulta muy útil a la hora de evaluar a miles de participantes y las estadísticas muestran su incidencia en que un mayor porcentaje de participantes finalicen el curso. Los cuestionarios de autoevaluación pueden formar parte de la evaluación inicial, en la que los participantes tienen la posibilidad de comprobar sus conocimientos antes de iniciar el curso; de la evaluación formativa, en la que los alumnos pueden comprobar su nivel de comprensión del contenido del curso durante el proceso de aprendizaje, y de la evaluación sumativa al final de los módulos y del curso. Hay ciertos aspectos de los cuestionarios que suelen ser configurables en función de la plataforma, entre ellos la tipología de las preguntas; el número de intentos; la visualización de las respuestas correctas; el umbral de superación, o el tiempo disponible para completarlos (UTEID, 2014; Castrillo et al., 2018).

En las tareas P2P, los participantes revisan y evalúan el trabajo de otros compañeros del curso, a menudo con ayuda de una rúbrica o matriz de evaluación. Castrillo et al. (2018) defienden que las tareas P2P enriquecen sumamente el proceso de aprendizaje y por lo tanto deben constituir un elemento central en ellos. El motivo es que, por un lado, estas tareas basan el aprendizaje en la interacción social y, por otro, presentan una solución a la inviabilidad de la evaluación de todas las tareas de los participantes por parte del equipo docente. En las tareas P2P, el número de trabajos que cada participante se encarga de evaluar depende en última instancia del número total de participantes en el curso y del número de trabajos entregados, aunque lo habitual es que cada participante valore el trabajo de tres compañeros. Para facilitar el proceso, es importante que el profesor proporcione a los alumnos rúbricas sencillas y bien construidas, además de instrucciones claras sobre los aspectos que deben ser valorados (UTEID, 2014). A la hora de programar este tipo de actividades, es importante tener en cuenta el mayor esfuerzo que requieren de parte de los estudiantes, ya que además de completar sus propias tareas los alumnos deben dedicar tiempo a corregir las tareas que les sean asignadas. Por este motivo, a la hora de diseñar el curso, es muy importante que las tareas P2P se incorporen en los momentos correctos. En lugar de sobrecargar todos los módulos con ellas, Castrillo et al. (2018) aconsejan incluirlas en los módulos intermedios en vez de próximas al cierre del curso, además de programar tiempo suficiente entre la entrega y la evaluación de las tareas para garantizar que a los alumnos les dé tiempo suficiente a completarlas. A pesar de sus múltiples beneficios, la evaluación por pares presenta varios inconvenientes, como la falta de disposición de algunos participantes, las disparidades de nivel o incluso sesgo hacia alumnos de la misma nacionalidad. Por ese motivo, una conjunción de ejercicios con autocorrección, autoevaluación estructurada y valoraciones voluntarias por parte de compañeros se presenta como la forma más adecuada de evaluar los LMOOC (Martín-Monje y Bárcena, 2014). En la figura 11 se muestra un diagrama resumen con los principales criterios a tener en cuenta en el diseño de los LMOOC.

FIGURA 11. CRITERIOS PARA EL DISEÑO DE LOS MOOC.



Nota. Adaptado de *What Drives a Successful MOOC? An Empirical Examination of Criteria to Assure Design Quality of MOOCs*, por A. Yousef, M. Chatti, U. Schroeder y M. Wosnitza, 2014, *2014 IEEE 14th International Conference on Advanced Learning Technologies*, 44-48. Copyright 2014 by The Institute of Electrical and Electronics Engineers, Inc.

Entre los criterios técnicos, algunos de los elementos más importantes del diseño son el contenido de vídeo; la interfaz de usuario; las herramientas de aprendizaje y sociales, y LA (Yousef et al., 2014). El papel de los recursos audiovisuales es fundamental para acrecentar el compromiso en los LMOOC, ya que constituyen un marco metodológico apropiado para explicar los objetos de aprendizaje a nivel práctico. El uso de vídeos implica la implementación de estrategias heurísticas, tales como el razonamiento por analogía, para ayudar a los estudiantes a desarrollar la competencia lingüística. De este modo, los vídeos empleados en los LMOOC deben sumergir a los participantes en la cultura y la lengua estudiadas mediante un entorno que fomente la reflexión y la conversación. Para ello, los vídeos deben contener materiales auténticos, y situaciones y objetos de la vida real usados como ayudas para el aprendizaje, lo cual implica evitar recurrir a las clases magistrales (Castrillo, 2014; Martín-Monje y Bárcena, 2014; Gimeno-Sanz et al., 2017).

Castrillo et al. (2018) dividen los tipos de vídeos empleados en un MOOC en tres tipos:

- «Busto parlante» representa el formato más tradicional y consiste en un vídeo que contiene un plano frontal del presentador ofreciendo una clase magistral.
- Los *screencast* son superposiciones de capturas de pantalla y narraciones en formato audio que explican el procedimiento representado por las imágenes. Este es el tipo de vídeo más indicado para demostraciones y simulaciones.
- Las animaciones e infografías representan de manera gráfica situaciones y relaciones conceptuales y, además de para la enseñanza de lenguas extranjeras, suelen emplearse para profundizar en conceptos complejos y procesos.

A esta clasificación puede añadirse un cuarto tipo de vídeo empleado específicamente en la enseñanza de lenguas: las representaciones de diálogos y conversaciones. Este tipo de vídeos están adaptados al nivel de los participantes del curso, ya que utilizan un vocabulario acotado y una velocidad de habla más pausada, además de presentar situaciones extraídas de la vida real.

Los elementos gráficos son una parte fundamental de los vídeos, por lo que se recomienda incluir tanto texto explicativo que los contextualice en pantalla como elementos gráficos que destaquen las ideas principales del contenido. También es recomendable añadir material complementario que permita al alumno diferentes niveles de profundización en forma de enlaces web o documentos descargables, lo cual requiere una revisión periódica para garantizar que el acceso a los recursos esté actualizado. Es aconsejable que el LMOOC cuente con un vídeo de presentación que explique los objetivos genéricos y específicos del curso, así como la metodología; el modelo de evaluación, y la tipología de actividades. De manera adicional, en los vídeos de LMOOC también es crucial proporcionar subtítulos detallados que permitan a los estudiantes seguir el contenido (Castrillo et al., 2018).

La *Guía para creación de MOOC de la Universidad Carlos III de Madrid* (UTEID, 2014) recomienda que los vídeos estén preparados un mes antes del inicio de curso, para lo que se debe comenzar a trabajar en los contenidos con

cuatro meses de antelación. Para que en los vídeos tipo «busto parlante» el contenido se transmita con entusiasmo; naturalidad, y de forma relajada y clara, Castrillo et al. (2018) recomiendan elaborar un esquema con puntos clave, en vez de un guion completo; incluir preguntas retóricas que ayuden a reflexionar sobre lo que está escuchando, y comenzar el vídeo con un resumen del contenido que incluya toda la información relevante para los alumnos.

Respecto a la duración de los vídeos, parece existir consenso en que no deben superar los 15 minutos. Por ejemplo, la *Guía para el docente de la Universidad Politécnica de Madrid* (GATE, s.f.) recomienda una duración de cada vídeo de entre 3 y 15 minutos, con una duración total de todos los vídeos de un módulo de en torno a 60 minutos. De manera similar, Khan (2012) afirma que el límite de atención de los estudiantes se sitúa entre los 10 y los 15 minutos. Por este motivo, Castrillo et al. (2018) recomiendan que la duración de los vídeos no supere los 5 minutos, ya que después de este tiempo la atención del individuo decae.

Otro de los criterios técnicos del diseño de un MOOC es la interfaz de usuario, la cual depende de la elección de la plataforma en la que se implementará el LMOOC, elemento ya analizado anteriormente. Relativo a este criterio, al ser los MOOC medios de instrucción basados principalmente en los vídeos, es importante que la plataforma elegida permita al estudiante el control sobre las opciones de visualización de los vídeos, de tal modo que los estudiantes puedan modificar la velocidad de reproducción; detenerse en ciertos pasajes; volver a visualizar otros, y revisar los contenidos. También es importante que la interfaz incluya una herramienta de búsqueda potente que ayude a los alumnos a encontrar fácilmente los materiales del curso (Yousef et al., 2014; Castrillo et al., 2018). Respecto a los foros, es importante que los estudiantes sepan que hay un instructor participando en el LMOOC, ya que de esta forma se establecen vínculos más estrechos entre docentes y participantes, y se transmite la sensación de presencia continuada por parte del equipo docente. Por consiguiente, es conveniente que la plataforma de LMOOC incluya la opción de visualizar los momentos en los que los instructores y facilitadores están conectados y disponibles para contactar con ellos (Martín-Monje y Bárcena, 2014).

Las herramientas sociales y las herramientas de aprendizaje también forman parte de los criterios técnicos del diseño de un MOOC. Dentro de las herramientas de aprendizaje, se recomienda la elaboración de materiales complementarios que, por un lado, ayuden a reforzar el conocimiento de los estudiantes avanzados y, por otro, faciliten el aprendizaje de estructuras lingüísticas para los estudiantes que no están familiarizados con el aprendizaje en línea. La heterogeneidad del alumnado de los LMOOC requiere el diseño de una estructura de andamiaje que tenga en cuenta la escalabilidad del conocimiento y que a la vez sea personalizable, ya que es conveniente que los estudiantes tengan la opción de seleccionar los contenidos de su interés y puedan así crear su propio camino de aprendizaje (Castrillo, 2014). A este respecto, Castrillo et al. (2018) recomiendan incluir recursos de apoyo gratuitos que primen la calidad sobre la cantidad y que sean accesibles para el mayor número posible de sistemas operativos de los dispositivos móviles. De manera similar, Stacey (2014) recomienda implementar el mayor grado posible de apertura al recurrir a pedagogías abiertas que aprovechen toda la web y utilicen los REA.

La diferencia principal entre los LMOOC y otros tipos de MOOC son las herramientas requeridas para el desarrollo de las competencias comunicativas. En la pedagogía de la enseñanza de lenguas extranjeras, la interacción en forma de comunicación auténtica, principalmente entre los propios estudiantes, es un elemento esencial, que además aumenta el grado de compromiso. Por este motivo, las herramientas sociales son un elemento particularmente importante en los LMOOC (Martín-Monje y Bárcena, 2014; Read, 2014). En cualquier curso en línea y especialmente en un LMOOC, es fundamental establecer diferentes vías de comunicación entre profesor y participante, y también entre los propios estudiantes. Las herramientas de comunicación deben ser cuidadosamente diseñadas y planificadas por parte del equipo docente, de modo que la comunicación fluya de manera organizada y estructurada, y los estudiantes puedan progresar sin incidentes y de manera flexible, además de poder plantear preguntas relacionadas con su proceso de aprendizaje (Castrillo, 2014) Martín-Monje y Bárcena (2014) destacan la importancia de crear presencia docente en el curso, ya que la figura del profesor cobra un papel destacado en los LMOOC al impulsar la creación de

una comunidad entre los participantes; comentar las intervenciones en los foros, y hacer anuncios de forma regular.

Las herramientas de comunicación pueden dividirse en dos tipos: sincrónicas y asincrónicas. Dentro de las herramientas sincrónicas, las más usadas son las aplicaciones de videoconferencia, como Zoom; Google Hangouts, o Skype. La ventaja de este tipo de herramientas es que permiten programar encuentros en directo entre profesores y participantes del MOOC. Sin embargo, las herramientas sincrónicas presentan serios inconvenientes que deben tenerse en cuenta. Debido al elemento masivo de los MOOC, estas herramientas no son capaces de soportar la conexión de un número elevado de estudiantes, cuyos horarios, por otra parte, pueden ser difíciles de coordinar debido a las diferentes zonas horarias (Castrillo et al., 2018). Castrillo (2014) considera que las herramientas síncronas en entornos masivos no suelen ser efectivas debido a la dificultad de originar interacciones cognitivas de alto nivel. Por este motivo, algunas alternativas son grabar la sesión como un webinar o seminario virtual, o realizar un debate virtual con un experto a través de Twitter u otro canal (Castrillo et al., 2018).

Las herramientas asincrónicas más utilizadas en los LMOOC son las siguientes:

- Los **foros** son espacios que se utilizan tanto para el intercambio de ideas como para plantear dudas. Constituyen la herramienta de comunicación más utilizada en los MOOC, ya que permiten la participación masiva de los estudiantes (UTEID, 2014). Los foros pueden llegar a ser una herramienta dinamizadora, ya que poseen el potencial de generar debates entre la comunidad de estudiantes, lo cual enriquece los contenidos con distintos puntos de vista (Castrillo et al., 2018). Es importante realizar un diseño correcto de los foros, de tal forma que sean capaces de abarcar todos los tipos posibles de problemas planteados. Castrillo (2014) recomienda diseñar una estructura rígida basada en categorías, subcategorías e hilos de conversación, para que así los temas estén distribuidos y se evite que el foro derive en caos. Algunas prácticas recomendables son incluir un foro general en el que se

consulten tanto cuestiones generales relacionadas con la organización del curso como problemas técnicos; crear categorías para cada una de las unidades temáticas; incluir al principio del foro consejos y normas de conducta; animar a los participantes a que valoren positiva o negativamente los mensajes, con el objetivo de facilitar su selección cualitativa, o impedir que los participantes puedan crear nuevas categorías (UTEID, 2014; Castrillo et al., 2018).

- Los **blogs** suelen utilizarse como complemento y diario de aprendizaje, y como tablón de anuncios en el que se publica información como las novedades; las fechas relevantes, o reflexiones sobre algún contenido. Las diferentes publicaciones pueden tratar sobre el nuevo material a disposición de los participantes; la forma de realizar las distintas actividades, o la aclaración de preguntas frecuentes planteadas en los foros o en las redes sociales. En los blogs, los contenidos suelen estar organizados en orden cronológico inverso, de modo que se muestra primero el contenido más reciente, y la comunicación se produce por medio de comentarios (UTEID, 2014; Castrillo et al., 2018).
- Las **wikis** son espacios de trabajo colaborativo que pueden ser editados por los usuarios, lo cual permite fomentar la participación y la comunicación entre los alumnos. Lo más habitual es proporcionar una estructura definida que los estudiantes van dotando de contenido a través de sus aportaciones. La mayoría de los proveedores de MOOC permiten integrar tanto blogs como wikis dentro de la plataforma, aunque también existe la posibilidad de recurrir a servicios externos (UTEID, 2014; Castrillo et al., 2018).
- Las **redes sociales** fomentan la interacción en línea y la implicación del alumnado en el curso, y generan verdaderas comunidades virtuales de aprendizaje. Las más utilizadas en los LMOOC son Twitter y Facebook. Una de las principales razones del éxito de Twitter es que limita los mensajes a 280 caracteres, lo cual obliga a los participantes a condensar lo que quieren transmitir. Para integrar esta red social, se recomienda abrir una cuenta a nombre del LMOOC desde la que se publicitarán contenidos, se fomentará el intercambio de opiniones y se monitorizará el comportamiento de los participantes. También es aconsejable utilizar un

acrónimo o las siglas del LMOOC a modo de hashtag que aglutine todos los tuits relacionados con los contenidos del curso. De esta forma, los usuarios serán capaces de encontrar fácilmente los mensajes o tuits más relevantes. La otra gran red social incluida frecuentemente en los LMOOC es Facebook, desde la cual existe la posibilidad de alertar de información relevante, como nuevos posts en el blog; nuevos contenidos en el MOOC, o fechas de entregas. De hecho, ante la tendencia entre los participantes de organizar grupos de discusión ajenos a la plataforma del MOOC en esta red social, algunos autores consideran conveniente que las plataformas de LMOOC propicien la organización de estos grupos; sugieran encuentros en persona; permitan el seguimiento de las publicaciones creadas por los participantes, y se habilite la posibilidad de que los estudiantes compartan sus audios y vídeos, ya que la conversación y la interacción son parte fundamental del aprendizaje en los LMOOC (UTEID, 2014; Martín-Monje y Bárcena, 2014; Castrillo et al., 2018).

La elección del tipo de herramienta de comunicación que se empleará en el LMOOC debe determinarse en función de los contenidos; la tipología del curso, y el modelo de evaluación, teniendo en cuenta las ventajas y los inconvenientes de cada herramienta. Los xMOOC suelen emplear foros de discusión, mientras que los cMOOC recurren frecuentemente a redes sociales para propiciar la comunicación, las cuales pueden plantear problemas de privacidad o restricciones gubernamentales en ciertas regiones. Independientemente del tipo de herramienta elegido, se debe proporcionar a los participantes pautas de comunicación claras y precisas con relación al uso de las herramientas, y asegurarse de que los estudiantes sepan siempre a dónde acudir para resolver cualquier cuestión o problema que surja. De manera adicional, es conveniente informar a los participantes sobre el riesgo de sobrecarga de información inherente a los MOOC. Deben ser conscientes de que no es necesario leer todos los mensajes, sino que deben seleccionar los que mejor se alineen con sus intereses (Martín-Monje y Bárcena, 2014; Castrillo et al., 2018)

Por último, como se analizará en profundidad en el siguiente capítulo, las herramientas de LA permiten supervisar el proceso de enseñanza-aprendizaje; descubrir los patrones de rendimiento, uso y aprendizaje de los participantes; identificar sus principales dificultades, y proporcionar a los alumnos valoraciones que les ayuden a reflexionar sobre su propia experiencia educativa. Por este motivo, es en la fase de diseño del LMOOC cuando deben considerarse las diferentes posibilidades y herramientas de LA que posteriormente permitirán la medición y el análisis del rendimiento del estudiante (Yousef et al., 2014, Castrillo et al., 2018).

CAPÍTULO 4: *LEARNING ANALYTICS* (LA)

4.1. INTRODUCCIÓN

En la sociedad digitalizada actual, gran parte de nuestras transacciones comerciales y bancarias; modos de entretenimiento; búsquedas de información, y comunicaciones personales y profesionales tienen lugar a través de internet. Todos nuestros dispositivos electrónicos, como teléfonos inteligentes; tabletas; ordenadores; lectores de libros electrónicos, o rastreadores de actividad física, están facilitando la acumulación de grandes cantidades de datos en las empresas, la industria, el gobierno y otras áreas de la actividad humana. Estos conjuntos de datos, que sobrepasan la capacidad habitual de los programas informáticos para capturarlos; almacenarlos; administrarlos, y analizarlos, se denominan *big data* (Manyika et al., 2011; Sclater, 2017). En consecuencia, en los últimos años se ha producido un auge en el desarrollo de técnicas de procesamiento y análisis que permiten recopilar y extraer información del *big data*. Estas técnicas contribuyen a mejorar la comprensión de las actividades personales, con el objetivo de optimizar los procesos y resultados de las organizaciones. De esta forma, las empresas son capaces de identificar consumidores potenciales, ajustar sus productos a las necesidades de los consumidores o adaptar sus estrategias de marketing en función a preferencias individuales (Bienkowski et al., 2012; Sclater, 2017).

Como se ha analizado en los capítulos anteriores, la evolución de Internet ha cambiado drásticamente los métodos y sistemas educativos, y ha propiciado el que una gran cantidad de instituciones educativas, principalmente universidades, hayan desarrollado una vasta oferta de cursos en línea. De este modo, existen cada vez más conjuntos de *big data*, y más fácilmente accesibles, sobre los estudiantes; sus actividades de aprendizaje, y los entornos en los que estudian. En los últimos años, el sector educativo ha sabido ver el enorme potencial subyacente del análisis de estos datos generados por los estudiantes. Como resultado, los educadores están comenzando a comprender cómo aprovechar esta información para ayudar a resolver algunos de los desafíos a los que se enfrentan sus instituciones (Sclater, 2017).

En los últimos años, son varias las disciplinas que se han desarrollado en torno al análisis de los datos generados por los sistemas educativos. Una de ellas es LA, la cual se ha erigido como una de las tendencias de futuro en educación más citadas en los informes internacionales de los últimos años (Lang et al., 2017; León Urrutia et al., 2017; Alexander et al., 2019; Brown et al., 2020). El objetivo de las LA es emplear los datos sobre los estudiantes y sus contextos de aprendizaje para mejorar muchos aspectos de la experiencia educativa (Sclater, 2017; Lester et al., 2018; Maselena et al., 2018; Alexander et al., 2019; Ifenthaler et al., 2019; Brown et al., 2020; Romero y Ventura, 2020).

Este capítulo está dedicado a la descripción y el análisis de las características de las LA. Para ello, en el siguiente apartado se ofrecerá una definición de LA, y se comparará y delimitará con respecto a otros ámbitos y disciplinas afines. Posteriormente, se ofrecerá una breve descripción de los orígenes y la evolución histórica de las LA. A continuación, se analizarán aspectos como las aplicaciones educativas más importantes de las LA; su metodología, o los retos y limitaciones a los que se enfrenta la disciplina. El capítulo concluirá con un análisis de los proyectos y aplicaciones más relevantes de las LA dentro del campo de los LMOOC.

4.2. DEFINICIÓN

Una de las definiciones del término LA, y la más aceptada actualmente, fue la que se ofreció en la primera conferencia internacional sobre analíticas de aprendizaje, organizada por la Sociedad para la investigación de Analíticas de Aprendizaje o, en inglés, *Society for Learning Analytics Research* (SoLAR en adelante). Allí se definió LA como la medida, recogida, análisis e informe de datos sobre estudiantes y sus contextos, con el objetivo de comprender y optimizar el aprendizaje y los entornos en los que este ocurre (Gasevic et al., 2010). Para Durall et al., las LA consisten en «la interpretación de un amplio rango de datos producidos y recogidos acerca de los estudiantes para orientar su progresión académica; predecir actuaciones futuras, e identificar elementos problemáticos» (2012, p.13), mientras que para Maselena et al. (2018) son la aplicación de técnicas analíticas en el análisis de datos educativos, lo cual incluye proporcionar datos sobre las actividades de alumnos y docentes;

identificar patrones de comportamiento, y proporcionar información factible que permita mejorar el aprendizaje y las actividades relacionadas con el aprendizaje. Johnson et al. (2011) definen LA como la interpretación de una amplia gama de datos producidos y recopilados a partir de los estudiantes, con el objetivo de evaluar su progreso académico; predecir su desempeño futuro, y detectar problemas potenciales. Estos datos se recopilan a partir de, por un lado, acciones específicas por parte de los estudiantes, tales como la realización de tareas y exámenes, y, por otro lado, de otras acciones que se consideran tácitas, entre las que se incluyen interacciones sociales en línea; actividades extracurriculares; publicaciones en foros de discusión, y otras actividades que no son evaluadas directamente como parte del progreso educativo del estudiante.

Existen tres disciplinas próximas a LA cuyos ámbitos en ocasiones llegan a solaparse, por lo que es necesario diferenciarlas y delimitarlas. Estas son: Analíticas Académicas o *Academic Analytics* (AA en adelante); investigación-acción, y Minería de Datos Educativos o *Educational Data Mining* (EDM en adelante).

El término AA fue adoptado por Goldstein y Katz (2005) para describir la implementación de técnicas de inteligencia empresarial en la educación superior. En su estudio, estos autores usan el término AA para describir la recopilación, análisis y utilización de datos por parte de las instituciones académicas durante el proceso de toma de decisiones. La principal diferencia de AA con respecto a LA es que los datos de alumnos, académicos e institucionales, son utilizados para la mejora de los procesos organizativos; los flujos de trabajo; la asignación de recursos, y la medición institucional, mientras que los objetivos principales de LA están más alineados con aspectos enmarcados dentro del proceso de enseñanza-aprendizaje, tales como la reflexión; la adaptación; la personalización, y la transmisión de recomendaciones. Por lo tanto, las LA son más específicas que AA, ya que la primera se centra exclusivamente en el proceso de aprendizaje, mientras que la segunda lo hace en el análisis de datos a nivel institucional (Long y Siemens, 2011; Chatti, et al., 2012; Maselena et al., 2018). De hecho, el término AA se emplea particularmente en los Estados Unidos para describir el análisis de

funciones administrativas de alto nivel en las universidades, como por ejemplo el reclutamiento o la predicción del éxito académico, mientras que LA es más usado para aspectos relativos al aprendizaje y a la experiencia educativa de los estudiantes (Sclater, 2017).

LA también están relacionadas con la investigación-acción, ya que ambas disciplinas intentan mejorar la educación a través de investigaciones cíclicas (Dyckhoff, 2011). La investigación-acción es un proceso de investigación sistemática llevado a cabo por personas dentro de una comunidad educativa, en lugar de expertos externos, cuyo objetivo es identificar acciones conducentes a la generación de mejoras educativas (Hinchey, 2008). Aunque los objetivos de la investigación-acción y de las LA son muy similares, los proyectos de investigación-acción intentan abordar un problema que surge de la práctica docente, mientras que los proyectos que utilizan LA evolucionan a partir de observaciones sobre los datos recolectados. De manera adicional, la investigación-acción suele utilizar datos cualitativos para conformar una visión holística de la situación de aprendizaje, mientras que LA se basa principalmente en métodos cuantitativos. Finalmente, la investigación-acción involucra habitualmente solo a profesores y alumnos, mientras que en LA suele haber otras partes implicadas, como diseñadores de sistemas y personal institucional (Chatti et al., 2012).

Aunque en su origen la EDM y las LA estuvieron entroncadas, actualmente su comunidad de investigadores es distinta (Sclater, 2017). La EDM desarrolla y adapta métodos estadísticos; técnicas algorítmicas; de aprendizaje automático, y de minería de datos para estudiar los datos educativos generados por estudiantes e instructores, con el objetivo de comprender a los estudiantes y los entornos en los que estos aprenden. De esta forma, es posible mejorar el modelado de estudiantes y la estructura del conocimiento en el que se basan los procesos de enseñanza (Baker and Yacef, 2009; Baker et al., 2012; Calvet Liñán y Juan Pérez, 2015; International Educational Data Mining Society, s. f.). Por lo tanto, a diferencia de otras formas de minería de datos, la EDM tiene en cuenta los aspectos pedagógicos tanto del alumno como del sistema en el que se produce el aprendizaje (Sclater, 2017).

FIGURA 12. TABLA COMPARATIVA DE LAS PRINCIPALES DISCIPLINAS QUE REALIZAN ANÁLISIS DE DATOS EDUCATIVOS. ELABORACIÓN PROPIA.

	LEARNING ANALYTICS	EDUCATIONAL DATA MINING	ANALÍTICAS ACADÉMICAS	INVESTIGACIÓN- ACCIÓN
OBJETIVO	Mejora del proceso de enseñanza-aprendizaje	Mejora del proceso de enseñanza-aprendizaje	Mejora de los procesos organizativos	Mejoras educativas
AGENTES	Profesores, alumnos, técnicos, instituciones	Profesores, alumnos, técnicos, instituciones	Profesores, alumnos, técnicos, instituciones	Profesores y alumnos
TIPO DE DATOS	Cuantitativos	Cuantitativos	Cuantitativos	Cualitativos
TIPO DE INVESTIGACIÓN	Análisis del aprendizaje y el entorno	Búsqueda de patrones y desarrollo de algoritmos	Análisis de funciones administrativas	Problemas de la práctica docente

LA y EDM tienen muchos aspectos en común, ya que ambas pertenecen al ámbito educativo y trabajan con datos procedentes de dicho entorno; entienden los problemas en la educación de forma similar; tienen como objetivo la mejora del proceso de aprendizaje, la evaluación y la calidad del análisis de datos educativos a gran escala con el propósito de apoyar la investigación básica y la práctica educativa; se asemejan en el modo de planificar y seleccionar las intervenciones, ya que cuentan con una metodología de proyectos similar, y comparten algunos de los métodos empleados en el tratamiento de los datos (Chatti et al., 2012; Siemens y Baker, 2012). Sin embargo, existe una serie de diferencias que permiten delimitar ambos campos de conocimiento. Generalmente, la minería de datos educativos busca nuevos patrones en los datos y desarrolla nuevos algoritmos y modelos, mientras que las LA aplica métodos y modelos conocidos para responder a preguntas importantes que afectan al aprendizaje de los estudiantes y a los sistemas de aprendizaje organizacional. Por lo tanto, mientras que las LA se centran en la toma de decisiones basadas en la interpretación de los datos y en la integración de las dimensiones técnicas; sociales, y pedagógicas del aprendizaje, mediante la aplicación de modelos predictivos conocidos, la EDM está más orientada al desarrollo de métodos para explorar los tipos únicos de datos que provienen de entornos educativos; a la búsqueda de nuevos patrones en los datos, y al

desarrollo de nuevos algoritmos y modelos. Por lo tanto, podría decirse que las LA están enfocadas en el desafío educativo, mientras que la EDM lo está en el desafío tecnológico (Bienkowski, et al., 2012; Calvet Liñán y Juan Pérez, 2015; International Educational Data Mining Society, s.f.; Romero y Ventura, 2020). De este modo, según Siemens y Baker (2012), es posible identificar cinco diferencias clave entre EDM y LA:

- **Descubrimiento:** en ambas disciplinas se pueden encontrar investigaciones que utilizan tanto el juicio humano, mediante la visualización y otros métodos, como el descubrimiento automatizado. Sin embargo, en la EDM existe una clara preferencia por el descubrimiento automatizado, mientras que en las LA se tiende a priorizar el juicio humano.
- **Holismo:** la EDM reduce los sistemas a componentes, los cuales explora y analiza junto a las relaciones entre ellos, mientras que el objetivo de LA es comprender los sistemas de forma global, profundizando en su complejidad.
- **Orígenes:** la EDM tiene sus raíces en el *software* educativo y el modelado de estudiantes, mientras que los orígenes de las LA están relacionados con la web semántica, el «currículo inteligente», la predicción de resultados y las intervenciones sistémicas.
- **Adaptación y personalización:** los modelos de EDM se suelen emplear como base de adaptaciones automatizadas realizadas por un sistema informático (por ejemplo, un sistema de tutoría inteligente), mientras que los modelos de LA suelen estar diseñados para informar y empoderar tanto a instructores como a estudiantes.
- **Técnicas y métodos:** la EDM emplea más técnicas y métodos de clasificación; agrupamiento; modelado bayesiano; minería de relaciones; descubrimiento con modelos, y visualización, mientras que las LA se centran en el análisis de redes sociales; análisis de sentimiento; análisis de influencias; análisis del discurso; predicción del éxito del alumno; análisis de conceptos, y modelos de relaciones lógicas.

FIGURA 13. COMPARATIVA DE LA Y EDM. ELABORACIÓN PROPIA



Según Calvet Liñán y Juan Pérez (2015), la popularidad y el rápido desarrollo de las LA se debe a cuatro factores principales: (1) el gran interés en emplear un enfoque basado en datos para llevar a cabo un seguimiento y control de la actividad en el campus que desemboque en la toma de decisiones estratégicas que orienten la planificación y las intervenciones

(Daradoumis et al., 2010; Durall et al., 2012; Siemens y Baker, 2012); (2) la existencia de poderosos métodos y técnicas estadísticas; de aprendizaje automático, y de extracción de datos que permiten encontrar patrones en los datos y construir modelos predictivos o reglas de decisión que puedan adaptarse fácilmente a los datos educativos; (3) la relativa facilidad para la generación de datos y la capacidad informática actual que permite su almacenamiento y procesamiento, y (4) la presión a la que están sometidas las universidades y otras entidades educativas para reducir costes y aumentar los ingresos mediante la reducción de las tasas de abandono y la mejora de la calidad de los cursos, aprovechando también las crecientes demandas educativas de los países en desarrollo.

4.3. ORÍGENES DE LA

La EDM, cuyo origen puede datarse en 1995, nació de la necesidad de analizar los *logs* generados en los intercambios de información entre los estudiantes y los servidores de los VLE, especialmente desde el desarrollo de la web 2.0 (Romero y Ventura, 2007; Ferguson, 2012). A partir de 2003, además del análisis basado en datos, comenzaron a surgir enfoques de análisis con orientación social y pedagógica, con especial relevancia de la integración del Análisis de Redes Sociales o *Social Network Analysis* (SNA en adelante), un método desarrollado en las ciencias sociales que permitía investigar las relaciones entre los integrantes de redes. A partir de 2008, el SNA comenzó a tener una mayor orientación pedagógica, y a centrarse en la comprensión y la optimización del aprendizaje (Aviv et al., 2003; De Laat et al., 2007; Haythornthwaite y de Laat, 2010; Ferguson, 2012). En 2006 se publicó el primer libro sobre temas relacionados con la EDM y las LA, el cual se tituló *Data Mining in E-Learning* (Romero y Ventura, 2006). Para 2007, el campo de las LA, aún subsumido junto al de las AA dentro de la EDM, experimentó un mayor grado de desarrollo conforme los investigadores recibían fondos para abordar nuevos desafíos, tanto educativos como tecnológicos. La influencia de los impulsores políticos en el campo de las analíticas, junto con la creciente madurez del campo de la EDM, que celebró su primera conferencia internacional en Montreal en 2008, propiciaron una división entre las dos disciplinas y, por consiguiente, una divergencia de ambas literaturas. En 2010 el campo de las

analíticas comenzó a dividirse una vez más y las LA se separaron gradualmente de las AA, división que culminó con la celebración de las primeras conferencias sobre LA, en las cuales se inició el desarrollo de comités directivos y de programas globales (Ferguson, 2012). Como resultado, el campo de las LA actualmente alcanza a cubrir más disciplinas que el originario de la EDM, ya que, además de la informática; la estadística; la psicología, y las ciencias del aprendizaje, las LA están relacionadas con la ciencia de la información y la sociología, e incluye algunos de sus métodos y técnicas (Bienkowski et al., 2012; Chatti et al., 2012). De hecho, durante los primeros años de desarrollo de ambas disciplinas, de 2006 a 2014, en los títulos de las obras de investigación abundaban los términos «Minería de Datos en Educación» y «Minería de Datos Educativos», pero posteriormente, de 2015 a 2017, comenzó a emplearse ampliamente el término «Learning Analytics», el cual se ha convertido en el más utilizado en la bibliografía de los últimos años. Por este motivo, autores como Dormezil et al. (2019) concluyen que actualmente es más preciso describir lo que parece ser dos dominios como un solo dominio, el de las LA, con el subconjunto destacado de la EDM.

En 2011, más de 100 participantes asistieron en Banff, Canadá, a la primera Conferencia Internacional sobre LA, cuyas actas fueron publicadas y se convirtieron en una de las primeras publicaciones en analizar este nuevo campo de estudio (Gasevic et al., 2011). Uno de los principales impulsores de la conferencia fue SoLAR¹¹, una sociedad profesional fundada en verano de 2011 para supervisar la conferencia; desarrollar y promover una agenda de investigación en LA, y educar en el uso de LA. Hasta la fecha, SoLAR se ha continuado encargando de la organización de las subsiguientes conferencias anuales sobre LA (Romero y Ventura, 2020).

El desarrollo de las LA se vio impulsado por su inclusión en el Informe *Horizon* del influyente *New Media Consortium*, cuyo número de 2011, centrado en las tecnologías emergentes; su impacto potencial, y su uso en la enseñanza; el aprendizaje, y la investigación creativa, identificó las LA como una de las nuevas tecnologías a observar (Johnson et al., 2011). El Informe *Horizon* de 2012

¹¹ <http://www.solaresearch.org/>

(Johnson et al., 2012) volvió a incluir las LA, considerándolas de adopción generalizada por las instituciones educativas en un periodo de dos a tres años. Por este motivo, agencias nacionales como *Jisc* en el Reino Unido; SURF en los Países Bajos, y la Oficina de Aprendizaje y Enseñanza del gobierno australiano, o transnacionales, como la Comisión Europea, comenzaron a organizar colaboraciones y eventos, y a financiar proyectos y actividades de investigación y desarrollo para desarrollar y promover las LA (Sclater, 2017).

A partir de 2014, y para complementar sus conferencias anuales, SoLAR comenzó a publicar el *Journal of Learning Analytics*. Asimismo, varias revistas de tecnología educativa y otros campos han publicado números especiales sobre LA a lo largo de los últimos años, entre ellas la serie de publicaciones producidas por EDUCAUSE (Siemens y Baker, 2012; Calvet Liñán y Juan Pérez, 2015; Sclater, 2017; Romero y Ventura, 2020). A partir de 2014, también comenzaron a aparecer libros y manuales dedicados enteramente al campo de las LA, como *Learning Analytics: From research to practice* (Larsson y White, 2014); *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (Daniel, 2016); *Developing effective educational experiences through learning analytics* (Anderson y Gavan, 2016); *Data Mining and Learning Analytics: Applications in Educational Research* (ElAtia et al., 2016); *Learning Analytics Explained* (Sclater, 2017); *Handbook of Learning Analytics* (Lang et al., 2017), una de las obras más importantes sobre LA hasta la fecha; *Learning Analytics Goes to School. A Collaborative Approach to Improving Education* (Krumm et al., 2018); *Learning Analytics in the Classroom. Translating Learning Analytics Research for Teachers* (Lodge et al., 2018); *Learning Analytics in Education* (Niemi et al., 2018); *Learning Analytics in Higher Education: Current Innovations, Future Potential, and Practical Applications* (Lester et al., 2018); *Utilizing Learning Analytics to Support Study Success* (Ifenthaler et al., 2019), y *Machine Learning Paradigms: Advances in Learning Analytics* (Virvou et al., 2019).

4.4. APLICACIONES DE LAS LA

Las nuevas formas de aprendizaje en línea generan un gran volumen de datos, entre los que se encuentran los datos relativos a los antecedentes de los participantes; datos de rendimiento; entrega de tareas; lectura de archivos;

visionado de vídeos; participación en foros; envío de mensajes, o consulta de enlaces recomendados. Gracias a las LA, esta ingente huella digital ofrece a investigadores, y a las ciencias del aprendizaje y de la educación la oportunidad de explorar el aprendizaje desde múltiples ángulos nuevos. La aplicación de técnicas de LA sobre los datos generados por los estudiantes permite trazar sus redes sociales; sus patrones de navegación y de pausas; sus hábitos de lectura y escritura, y sus mecanismos de aprendizaje y comprensión de los conceptos básicos de un curso. De manera adicional, estos datos ofrecen la posibilidad de extraer información valiosa para la mejora del rendimiento de los estudiantes, además de proporcionar valor predictivo sobre los tipos de patrones de aprendizaje y de actividad que podrían indicar riesgo de fracaso o abandono (Siemens y Gasevic, 2012; Siemens, 2013; Calvet Liñán y Juan Pérez, 2015; Romero y Ventura, 2020).

El objetivo de las LA es posibilitar que los docentes adapten las estrategias educativas al nivel de necesidad y capacidad de cada alumno de forma rápida y eficaz. En consecuencia, las LA proporcionan un nuevo modelo para que los líderes de instituciones educativas diseñen una mejor asignación de recursos humanos y materiales, y así mejoren la enseñanza; el aprendizaje; la eficiencia organizacional, y la toma de decisiones, elementos fundamentales para el cambio sistémico que se requiere de la educación actualmente (Long y Siemens, 2011; Durall et al., 2012; Calvet Liñán y Juan Pérez, 2015). De esta forma, las LA resultan sumamente provechosas para docentes; instituciones educativas, e investigadores, ya que los primeros las utilizan para obtener conocimientos y optimizar los procesos de aprendizaje de sus alumnos; los segundos consiguen aumentar la tasa de retención de estudiantes, mejorar su éxito en el aprendizaje y aliviar la carga de responsabilidad, y los últimos pueden probar y adaptar sus teorías basándose en datos educativos (Calvet Liñán y Juan Pérez, 2015; Sclater, 2017; Maselena et al., 2018).

Para Durall et al. (2012), la relevancia de las LA para la docencia, el aprendizaje, la investigación y la gestión de la información radica en tres puntos principales. En primer lugar, la información aportada por las LA contribuye a la personalización de la acción formativa y al diseño de entornos de aprendizaje acordes con las necesidades; intereses, y formas de interacción entre profesores

y estudiantes. Las LA pueden, a partir del estudio de los resultados iniciales, identificar estudiantes con un alto grado de motivación. De este modo, desde el primer momento el diseño del curso puede modificarse y adaptarse a los casos particulares de los estudiantes, y así dar una mejor respuesta a sus diferentes patrones de aprendizaje. Adicionalmente, los estudiantes pueden recibir recomendaciones sobre recursos de acuerdo con su desempeño, metas y motivaciones; pueden analizar gráficamente los resultados de su proceso de aprendizaje; compararlos con los del resto de la clase, y observar el desempeño y las contribuciones relacionadas con las actividades colaborativas (Johnson et al., 2011; Durall et al., 2012; Calvet Liñán y Juan Pérez, 2015; León Urrutia et al., 2017; Sclater, 2017). En segundo lugar, el registro de los procesos de aprendizaje posibilita que los docentes centren sus esfuerzos en el diseño y análisis de los procesos formativos en lugar de en los materiales (Durall et al., 2012). Por último, el registro estadístico de la actividad de estudiantes y docentes permite la identificación de puntos conflictivos dentro del proceso de enseñanza-aprendizaje, lo cual contribuye a su mejora continua. De esta forma, al ser capaces de combinar información de fuentes dispares para crear un perfil de estudiantes mucho más sólido y matizado, las LA ofrecen a los profesores conocimientos más detallados y completos acerca de sus estudiantes. Esta información puede ayudar a predecir cuándo un estudiante va a abandonar el curso, mediante la identificación de puntos de intervención en los que los estudiantes dejan de participar y corren el riesgo de abandonar el curso, además de contribuir a la rápida identificación de aquellos alumnos que tienen dificultades para seguir el curso, y de las unidades que generan más confusión, (Durall et al., 2012; Nunn et al., 2016; Sclater, 2017; Romero y Ventura, 2020).

Long y Siemens (2011) identifican nueve formas en las que las LA pueden proporcionar valor añadido a la educación superior:

- 1) Mejora de la toma de decisiones administrativas y la asignación de recursos organizacionales.
- 2) Identificación de los estudiantes en riesgo de abandono con el objetivo de intervenir para ayudarlos a finalizar el curso con éxito.
- 3) Creación de un entendimiento compartido de los éxitos y desafíos de la institución.

- 4) Innovación y transformación del sistema de colegios y universidades; de los modelos académicos, y de los enfoques pedagógicos.
- 5) Aporte de claridad a desafíos y temas complejos, mediante la combinación de redes sociales; técnicas, e información.
- 6) Ayuda a los líderes en la toma holística de decisiones, a través del análisis de escenarios hipotéticos que exploran cómo se conectan los diversos elementos dentro de una disciplina compleja.
- 7) Incremento de la productividad y la eficacia de las organizaciones, al proporcionar información actualizada y permitir una respuesta rápida a los desafíos.
- 8) Ayuda a los líderes institucionales a determinar el valor tangible (patentes, investigaciones) e intangible (reputación, perfil, calidad de la enseñanza) generado por la actividad docente.
- 9) Ofrecimiento a los alumnos de una perspectiva de sus propios hábitos de aprendizaje mediante recomendaciones sobre cómo mejorar.

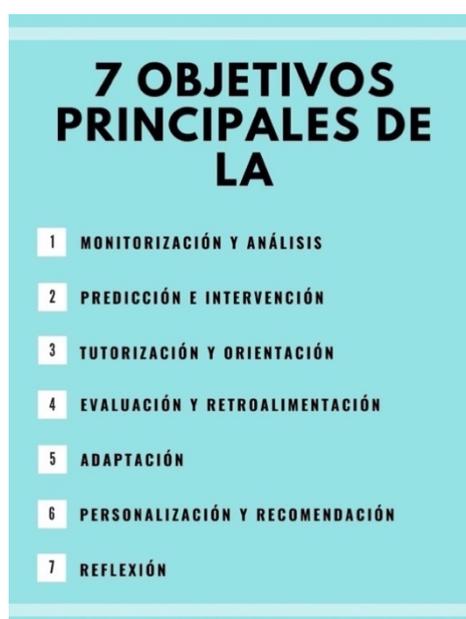
Por su parte, Chatti et al. (2012) identifican siete objetivos de las LA:

- 1) **Monitorización y análisis:** el propósito de la monitorización es el rastreo de las actividades de los estudiantes y la generación de informes que contribuyan a la toma de decisiones por parte del docente o de la institución educativa. La monitorización también está relacionada con el diseño instruccional y se refiere a la evaluación del proceso de aprendizaje por parte del docente con el propósito de contribuir a su mejora continua. Examinar cómo los estudiantes usan un sistema de aprendizaje y analizar los logros de los estudiantes puede ayudar a los profesores a detectar patrones y a tomar decisiones sobre el diseño futuro de la actividad de aprendizaje.
- 2) **Predicción e intervención:** el objetivo de la predicción es el desarrollo de un modelo que se base en las actividades y logros de los alumnos para intentar predecir sus conocimientos y desempeños futuros. De este modo, en el caso de los alumnos que requieran una ayuda adicional, se pueden utilizar modelos predictivos para sugerir acciones que proporcionen una intervención temprana y proactiva que contribuya a que tales alumnos puedan mejorar su desempeño.

- 3) **Tutorización y orientación académica:** el foco principal de la tutorización es ayudar en el aprendizaje y las tareas de los estudiantes, y se limita a menudo a un dominio concreto y a un solo curso académico. Un tutor, por ejemplo, orienta a los alumnos, proporciona instrucciones respecto a las distintas áreas temáticas de un curso e introduce nuevos módulos de aprendizaje. En los procesos de tutoría, el tutor es quien asume el control del proceso y el foco se orienta hacia el proceso de enseñanza. Por su parte, la orientación académica va más allá de la tutorización, y proporciona al alumno apoyo y guía durante todo el proceso de enseñanza-aprendizaje, habitualmente durante el tiempo en que el mentor y el alumno forman parte de la misma institución educativa. Entre las acciones que se incluyen en este proceso están la orientación en la planificación de la carrera; la supervisión del grado de logro de las metas académicas, o la preparación del alumno para desafíos específicos dentro del contexto institucional y educativo. En los procesos de orientación académica, el alumno es quien asume el control del proceso y el foco se orienta hacia el proceso de aprendizaje.
- 4) **Evaluación y retroalimentación:** las LA pueden mejorar la eficiencia de los mecanismos de autoevaluación por parte de los alumnos durante su proceso de aprendizaje, especialmente mediante la retroalimentación inteligente. A partir de los datos relativos a los intereses del usuario y su contexto de aprendizaje, la retroalimentación inteligente proporciona información valiosa sobre el proceso de aprendizaje, tanto para los estudiantes como para los profesores y mentores.
- 5) **Adaptación:** mediante las LA es posible realizar una organización adaptativa de los recursos de aprendizaje y las actividades de instrucción de acuerdo con las necesidades de cada alumno, de tal forma que los sistemas de tutorización o las instituciones educativas puedan indicar a los alumnos cuáles son los siguientes pasos dentro de su proceso formativo.

- 6) **Personalización y recomendación:** las LA son una poderosa herramienta para la personalización del aprendizaje, ya que permiten que este se centre en gran medida en los alumnos al ayudarles a decidir sobre su propio aprendizaje y a moldear continuamente sus entornos de aprendizaje para lograr sus objetivos particulares. De este modo, las LA contribuyen a un cambio de énfasis desde un modelo de aprendizaje en el que el flujo de información es dirigido por la institución y el docente a otro modelo en el que es el propio alumno quien navega hacia el conocimiento. A este respecto, las LA permiten gestionar la posible sobrecarga de información mediante la implementación de sistemas de recomendación que fomenten el aprendizaje auto dirigido. Estos sistemas son capaces de recomendar a los estudiantes, en función de sus preferencias y las actividades de otros estudiantes con preferencias similares, nodos de conocimiento explícito (recursos de aprendizaje) y tácito (personas).
- 7) **Reflexión:** las LA se erigen como una herramienta valiosa para promover la reflexión por parte de estudiantes y profesores, ya que les permite comparar datos dentro de un mismo curso; entre clases, o incluso entre instituciones, y de este modo obtener conclusiones y reflexionar sobre la eficacia de su proceso de aprendizaje y de su práctica docente, respectivamente.

FIGURA 14. OBJETIVOS PRINCIPALES DE LA.

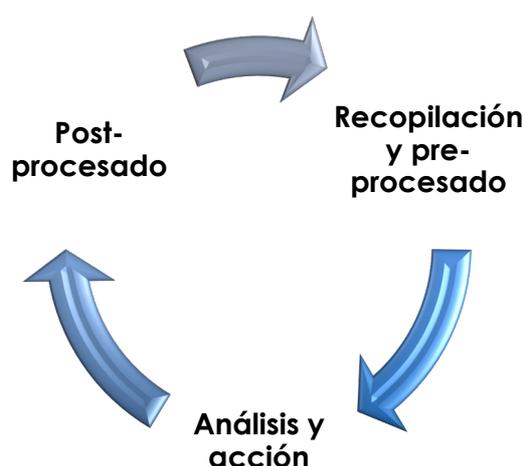


Nota. Adaptado de A Reference Model for Learning Analytics, por M. A. Chatti, A.L. Dyckhoff, U. Schroeder y H. Thüs, 2012, *International Journal of Technology Enhanced Learning*, 4(5/6), 326–327. Copyright 2012 by M.A. Chatti, A.L. Dyckhoff, U. Schroeder and H. Thüs.

4.5. METODOLOGÍA DE LA

Como en cualquier otro proyecto de investigación, una clara identificación del objetivo principal del estudio es primordial en los proyectos de LA, ya que ello determinará el tratamiento de los datos y las técnicas específicas de exploración. Los proyectos de LA consisten habitualmente en un ciclo iterativo dividido en tres pasos principales: (1) recopilación de datos y pre-procesado; (2) análisis y acción, y (3) post-procesado (Romero y Ventura, 2007; Chatti et al., 2012; Calvet Liñán y Juan Pérez, 2015; Romero y Ventura, 2020). En cada una de estas fases, existen una serie de métodos que suelen requerir ciertas herramientas informáticas para su implementación. Estas herramientas utilizadas por las LA pueden agruparse en dos grandes categorías: herramientas comerciales y herramientas abiertas. En los siguientes subapartados se analizarán en detalle cada uno de los pasos del proceso de LA, junto con los métodos y herramientas más característicos.

FIGURA 15. ETAPAS DEL PROCESO DE LA.



Nota. Adaptado de A Reference Model for Learning Analytics, por M. A. Chatti, A.L. Dyckhoff, U. Schroeder y H. Thüs, 2012, *International Journal of Technology Enhanced Learning*, 4(5/6), 326–327. Copyright 2012 by M.A. Chatti, A.L. Dyckhoff, U. Schroeder and H. Thüs.

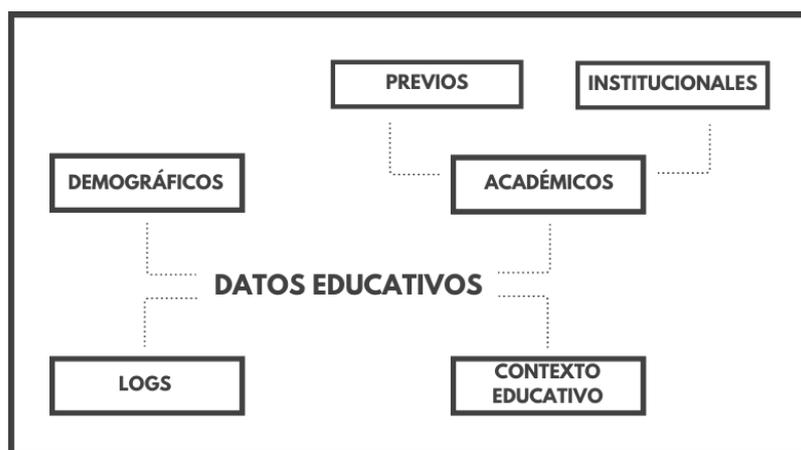
4.5.1. RECOPIACIÓN Y PRE-PROCESADO DE DATOS

El primer paso dentro de un proyecto de LA es recopilar datos de varios entornos y sistemas educativos. Según Chatti et al. (2012), estas fuentes de datos pueden dividirse en dos grandes categorías: sistemas educativos centralizados y entornos de aprendizaje distribuidos. Los LMS, los cuales son la tipología más relevante de sistemas educativos centralizados, acumulan grandes registros de actividades e interacciones de los estudiantes tales como lectura; escritura; acceso y carga de material de aprendizaje, y realización de exámenes. Algunos de los LMS, entre los que se incluyen los LMS comerciales; los desarrollos propios, y los LMS de código abierto, a veces integran herramientas que permiten la generación de informes simples a partir de los datos educativos (Romero y Ventura, 2007). El aprendizaje y la creación de conocimientos pueden presentarse distribuidos a través de múltiples medios y entornos web, por lo que los datos de dicha actividad pueden estar fragmentados en múltiples registros. Estos datos pueden provenir de canales de aprendizaje distintos, tanto formales como informales; presentar diferentes formatos, y estar distribuidos en el espacio y en el tiempo (Suthers y Rosen, 2011; Chatti et al., 2012; Sclater, 2017).

Sclater (2017) divide los tipos de datos en cuatro categorías distintas. La primera son los datos demográficos que la institución tiene sobre los estudiantes, como la fecha de nacimiento; el sexo; el estatus socioeconómico, o la etnia. La segunda categoría son datos académicos que la institución almacena de cada estudiante, por ejemplo, opciones de materias; tareas, y calificaciones. Dentro de esta categoría existen dos subconjuntos de datos: los datos previos al desempeño dentro de la institución, que pueden incluir el informe académico de la etapa de educación secundaria o calificaciones obtenidas en otras instituciones educativas, y el contenido generado por el alumno, que puede consistir en el texto de redacciones; informes de grupo; blogs, o cualquier otro contenido académico generado por los alumnos durante su proceso de aprendizaje. La tercera categoría son los datos de la actividad de aprendizaje que se registran en forma de *logs* por los sistemas informáticos de las instituciones, tales como inicios de sesión; descargas de contenido de aprendizaje, o publicaciones de comentarios en los foros. Esta categoría de datos es clave para las LA y también incluye datos de otros sistemas informáticos

asociados a las instituciones, por ejemplo, registros de préstamos de libros en la biblioteca o de asistencia. La última categoría son los datos del contexto educativo, los cuales incluyen información sobre el plan de estudios; el contenido y las actividades incluidos en el diseño de aprendizaje; las evaluaciones, y eventos como el inicio y la duración del curso, o las fechas de seminarios y exámenes.

FIGURA 16. TIPOS DE DATOS EDUCATIVOS. ELABORACIÓN PROPIA.



Una vez recopilados los datos, la transformación o pre-procesado de estos datos es un elemento esencial de las LA que puede requerir una amplia gama de procesos, ya que los datos pueden provenir de varias fuentes; tener diferentes formatos, o necesitar una transformación a un formato específico que se pueda utilizar como entrada para un método particular de LA. También existen casos en los que es necesario unificar registros y datos que, en sí mismos pueden no ser útiles, pero que cuando se combinan pueden contribuir a la identificación de patrones. Cuando la información proviene de sistemas distintos, pueden surgir problemas en el pre-procesado de estos datos no coincidentes, por ejemplo, al utilizar datos recopilados en diferentes países que sigan distintas convenciones para las fechas (Chatti et al., 2012; Waterman y Bruening, 2014; Sclater, 2017).

McKay (2015) recomienda tener en cuenta tres aspectos respecto a los datos de los proyectos de LA. En primer lugar, los datos pueden ser inválidos, ya que pueden llegar a corromperse mediante procesos de producción y manipulación ejecutados de forma incorrecta, o pueden haberse introducido

erróneamente, por ejemplo, si un usuario rellena información en un campo equivocado. En segundo lugar, los datos pueden estar incompletos, ya que los sistemas de registro de información a menudo contienen campos vacíos. Por último, puede que los datos no reflejen lo que realmente se espera de ellos; por ejemplo, las calificaciones pueden no representar el desempeño real de un alumno en una clase, sino la evaluación específica del desempeño del estudiante por parte de un instructor en particular. Adicionalmente, aunque algunos defensores del *big data* argumentan que, cuando se trabaja en conjuntos de datos masivamente grandes, los posibles fallos en los datos subyacentes no afectan significativamente a los resultados, la experiencia práctica muestra que pueden existir franjas significativas de estos fallos que pasen inadvertidas en las pruebas de calidad programadas para la verificación de los datos, lo cual suele producir una representación insuficiente de una característica o grupo, y una representación excesiva de otro (Waterman y Bruening, 2014).

FIGURA 17. PRINCIPALES RIESGOS DERIVADOS DE LOS DATOS EN PROYECTOS LA. ELABORACIÓN PROPIA.



Por este motivo, y especialmente en el caso de los grandes conjuntos de datos, la fase de pre-procesado incluye la depuración de los registros falsos para que los análisis sean precisos. Un ejemplo sería el de los instructores que inician sesión en su curso con un perfil de estudiante para realizar pruebas y que obtienen un 100% de acierto en todas las preguntas. Estos datos pueden llegar a sesgar el promedio de puntuaciones de la clase, por lo que estos registros deberían filtrarse en la etapa de pre-procesado. Los procesos analíticos que se realizarán posteriormente sobre los datos también pueden requerir la eliminación de valores atípicos que alteren significativamente las métricas y predicciones (Sclater, 2017).

La ingeniería de características es también una parte importante del pre-procesado, la cual permite generar y seleccionar atributos y variables con información sobre los estudiantes. Normalmente, todos los atributos disponibles se pueden reducir y transformar en una tabla resumen para un mejor análisis. Los atributos continuos habitualmente se transforman y convierten en atributos discretos o categóricos para mejorar su comprensión. Finalmente, es importante mantener y proteger la confidencialidad del estudiante anonimizando los datos y borrando toda la información personal, como el nombre, correo electrónico, número de teléfono, etc. (Romero y Ventura, 2020).

Las técnicas empleadas en el pre-procesado de datos suelen ser técnicas procedentes de la EDM, tales como la limpieza de datos; la integración de datos; la transformación de datos; la reducción de datos; el modelado de datos; la identificación de usuarios y sesiones, y el completado de rutas (Liu, 2006; Romero y Ventura, 2007; Chatti et al., 2012; Calvet Liñán y Juan Pérez, 2015). Para implementar estas técnicas, existen una serie de herramientas, tanto comerciales como libres, que suelen emplearse en los proyectos de LA. Algunas de las herramientas comerciales básicas frecuentemente empleadas en el pre-procesado de datos son *Microsoft Excel* y *Google Sheets*. Estas herramientas son muy útiles para presentar los datos claramente dentro de una interfaz completamente visual, lo cual facilita la identificación de problemas estructurales o semánticos en los datos, tales como valores inusuales o ausentes, o entradas duplicadas. Sin embargo, estas herramientas tienen límites en la cantidad de datos que se pueden cargar y operar, por lo que no son útiles para

conjuntos de datos extremadamente grandes (más de 1 millón de filas). Una herramienta libre, comúnmente empleada en el pre-procesado de datos, es *Jupyter notebook*, la cual mantiene un registro de los análisis realizados y los resultados intermedios, mostrando cada acción del usuario y su resultado. En esta fase también suelen emplearse ciertos lenguajes de programación, como Python o *Structured Query Language* (SQL en adelante). Python es capaz de manejar muchos tipos diferentes de formatos, como los archivos *JavaScript Object Notation* (JSON en adelante) generados por los MOOC y las plataformas de aprendizaje en línea. SQL se utiliza para realizar consultas en bases de datos y extraer exactamente los datos deseados, a veces integrándolos en múltiples tablas de bases de datos (Slater et al., 2017).

4.5.2. ANÁLISIS Y ACCIÓN

Una vez realizada la recopilación y pre-procesado de los datos, se pueden aplicar diferentes técnicas de LA para su exploración, con el objetivo de identificar patrones recurrentes. Existen numerosas clasificaciones de las posibles técnicas empleadas en este paso. Una de las clasificaciones más amplias es la de Khalil y Ebner (2016), que dividen las técnicas empleadas en LA en siete grupos distintos:

- 1) La **visualización** de la información, en inglés *Information Visualization* (IV en adelante), consiste en propiciar las formas de presentación visual de la información, de tal forma que se incremente la eficiencia de las capacidades humanas para percibir y procesar los datos, y se facilite la interpretación y el análisis de los datos educativos. Las representaciones visuales y las técnicas de interacción aprovechan la amplitud de la vía que conecta el ojo humano con la mente para permitir a los usuarios ver, explorar y comprender grandes cantidades de información. De este modo, entre las diferentes técnicas de IV que permiten representar la información en un formato claro y comprensible se encuentran las gráficas, los diagramas de dispersión, los diagramas de flujo, las representaciones en 3D o los mapas. En los últimos tiempos, los informes tradicionales basados en tablas de datos están siendo reemplazados por paneles que muestran gráficamente

diferentes indicadores de rendimiento (Kapler y Wright, 2005; Romero y Ventura, 2007; Mazza, 2009; Chatti et al., 2012).

Aunque existen muchas herramientas de visualización de datos, como *Chart.js*¹²; *JavaScript InfoVis Tool kit*¹³; *jpGraph*¹⁴, y *Google Charts*¹⁵, las más extendidas entre los investigadores de LA son *Tableau*¹⁶, una herramienta comercial, y *D3.js*¹⁷, una herramienta de código libre. *Tableau* es una familia de productos para el análisis y la visualización de datos interactivos que, aunque orientada principalmente al campo de la inteligencia empresarial, se emplea comúnmente en entornos educativos. La principal ventaja de *Tableau* es que no se necesitan conocimientos de programación para analizar grandes cantidades de datos de diversas fuentes y en varios formatos. Además, ofrece la posibilidad de crear paneles interactivos, con visualizaciones dinámicas en tiempo real. *D3.js* es una biblioteca de JavaScript para la manipulación de documentos basados en datos, lo cual permite crear complejas visualizaciones interactivas de datos. Algunos de sus inconvenientes son la necesidad de poseer conocimientos de programación; sus limitaciones de rendimiento para grandes conjuntos de datos, y la necesidad de realizar un pre-procesado previo de datos para garantizar la privacidad y la seguridad, ya que *D3.js* no proporciona mecanismos para ocultar datos de usuarios (Siemens, 2013; Slater et al., 2017).

- 2) **Técnicas de minería de datos:** el objetivo de la minería de datos es el descubrimiento de patrones o conocimientos útiles, haciendo legibles y utilizables para el beneficio de estudiantes; profesores, y universidades datos educativos brutos e ininteligibles procedentes de fuentes tales como bases de datos, textos, imágenes o internet (Liu, 2006; Romero et al., 2016). De hecho, según Khalil y Ebner (2016), las técnicas de minería de datos son el método más utilizado para analizar e interpretar los datos de registro de los alumnos. Los métodos

¹² <https://www.chartjs.org/>

¹³ <http://philogb.github.io/jit/>

¹⁴ <https://jpgraph.net/>

¹⁵ <https://developers.google.com/chart>

¹⁶ <https://www.tableau.com/>

¹⁷ <https://d3js.org/>

de minería de datos, a su vez, pueden clasificarse en las siguientes categorías: (a) modelos de predicción, (b) minería de relaciones, (c) descubrimiento de estructuras y (d) descubrimiento con modelos (Baker e Inventado, 2014).

a) En los **modelos de predicción**, el objetivo es desarrollar un modelo que pueda inferir un solo aspecto de los datos denominado variable predicha (similar a las variables dependientes en el análisis estadístico tradicional) a partir de alguna combinación de otros aspectos de los datos denominada variable predictora (similar a las variables independientes en análisis estadístico tradicional). La predicción requiere que, para un conjunto de datos limitado, existan para la variable de salida etiquetas basadas en información confiable y confirmada para casos específicos. Los modelos de predicción se utilizan con frecuencia para predecir cuál será el valor de una variable en contextos en los que no es deseable obtener directamente una etiqueta para esa estructura de datos y también para estudiar qué estructuras específicas de datos influyen en la predicción de otras (Chatti et al., 2012; Baker e Inventado, 2014). Los modelos de predicción más habitualmente empleados son los siguientes:

- o **Clasificación.** La clasificación es el proceso de encontrar un modelo que describa y distinga clases de datos o conceptos. El objetivo de la clasificación es emplear una función, también denominada modelo de clasificación; modelo predictivo, o simplemente clasificador, que sea capaz de predecir la clase de objetos cuya etiqueta de clase se desconoce, siendo esta etiqueta una variable de tipo binario o categórico. Entre los métodos de clasificación más populares en el ámbito educativo se incluyen los árboles de decisión, las redes neuronales, los bosques aleatorios, las reglas de decisión, la regresión por pasos, la regresión logística, el clasificador bayesiano

ingenuo, las máquinas de vectores de soporte, los k-vecinos más próximos (KNN en adelante) y la estimación latente del conocimiento. (Chatti et al., 2012; Baker e Inventado, 2014; Aldowah, 2019).

- o **Regresión.** La clasificación y la regresión también se denominan aprendizaje supervisado, debido a que los objetos de datos utilizados para el aprendizaje (llamados datos de entrenamiento) están etiquetados con clases predefinidas. Sin embargo, mientras que la clasificación predice etiquetas categóricas o discretas, la regresión modela funciones de valor continuo, es decir, se utiliza para predecir valores de datos numéricos ausentes en lugar de etiquetas de clase. Algunas de las metodologías más frecuentemente utilizadas para la predicción numérica son la regresión lineal, los árboles de regresión y, compartido con la clasificación, las redes neuronales y las máquinas de vectores de soporte (Chatti et al., 2012; Baker e Inventado, 2014; Aldowah, 2019).
- b) En la **minería de relaciones**, el objetivo es descubrir relaciones entre las variables de un conjunto de datos con una gran cantidad de variables. Para ello, se puede intentar descubrir qué variables están más fuertemente asociadas con una sola variable de interés particular, o se puede intentar descubrir qué relaciones entre dos variables son más fuertes. Los cuatro tipos de minería de relaciones de uso común en EDM son:
 - o **Minería mediante reglas de asociación.** Su objetivo es el descubrimiento de asociaciones y correlaciones relevantes dentro de los datos de tal forma que, si se encuentra un conjunto de valores en una variable, esto generalmente supondrá que otra variable tenga un valor específico. Algunos de los métodos más populares de reglas de asociación son el algoritmo *a priori*, y el árbol de patrones frecuentes o FP-tree (Chatti et al., 2012; Baker e Inventado, 2014; Aldowah, 2019).

- **Minería de patrones secuenciales.** El objetivo de estos algoritmos es encontrar asociaciones temporales entre eventos (Baker e Inventado, 2014; Aldowah, 2019).
 - **Minería de correlación.** El objetivo de estos algoritmos es encontrar correlaciones lineales positivas o negativas entre variables (Baker e Inventado, 2014).
 - **Minería de datos causales.** El objetivo de estos algoritmos es encontrar si un evento o conjunto observado de datos es la causa de otro. La minería de datos causales se distingue de la predicción en su intento de encontrar, no solo predictores, sino relaciones causales reales mediante la observación de los patrones de covarianza entre las distintas variables del conjunto de datos (Spirtes et al., 2000; Baker e Inventado, 2014).
- c) Los algoritmos de **descubrimiento de estructuras** intentan encontrar una estructura en los datos, sin ninguna idea *a priori* de lo que se debería encontrar. Estos algoritmos contrastan fuertemente con los modelos de predicción anteriores, donde etiquetas de base eran aplicadas a un subconjunto de los datos con antelación. Los algoritmos de descubrimiento de estructuras más usados en educación son:
- **Agrupamiento.** En el agrupamiento o *clustering*, a diferencia de en la clasificación, la etiqueta de clase de cada objeto de entrenamiento no se conoce de antemano. Mediante una función de distancia que define la cercanía espacial de los objetos, el agrupamiento organiza los objetos de datos en grupos o clústeres, de modo que los objetos dentro de un clúster sean similares entre sí y a la vez diferentes de los objetos de otros clústeres. Los principales métodos de agrupación en clústeres se pueden clasificar en las siguientes categorías: métodos de partición, métodos

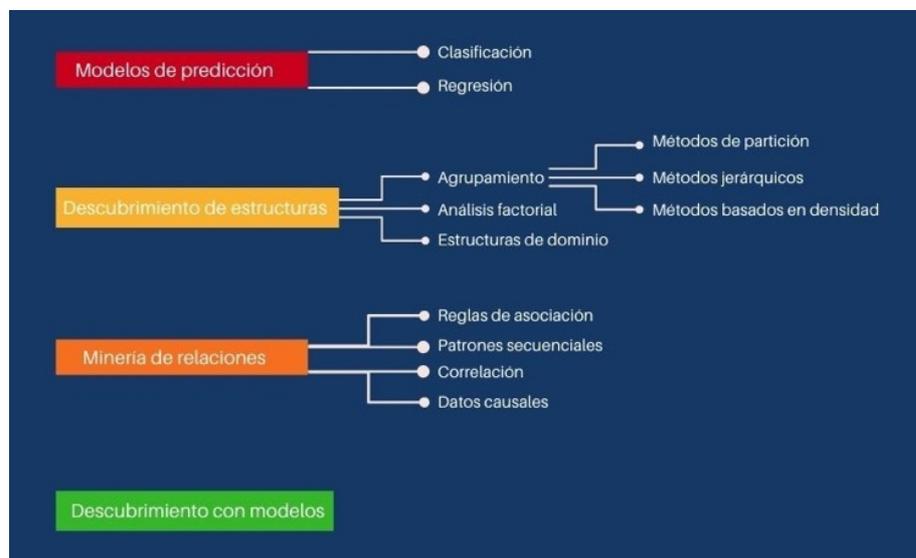
jerárquicos y métodos basados en densidad (Chatti et al., 2012; Baker e Inventado, 2014; Aldowah, 2019).

- ❖ Los métodos de partición comienzan con una división en clústeres aleatoria que el método intenta mejorar moviendo objetos de un grupo a otro mediante una técnica de reubicación iterativa. Algunas de estas técnicas de partición están basadas en métodos heurísticos populares, como el algoritmo k-medias, en el que cada grupo está representado por el valor medio de los objetos en el grupo, o el algoritmo k-medoids, en el que cada grupo está representado por uno de los objetos en el grupo (Chatti et al., 2012).
- ❖ Los métodos jerárquicos crean, a partir del conjunto de objetos de datos, una descomposición jerárquica o árbol de grupos mediante métodos que pueden ser aglomerantes, si la descomposición jerárquica se forma de abajo hacia arriba, o divisivos, si se forma de arriba hacia abajo (Chatti et al., 2012).
- ❖ Los métodos basados en densidad consideran que los clústeres son regiones densas de objetos en el espacio de datos que están separadas por regiones de baja densidad que representan ruido. Por lo tanto, un método basado en densidad continúa incrementando un clúster siempre y cuando la densidad, o número de objetos, en el vecindario exceda un cierto umbral; es decir, para cada objeto dentro de un grupo determinado, la vecindad de un radio determinado debe contener al menos un número mínimo de objetos. Este método se emplea en el filtrado de los valores atípicos (o ruido) y en el descubrimiento arbitrario de grupos de forma. Los

métodos DBSCAN y su extensión OPTICS son ejemplos característicos de métodos basados en densidad (Chatti et al., 2012).

- o **Análisis factorial.** En este tipo de análisis, el objetivo es encontrar variables que se agrupen naturalmente, dividiendo el conjunto de variables en un conjunto de factores latentes o no observables directamente. El análisis factorial incluye algoritmos como el análisis de componentes principales y el análisis de componentes principales de familias exponenciales (Kline, 1993; Baker e Inventado, 2014).
 - o **Descubrimiento de estructuras de dominio.** El objetivo de estos algoritmos es encontrar qué elementos se asignan a habilidades específicas entre los estudiantes (Baker e Inventado, 2014).
- d) En el **descubrimiento con modelos**, primero se desarrolla un modelo mediante uno de los métodos analizados anteriormente (predicción, agrupamiento o usando el razonamiento humano), el cual posteriormente se utiliza como componente de un segundo análisis o modelo, por ejemplo, de predicción o minería de relaciones (Baker e Inventado, 2014).

FIGURA 18. ESQUEMA DE LOS PRINCIPALES MÉTODOS Y ALGORITMOS DE LA MINERÍA DE DATOS. ELABORACIÓN PROPIA



Existen una gran cantidad de herramientas de código abierto que permiten implementar los distintos algoritmos de la EDM. *WEKA*¹⁸ es un paquete de *software* gratuito que reúne una amplia gama de algoritmos de minería de datos, entre ellos de clasificación; agrupamiento; minería de asociación, y creación de modelos. Los algoritmos pueden invocarse desde la línea de comandos, mediante una interfaz gráfica de usuario o invocando algoritmos de una API de Java. *Orange*¹⁹ es un paquete de análisis y visualización de datos que, aunque contiene considerablemente menos algoritmos y herramientas que *WEKA* (o que otras herramientas comerciales que se analizarán posteriormente), y tiene limitaciones respecto a la escala de datos con los que puede trabajar, su interfaz más sencilla y fácil de entender, con *widgets* codificados por colores que diferencian entre entrada de datos y limpieza; visualización; regresión, y agrupamiento, la convierte en una herramienta ideal para proyectos más pequeños o investigadores con experiencia limitada. Entre los algoritmos que ofrece están muchos de uso común, como KNN; bosques aleatorios; clasificador bayesiano ingenuo, y máquinas de vectores de soporte. *KEEL*²⁰ es una herramienta Java de código abierto desarrollada específicamente para la investigación y la educación por la Universidad de Granada, en colaboración con el gobierno de España. *KEEL* es especialmente potente para algunos tipos específicos de algoritmos, como algoritmos de clasificación y regresión; algoritmos evolutivos, o algoritmos de discretización, pero presenta más limitaciones que otras herramientas en cuanto a otros algoritmos de EDM, tales como el agrupamiento; el análisis de factores, o la minería de reglas de asociación. *Spark MLlib*²¹ es un motor de análisis para el procesamiento distribuido de datos a gran escala a través de múltiples procesadores, que permite la integración con varios lenguajes de programación, incluidos Java; Python, y SQL (Siemens, 2013; Slater et al., 2017; Triguero et al., 2017; Romero y Ventura, 2020).

¹⁸ <https://www.cs.waikato.ac.nz/ml/weka/>

¹⁹ <https://orange.biolab.si>

²⁰ <https://sci2s.ugr.es/keel/>

²¹ <http://spark.apache.org/ml/lib/>

Las herramientas comerciales más utilizadas en EDM incluyen *RapidMiner* y *KNIME*. Aunque *RapidMiner*²² es una herramienta comercial, cuenta con un programa educativo que proporciona licencias gratuitas para uso académico tanto a estudiantes como a profesores. *RapidMiner* incorpora todos los algoritmos disponibles en WEKA y posee un extenso conjunto propio de algoritmos de clasificación; regresión; agrupamiento, y minería de reglas de asociación, entre otros, además de una interfaz de programación (API) que permite la integración con programas escritos en Java o Python. *KNIME*²³, anteriormente *Hades*, es un paquete de análisis y limpieza de datos que ofrece muchas de las capacidades de *RapidMiner* y *WEKA*, incluidos todos los algoritmos de *WEKA*. Algunos aspectos especialmente potentes de *KNIME* son su capacidad para integrar datos de múltiples fuentes y su gran número de extensiones, lo que permite interactuar con R, Python, Java y SQL (Slater et al., 2017; Romero y Ventura, 2020).

3) **Estadística y matemáticas.** La estadística es la ciencia de medir, controlar, comunicar y comprender los datos (Davidian y Louis, 2012). En LA es muy frecuente el uso de operaciones de estadística descriptiva, como la media; la mediana, y la desviación estándar, para procesar la información contenida en los datos educativos. De hecho, la mayoría de los sistemas de gestión del aprendizaje existentes implementan herramientas de informes que proporcionan estadísticas básicas de la interacción de los estudiantes con el sistema. Algunos de los ejemplos más frecuentes de estadísticas de uso incluyen el tiempo en línea; el número total de visitas; el número de visitas por página; la distribución de las visitas a lo largo del tiempo; el número de clics sobre los elementos de una página; el tiempo dedicado a una tarea; la frecuencia de las publicaciones y respuestas de los estudiantes, o el porcentaje de material leído. Estas herramientas a menudo generan operaciones estadísticas simples

²² <https://rapidminer.com/products/studio/>

²³ <https://www.knime.com/>

tales como promedio, media y desviación estándar (Chatti et al., 2012; Khalil y Ebner, 2016; Aldowah, 2019; Romero y Ventura, 2020). Entre las herramientas utilizadas para realizar análisis estadísticos están las ya mencionadas Microsoft Excel y Google Sheets, las cuales permiten crear prototipos de nuevas variables en subconjuntos de un conjunto mayor de datos; diseñar nuevas funciones y aplicarlas a toda la hoja, y vincular conjuntos de datos y crear niveles de agregación. Otras herramientas comúnmente empleadas son R²⁴, MATLAB²⁵, SPSS²⁶ y su alternativa libre PSPP²⁷, paquetes estadísticos que ofrecen una gran variedad de pruebas estadísticas; marcos de regresión; correlaciones, y análisis de factores (Siemens, 2013; Slater et al., 2017).

- 4) **Análisis de textos, semántica y análisis lingüístico.** El análisis de texto aplica una variedad de técnicas del procesamiento del lenguaje natural y del análisis del discurso, junto con técnicas lingüísticas; estadísticas, y de extracción de datos, para extraer conceptos y patrones empleados para la categorización y clasificación de documentos textuales. Hay muchos tipos de modelos y estrategias para categorizaciones de texto, pero, en general, todos los algoritmos se pueden dividir en: categorización supervisada, la cual utiliza información previa obtenida a partir de la correcta categorización de documentos que ya han sido verificados; la categorización no supervisada, la cual no utiliza ninguna información previa o externa, por lo que las decisiones se toman internamente en base a lógicas o modelos predefinidos, y la categorización semi-supervisada, que combina enfoques de las categorizaciones supervisada y no supervisada (Chatti et al., 2012; Fiaidhi, 2014; Aldowah, 2019).

Además de la categorización, otros de los objetivos del análisis de texto son la transformación de la información no estructurada en datos que puedan ser tratados mediante técnicas más tradicionales,

²⁴ <https://www.r-project.org/>

²⁵ <https://es.mathworks.com/products/matlab.html>

²⁶ <https://www.ibm.com/es-es/analytics/spss-statistics-software>

²⁷ <https://www.gnu.org/software/pspp/>

y la identificación de significados y relaciones en grandes volúmenes de información. Al combinar el análisis de texto y el semántico, es posible identificar una serie de características dentro de los textos, entre ellas entidades nombradas (objetos de aprendizaje, alumnos, etc.); entidades basadas en patrones (por ejemplo, direcciones de correo electrónico); conceptos o abstracciones de entidades, hechos y relaciones; eventos; atributos concretos y abstractos, y elementos subjetivos en forma de opiniones, sentimientos y emociones. El análisis textual también se emplea para la anotación de información sensible; la verificación de transmisiones en vivo; el filtrado de contenido mediante metadatos; la generación de resúmenes de alto nivel e informes detallados; la recomendación de recursos; el análisis de comentarios y opiniones; la construcción de modelos de temas y la generación de agrupaciones; la medición y validación de resultados, y la identificación de documentos y respuestas similares para detectar posibles plagios (Grimes, 2013; Fiaidhi, 2014).

Existe una ingente cantidad de herramientas comerciales y de código libre, entre las que destacan *LIWC*²⁸, *WMatrix*²⁹ y *Coh-Metrix*³⁰ para el análisis de textos; *MALLET*³¹ y el paquete *stm R*³² para el modelado probabilístico de temas; *CoreNLP*³³, *NLTK*³⁴ y *Apache OpenNLP*³⁵, para el procesado de lenguaje natural; *LightSIDE*³⁶ (basado en *WEKA*), para la minería de textos; *ConceptNet*³⁷, para redes semánticas, y *TAGME*³⁸, para anotaciones semánticas (Rayson, 2008; McNamara et al., 2014; Slater et al., 2017; Roberts et al., 2019; Leedham et al., 2020).

- 5) El **SNA** es el estudio cuantitativo de las relaciones entre individuos u organizaciones a través de herramientas que permiten administrar,

²⁸ <http://liwc.wpengine.com/>

²⁹ <http://ucrel.lancs.ac.uk/wmatrix/>

³⁰ <http://cohmetrix.com/>

³¹ <http://mallet.cs.umass.edu/>

³² <https://cran.r-project.org/web/packages/stm/index.html>

³³ <https://stanfordnlp.github.io/CoreNLP/>

³⁴ <https://www.nltk.org/>

³⁵ <https://opennlp.apache.org/>

³⁶ <http://ankara.lti.cs.cmu.edu/side/>

³⁷ <https://conceptnet.io/>

³⁸ <https://tagme.d4science.org/tagme/>

representar visualmente, analizar y explorar fácilmente las conexiones de las redes sociales (Dawson, 2010). SNA se basa en varios conceptos de la teoría de grafos y la teoría estructural para evaluar propiedades de una red tales como la densidad, la centralidad, la conectividad, la intermediación y sus grados. Estas medidas proporcionan un marco para interpretar y desarrollar una comprensión de los patrones observados en los intercambios que ocurren entre los actores sociales. En SNA, una red social se modela mediante un gráfico $G = (V, E)$, donde V es el conjunto de nodos o vértices que representan a los distintos actores, y E es un conjunto de aristas (también conocidas como arcos, enlaces o vínculos) que representan un cierto tipo de vínculo entre actores. Al cuantificar las estructuras sociales, es posible determinar los nodos más importantes de la red mediante el concepto de centralidad (Wasserman y Faust, 1994; Chatti et al., 2012). A este respecto, la función de las herramientas de SNA es la automatización del proceso de extracción; recopilación; evaluación, y visualización de los datos de la red de los estudiantes y la presentación rápida y eficiente de la información de la red en un formato que los instructores puedan manejar. De este modo, en LA, el SNA se puede utilizar para promover el aprendizaje colaborativo e investigar las conexiones entre alumnos; profesores, y recursos, además de para examinar relaciones de vínculos fuertes o débiles entre los distintos agentes (Elias, 2011; Ferguson, 2012; Khalil y Ebner, 2015; Khalil y Ebner, 2016).

Existen numerosas herramientas diseñadas para el análisis de redes, particularmente herramientas libres. *Gephi*³⁹ es una herramienta ampliamente empleada en LA para el análisis y la visualización de diferentes tipos de redes. *Gephi* posee un conjunto de herramientas gráficas que permiten colorear los nodos y los bordes de las redes en función de sus atributos, propiedades y posición dentro de la red; un API de Java para la manipulación de gráficos de redes y el cálculo de múltiples medidas, y algoritmos de ejecución comúnmente

³⁹ <https://gephi.org/>

utilizados en SNA. *EgoNet*⁴⁰ es una herramienta orientada al análisis de redes egocéntricas (desde la perspectiva de los actores individuales de la red) a través de instrumentos de encuesta. *NodeXL*⁴¹ es una extensión de código libre para *Microsoft Excel* que acepta una amplia variedad de formatos de datos de entrada y proporciona herramientas para filtrar y visualizar los datos, y calcular las propiedades básicas de la red y de los nodos. *Pajek*⁴² es una herramienta de escritorio gratuita para el análisis complejo de una amplia variedad de redes grandes (hasta cientos de miles de nodos), incluido el análisis de redes de interacciones sociales. *Cytoscape*⁴³ es una herramienta multiplataforma desarrollada en Java con numerosas funciones para la visualización y el análisis de varios tipos de redes. *SoNIA*⁴⁴ es otra herramienta Java para el análisis de datos de redes longitudinales, desarrollada por la Universidad de Stanford, que permite la visualización de los cambios de una red a lo largo del tiempo. *Social Networks Visualizer*⁴⁵ es también una herramienta de código abierto para el análisis y visualización de redes sociales que cuenta con opciones para la carga de datos de varios formatos; el cálculo de las propiedades típicas de nodos y gráficos, y la visualización flexible de los datos en red. Una característica interesante es que integra un rastreador web que, comenzando desde una página inicial con enlaces a otras páginas, permite extraer automáticamente la estructura anidada de enlaces de una colección de documentos HTML. *NetworkX*⁴⁶ es una biblioteca de *software* de código abierto para Python, muy utilizada en el mundo académico, que proporciona un amplio conjunto de funcionalidades avanzadas para el análisis de redes complejas de datos, incluidas la reducción de gráficos mediante técnicas de modelado de bloques; agrupación de gráficos; detección de comunidades; predicción de

⁴⁰ <https://sourceforge.net/projects/egonet/>

⁴¹ <https://archive.codeplex.com/?p=nodexl>

⁴² <http://mrvar.fdv.uni-lj.si/pajek/>

⁴³ <https://cytoscape.org/>

⁴⁴ <https://sourceforge.net/projects/sonia/>

⁴⁵ <https://socnetv.org/>

⁴⁶ <https://networkx.org/>

enlaces, o análisis de tríadas de redes. *SNAPP*⁴⁷ es un *bookmarklet* (un programa JavaScript utilizado como un botón en la barra de marcadores del navegador) para el análisis de las redes sociales que los estudiantes crean en algunos de los LMS más comunes, por ejemplo, Blackboard, Desire2Learn y Moodle. *SNAPP* extrae la red social de un estudiante, formada a través de las publicaciones y respuestas de los estudiantes, a partir de las páginas HTML de las discusiones en los LMS. *SNAPP* también permite explorar la evolución de las redes sociales de los estudiantes; analizar usuarios altamente activos o inactivos; identificar agujeros estructurales, y realizar un análisis comparativo de varios foros de discusión. Entre las herramientas comerciales cabe destacar *KNIME*, mencionado anteriormente, que también ofrece una gran cantidad de algoritmos especializados en SNA y análisis de sentimientos, y *NetMiner*⁴⁸, una herramienta gráfica que permite la importación de datos de red en varios formatos; el acceso a visualizaciones; el cálculo de estadísticas comunes basadas en gráficos y nodos, y la ejecución de algoritmos de minería de datos (Bakharia y Dawson, 2011; Siemens, 2013; Slater et al., 2017; de Nooy, 2018; Romero y Ventura, 2020).

- 6) El **análisis cualitativo** trata de procesar datos y tomar decisiones basándose en una descripción que incluya un componente explicativo por parte de los analistas y expertos. Algunos ejemplos de métodos de análisis cualitativo en LA son el uso de entrevistas, encuestas y observaciones. Los resultados de un análisis cualitativo podrían proporcionarse a los alumnos a modo de recomendaciones basadas en sus respuestas y en sus actividades de aprendizaje anteriores (Khalil y Ebner, 2015; Khalil y Ebner, 2016).
- 7) La **gamificación** es el uso de mecánicas, elementos visuales y modos de pensamientos lúdicos para captar la atención de los estudiantes; motivarlos a la acción, y promover el aprendizaje y la resolución de problemas (Kapp, 2012). Khalil y Ebner (2016) consideran la gamificación como una técnica en sí misma y destacan su reciente

⁴⁷ <https://apps-smah.uow.edu.au/SNAPP/getsnappresults.html>

⁴⁸ <http://www.netminer.com/main/main-read.do>

impacto en talleres educativos y en nuevas metodologías que buscan hacer entretenido el aprendizaje. En esencia, la gamificación está basada en la recopilación y procesamiento de datos relevantes, por ejemplo, al asignar puntos de experiencia; subir de nivel; presentar recompensas; ganar y gastar monedas virtuales, o mantener tablas de clasificación. Al examinar cómo interactúan los alumnos en un entorno gamificado durante todo el proceso de aprendizaje es posible generar comentarios que puedan ayudar tanto a los alumnos como a los profesores a maximizar las oportunidades de aprendizaje. La gamificación también puede aportar al proceso de aprendizaje mediciones; informes; métodos analíticos y predictivos útiles, y una mayor cantidad de indicadores que reflejen tanto las interacciones del alumno dentro del juego como sus resultados específicos (Khalil y Ebner, 2016; Sclater, 2017; Dichev et al., 2018).

FIGURA 19. PRINCIPALES GRUPOS DE TÉCNICAS DE ANÁLISIS Y HERRAMIENTAS EN LA. ELABORACIÓN PROPIA.



El segundo paso dentro del proceso de LA no solo incluye el análisis y la visualización de la información, sino también acciones sobre dicha información, ya que la toma de decisiones es el objetivo principal de todo proceso de LA. Los instructores y las autoridades académicas deben utilizar los nuevos conocimientos descubiertos mediante los métodos de LA para realizar intervenciones y tomar decisiones, con el fin de mejorar el rendimiento y el aprendizaje de los estudiantes. Para ello, es importante que los modelos obtenidos por el proceso de LA sean comprensibles, ya que solo de esta forma podrán llegar a ser útiles en la toma de decisiones (Romero y Ventura, 2020).

4.5.3. POST-PROCESADO

La fase de post-procesado es fundamental para la mejora continua del proceso de análisis. Si los resultados no fuesen concluyentes, debido entre otros motivos a deficiencias en la calidad de los datos o en la metodología, sería necesario modificar el diseño del estudio y repetir el análisis. Esto puede obligar a compilar nuevos datos de fuentes adicionales; refinar el conjunto de datos; determinar nuevos atributos requeridos para la nueva iteración; identificar nuevos indicadores o métricas; modificar las variables de análisis, o elegir un nuevo método de análisis. Asimismo, en caso de que las conclusiones sugirieran la aplicación de cambios al proceso de enseñanza-aprendizaje, el análisis debe realizarse de nuevo después de la implementación de las modificaciones (Chatti et al., 2012; Calvet Liñán y Juan Pérez, 2015).

4.6. LIMITACIONES DE LA

Las LA son un campo relativamente reciente, el cual, aunque ha avanzado significativamente en los últimos años, aún debe enfrentarse a una serie de retos y de limitaciones. Khalil y Ebner (2015) dividen estas limitaciones actuales de las LA en ocho dimensiones distintas:

- 1) **Privacidad:** es importante considerar detenidamente los posibles problemas de privacidad que puedan surgir al analizar los datos de los estudiantes, ya que los conjuntos de datos podrían incluir información confidencial y el análisis de los datos podría revelar información personal; actitudes, y actividades de los alumnos. De

manera adicional, en el análisis de los datos relativos a grupos pequeños, la identificación individual de estudiantes podría resultar evidente. Por lo tanto, es posible que se requiera llevar a cabo procesos de desidentificación de los datos para preservar la información de los alumnos, para lo cual existen diversas soluciones criptográficas; técnicas de anonimato, y métodos estadísticos (Fung et al., 2010; Petersen, 2012; Slade y Prinsloo, 2013; Khalil y Ebner, 2015; Sclater, 2017; Leitner et al., 2019)

- 2) **Acceso:** el control de los niveles de acceso para todas las partes involucradas en un proyecto de LA es importante para proteger la privacidad de los estudiantes. Para ello, unos mecanismos de autenticación rigurosos y adecuados garantizan que solo los usuarios legítimos tengan permiso para acceder a datos específicos. De este modo, se debe permitir a los estudiantes actualizar su información y proporcionar información adicional cuando sea necesario. Un docente deberá estar autorizado a acceder a los datos de los estudiantes sin la posibilidad de ver información sensible, como el origen étnico o la nacionalidad. De manera adicional, los tomadores de decisiones pueden necesitar proporcionar datos a los sistemas para cumplir con los objetivos institucionales. También relativo al acceso a los datos, el aumento del alcance de la recopilación de datos distribuidos a través de modelos de recolección alternativos, como la computación portátil y los dispositivos móviles, supone un desafío para los investigadores, principalmente debido a cuestiones de privacidad, y a la dificultad de integrar diversidad de conjuntos y fuentes de datos. Algunas de las posibles soluciones para la integración y el análisis de dichos datos son la construcción de infraestructuras que agreguen datos de múltiples fuentes, o el desarrollo de mecanismos que permitan capturar y evaluar datos distribuidos (Siemens, 2013; Khalil y Ebner, 2015).
- 3) **Transparencia:** los métodos de LA deberían tener como objetivo la transparencia ante el personal y los estudiantes de la institución. Para ello, las instituciones pueden optar por proporcionar acceso a información relativa a la recopilación de datos, y al uso y la

participación de terceros en el análisis de la información de los estudiantes (Khalil y Ebner, 2015). Oblinger (2012) sostiene que es fundamental que las instituciones informen a los estudiantes sobre qué información relativa a ellos se utilizará para qué fines, por quién y para qué beneficio. Esto no significa que los datos deban estar disponibles para el público, sino que puede resultar beneficioso el que sean accesibles a todos los actores (estudiantes, docentes e institución) y todos los sectores (educación, psicología e informática) involucrados en el proceso de LA, de modo que puedan desarrollarse los principios técnicos y éticos correctos en todo el ciclo de vida del proyecto. Una correcta planificación de las políticas de transparencia hará posible que la información apropiada sea compartida con las personas adecuadas (Khalil y Ebner, 2015; Leitner et al., 2019).

- 4) **Políticas institucionales:** con la adopción de LA en los campos educativos, las instituciones deben ajustar sus políticas a los marcos legislativos y definir una serie de regulaciones que contribuyan a establecer una política institucional que refleje de forma adecuada las implicaciones éticas del proyecto. Estas regulaciones pueden incluir limitaciones y directrices respecto a la recopilación de información personal de los alumnos; el uso de la información; la metodología de recogida de datos; los mecanismos de protección de los datos; el periodo de tiempo de conservación de los datos de los alumnos, y los procedimientos de borrado de los datos (Khalil y Ebner, 2015; Leitner, 2019).
- 5) **Seguridad:** todas las herramientas utilizadas en un proyecto de LA deben seguir unos principios de seguridad que aseguren que los resultados del análisis y los registros de los estudiantes estén salvo de amenazas, tanto internas como externas. De este modo, se debe garantizar la confidencialidad, esto es, que no pueda producirse un acceso no autorizado a los datos; la integridad, de tal modo que los datos no puedan alterarse ni reemplazarse, y la disponibilidad de acceso a los datos para todos los actores autorizados dentro del proyecto. El cifrado es un componente clave para proteger la información de los estudiantes, de tal forma que se garantice que solo

las personas autorizadas puedan utilizar los datos. Otros mecanismos que permiten proteger la confidencialidad son la modificación de los permisos de acceso a archivos, la configuración de medidas de seguridad en el entorno operativo o las funciones *hash* criptográficas aplicadas a conjuntos de datos (Chen y Wang, 2008; Khalil y Ebner, 2015; Leitner, 2019).

- 6) **Precisión:** un proyecto de LA debe garantizar que está analizando y seleccionando correctamente los datos; que se ajusta a unos criterios elevados de calidad, y que produce el nivel requerido de precisión, ya que los errores producidos al seleccionar un conjunto de datos incorrecto o ignorar un componente relevante para el análisis de los datos conllevarán un impacto negativo en la precisión del resultado. Por este motivo, es necesario un estrecho seguimiento de las fuentes de datos; las prácticas de recopilación y de integración, y los modelos resultantes, de tal forma que se garantice que los algoritmos analíticos tengan la calidad predictiva adecuada. Adicionalmente, puede ser necesaria una auditoría, especialmente durante la fase de recopilación y pre-procesado de datos, y la fase de análisis, ya que ambas actividades involucran procesos que pueden generar riesgos (Waterman y Bruening, 2014; Khalil y Ebner, 2015). Waterman y Bruening (2014) recomiendan formular preguntas fundamentales sobre los datos y los procesos de análisis del tipo «quién»; «qué»; «cuándo»; «dónde», y «cómo», por ejemplo, ¿cómo se obtuvieron los datos?; ¿qué contienen los datos?; ¿cómo se limpiaron, formatearon e integraron los datos?; ¿qué áreas geográficas están representadas en los datos?, o ¿qué sucedió durante el periodo en que se recopilaron los datos que deba tenerse en cuenta? Por último, en caso de realizarse predicciones a partir de un proceso de LA, es recomendable que la predicción vaya acompañada de una explicación clara de los riesgos en cada etapa; del método matemático de cálculo, y del riesgo agregado estimado (Waterman y Bruening, 2014).
- 7) **Restricciones:** los procesos de LA deben respetar las leyes y normativas relativas a protección de datos y derechos de autor vigentes en el

respectivo país. Tales restricciones legales incluyen limitaciones temporales para la conservación de los datos; la responsabilidad de mantener los datos seguros y a salvo de amenazas internas y externas, y la limitación de la utilización de los datos para fines específicos. Estas restricciones también aplican a los datos obtenidos de redes sociales, ya que estas contienen información personal sensible protegida por las correspondientes leyes de protección de datos (Khalil y Ebner, 2015; Leitner, 2019). A este respecto, existen problemas adicionales atribuibles al hecho de que los sistemas legales nacionales aún no han definido completamente los nuevos desafíos relativos a la privacidad, los derechos de autor, la propiedad intelectual y la propiedad de datos en entornos digitales. Las leyes de privacidad difieren de un país a otro y surgen cuestiones adicionales cuando un estudiante sigue un curso en línea con un proveedor extranjero (Siemens, 2013; Khalil y Ebner, 2015). Algunos autores consideran que en un futuro próximo será necesario una armonización de las legislaciones que rigen este tipo de datos, de manera similar a lo que ha ocurrido con las leyes de propiedad intelectual y derechos de autor en muchos países desarrollados durante las últimas décadas (Siemens, 2013; Slade y Prinsloo, 2013).

- 8) **Propiedad:** por un lado, los datos educativos empleados en los proyectos de LA se originan en los estudiantes, por ejemplo, al introducir datos personales en un formulario o interactuar con los sistemas de información. Por otro lado, el sistema proporciona un entorno estructurado con el que el estudiante interactúa y luego registra esos movimientos como datos. Esto implica que, sin las tecnologías para capturar las aportaciones de los estudiantes, dichos datos no podrían generarse, ya que es la propia institución la que instala; mantiene, y controla los sistemas informáticos que permiten el recopilado y almacenamiento de los datos. En consecuencia, hay dos perspectivas principales sobre quién es el propietario de los datos utilizados en un proyecto de LA: los estudiantes o las instituciones (Siemens, 2013; Jones et al., 2014; Khalil y Ebner, 2015). Algunos autores proponen la fusión de ambas perspectivas, de tal modo que ni los

estudiantes ni las instituciones ostenten en exclusiva la propiedad sobre los datos, sino que ambas partes se beneficien de ellos al colaborar como partes interesadas en el proceso de LA. Una forma sería mediante la implementación de un modelo de copropiedad en el que tanto los estudiantes como las instituciones compartan las oportunidades, obligaciones y responsabilidades de la propiedad de los datos. Este tipo de modelo respaldaría las necesidades de datos de la institución; protegería la privacidad de los estudiantes, y los convertiría en agentes activos con poder para determinar el alcance y el propósito de los datos recopilados, y qué derechos tienen sobre ellos (Slade y Prinsloo, 2013; Jones et al., 2014). Siemens (2013), por su parte, propone tratar los datos educativos como entidades transaccionales, al igual que el dinero, de tal forma que en el futuro puedan producirse escenarios en los que las instituciones animen a los estudiantes a facilitar sus datos a cambio de servicios de apoyo al aprendizaje personalizado.

Por último, con el fin de evitar un impacto personal negativo, Siemens (2013) destaca la necesidad de mantener los procesos humanos y sociales en el centro de las actividades de LA, ya que el proceso de aprendizaje es esencialmente social; creativo; requiere la generación de nuevas ideas, enfoques y conceptos, y por lo tanto no puede reducirse completamente a algoritmos. Los riesgos surgen en la fase de aplicación del proceso analítico principalmente de dos formas. Por un lado, el proceso de LA puede basarse en datos defectuosos, y conducir a resultados incorrectos o engañosos que sean tomados como base para hacer predicciones o tomar decisiones. En el caso de que tales decisiones tengan un efecto negativo sobre los alumnos, se plantean interrogantes sobre la forma de mitigar las posibles consecuencias y sobre la asunción de responsabilidades. Por otro lado, incluso cuando el procesamiento y la información en la que se basa son confiables, y el proceso de análisis produce resultados precisos y confiables, las predicciones y los resultados derivados pueden tener profundas consecuencias, y afectar la capacidad de un alumno para ejercer ciertas elecciones o aprovechar ciertas oportunidades. Por este motivo, las instituciones deben comprometerse a poner todos los medios a su alcance para evitar prejuicios y estereotipos, reconociendo siempre

la naturaleza incompleta; dinámica, y plural de la identidad y las experiencias individuales. De manera adicional, las instituciones deben brindar oportunidades para que los estudiantes puedan actuar fuera de los algoritmos y modelos impuestos; puedan demostrar que los análisis predictivos iniciales son incorrectos o incompletos, o puedan redimirse a pesar de cualquier duda institucional inicial con respecto a su potencial. Otra posible solución sería permitir que los estudiantes opten por no participar en iniciativas institucionales que personalicen el aprendizaje, con la condición de que sean conscientes de las consecuencias de su decisión (Slade y Prinsloo, 2013; Waterman y Bruening, 2014; Sclater, 2017; Leitner, 2019).

Slade y Prinsloo (2013) proponen un modelo de uso LA basado en la comprensión de su alcance; su papel en las instituciones; sus límites éticos, y un conjunto de creencias morales fundadas en los respectivos contextos regulatorios; culturales; geopolíticos, y socioeconómicos. Este modelo está basado en seis principios generales a partir de los cuales las instituciones pueden desarrollar sus propios conjuntos de directrices en función de sus contextos:

1. **LA como práctica moral.** Ya que la educación es principalmente una práctica moral, no causal, las LA no deben orientarse únicamente hacia la eficacia, sino que deben incluir entre sus objetivos la obtención de indicadores relevantes que ayuden a tomar decisiones basadas en aquello que se considera apropiado y necesario desde un punto de vista moral.
2. **Los estudiantes como agentes.** Para que las LA contribuyan al aprendizaje y al desarrollo de los estudiantes, y no solo a la eficiencia de las instituciones, los estudiantes deben colaborar voluntariamente en el suministro y acceso a los datos. Valorar a los estudiantes como agentes capaces de tomar decisiones y colaborar con la institución en la construcción de sus identidades puede ayudar a prevenir la mercantilización de la educación superior.
3. **La identidad y el desempeño de los estudiantes son construcciones dinámicas y temporales.** Ya que una parte integral de las LA se basan en la noción de identidad del estudiante, es fundamental considerar dicha identidad como una vista instantánea de un alumno en un

momento y contexto particulares, y una combinación de atributos permanentes y dinámicos. Por este motivo, se debe permitir que los estudiantes evolucionen, se ajusten y aprendan de las experiencias pasadas sin que esas experiencias se conviertan en lastres permanentes en su historial de desarrollo. Para ello, los datos recopilados deben tener una fecha de vencimiento determinada y mecanismos para que los estudiantes soliciten la eliminación de datos según criterios acordados.

4. **El éxito académico es un fenómeno complejo y multidimensional.** Aunque las LA ofrecen enormes oportunidades para obtener una comprensión más completa del aprendizaje de los estudiantes, siempre hay que recordar que los datos nunca podrán llegar a estar completos, por lo que los análisis son vulnerables a malas interpretaciones y a sesgos.
5. **Transparencia.** Las instituciones educativas deben ser transparentes en cuanto a los fines y las condiciones en los que se utilizarán los datos, sobre quién tendrá acceso a los datos y las medidas a través de las cuales se protegerá la identidad de las personas.
6. **Las instituciones educativas no pueden abstenerse de utilizar datos.** Ya sea que su propósito sea obtener ganancias o mejorar los resultados para los estudiantes, las instituciones tienen una responsabilidad ante los accionistas; los gobiernos, o los propios estudiantes, por lo que deben utilizar los datos disponibles para comprender mejor el aprendizaje de los estudiantes y posteriormente comprometerse a mejorar los resultados.

4.7. LA Y LOS LMOOC

Los MOOC son un tipo de cursos especialmente interesantes para el empleo de LA, ya que su apertura y su masividad permiten la generación de enormes cantidades de datos relacionados con los procesos de enseñanza-aprendizaje. De este modo, un conocimiento más profundo y exacto de las características de la interacción de los participantes con los distintos tipos de contenidos de los cursos permite la implementación y el diseño de modelos de MOOC que se adapten a las preferencias individuales de los estudiantes, y

acrecienten el grado de compromiso y vinculación de los participantes con el curso.

Aunque en los últimos años se ha incrementado el número de estudios que utilizan diversas técnicas de LA para analizar los datos generados por los MOOC, aún existen pocos estudios que apliquen LA en el subtipo específico de los LMOOC. Una de las investigaciones a destacar es el trabajo pionero llevado a cabo por Martín-Monje et al. (2018), el cual recoge los resultados de un estudio que empleó LA sobre el LMOOC *Cómo tener éxito en el examen de nivel B1 de inglés*, enmarcado en el proyecto europeo *Elearning, Communication and Open-data: Massive Mobile, Ubiquitous and Open Learning (ECO)*. Este LMOOC contó en su primera edición con más de 8.000 participantes de los que la plataforma de MOOC registró su actividad online, su acceso a los distintos objetos de aprendizaje y su interacción en los foros. Este conjunto de datos fue procesado utilizando el software Microsoft Excel y el paquete estadístico SPSS, para posteriormente realizar un análisis estadístico inferencial de la interacción en línea de los estudiantes, y de los perfiles y los tipos de objetos de aprendizaje más fuertemente relacionados con la finalización con éxito del curso. Los resultados mostraron que los vídeos cortos son los objetos de aprendizaje más poderosos en este tipo de cursos en línea; que la entrega regular de actividades de calificación automatizada es un indicador sólido de éxito, y que el perfil de estudiante más prominente en los LMOOC es el de los alumnos que acceden a los materiales de aprendizaje, pero no envían tareas ni interactúan activamente en los foros. Lee et al. (2018) también emplearon LA mediante un análisis de estadística descriptiva a partir de los datos recabados del LMOOC *Speak English Professionally: In Person, Online & On the Phone* un LMOOC ofrecido por el Instituto de Tecnología de Georgia a través de la plataforma Coursera y seguido por más de 250.000 alumnos. Los datos recopilados incluían exámenes y calificaciones del curso; publicaciones en foros de discusión, y datos de flujo de clics generados a medida que cada usuario interactuaba con los vídeos y demás contenido de la plataforma. Al igual que en el trabajo de Martín-Monje et al. (2018), los resultados de este estudio apuntan a una correlación entre el visualizado de vídeos y la entrega de tareas, y el éxito en la finalización de curso. Los autores concluyen que las técnicas de LA permiten realizar inferencias sobre el proceso de aprendizaje de los alumnos, predecir su comportamiento y

mejorar el entorno de aprendizaje. Castrillo y Sedano (2021) emplearon una herramienta de LA embebida dentro de la propia plataforma de MOOC y posteriormente utilizaron métodos estadísticos sobre los datos de los 2252 y 1233 participantes, respectivamente, de las dos ediciones del LMOOC PA. Los resultados de su estudio reflejan la importancia del diseño de los LMOOC para el éxito del curso, lo cual en su caso se evidencia por el alto compromiso de los participantes y su alto desempeño en el mismo. Santos et al. (2014) utilizaron métodos de visualización para analizar los datos de dos LMOOC de la plataforma COMA de la UNED, *Fundamentos de alemán para hispanohablantes* e IP, con 23.065 y 33.811 estudiantes respectivamente. Los autores observaron ciertas variables que se correlacionaron con el éxito en el curso, como la participación en los foros y la realización de actividades, y también identificaron que, cuando los estudiantes omitían alrededor del 10% de las actividades propuestas, la probabilidad de aprobar el curso disminuía en un 25%. También basado en la tercera edición del LMOOC de la UNED IP, se encuentra el estudio realizado por del Peral (2019), en el que se utilizaron técnicas de LA sobre los logs que registraron toda la actividad de los estudiantes que participaron en el curso. En este trabajo se emplearon, una vez más, métodos estadísticos para identificar una serie de patrones temporales de uso de la plataforma por parte de los estudiantes, como se analizará con más en detalle en el siguiente capítulo.

CAPÍTULO 5: PERFILES DE PARTICIPANTES DE MOOC

5.1. INTRODUCCIÓN

La gran cantidad de personas que se registran en los MOOC, que puede ir desde los cientos a las decenas de miles de participantes, implica la existencia de una gran diversidad de características entre el alumnado. Estos cursos atraen a estudiantes de diferentes niveles educativos, entre ellos estudiantes no universitarios; estudiantes preuniversitarios; estudiantes universitarios, o estudiantes de posgrado, lo cual implica que dentro de un mismo MOOC podemos encontrar diferentes perfiles de estudiantes, cada uno con un bagaje propio de conocimientos diferente del resto. Adicionalmente, los objetivos de aprendizaje también varían de un alumno a otro, de tal modo que, mientras que algunos de los estudiantes que se inscriben en un MOOC pueden estar interesados en profundizar en un tema, explorando todo el contenido y completando todas las evaluaciones, otros quizás prefieran explorar el curso de una forma más general. La existencia de tal heterogeneidad entre el alumnado de un MOOC tiene como consecuencia que entre los participantes de estos cursos haya una amplia gama de comportamientos de aprendizaje, destrezas y habilidades individuales. De manera adicional, actualmente existe un predominio de los enfoques de aprendizaje basados en plataformas (característicos de la tipología xMOOC), lo cual generalmente implica que la interacción directa entre el instructor y los participantes será reducida. Esto deposita sobre los alumnos una mayor responsabilidad respecto al proceso de aprendizaje, de tal forma que son ellos los encargados de crear y organizar su propio camino de aprendizaje. (McAuley et al., 2010; Littlejohn et al., 2016; Arora et al, 2017).

En este contexto, comprender las diferentes formas en que los participantes de un MOOC se relacionan con los distintos elementos del curso es esencial para que los desarrolladores e instructores de MOOC puedan diseñar cursos que se adapten a las diferentes necesidades, características y objetivos de los estudiantes. Esta comprensión es también importante en la evaluación del fenómeno de los MOOC y de su papel en el servicio de un aprendizaje

permanente a gran escala (Arora et al., 2017; Kahan et al., 2017). Así pues, la tipificación del alumnado de los MOOC permite a los investigadores clasificar a cada individuo como parte de una fracción del grupo completo, sin la necesidad de identificar con precisión el progreso del aprendizaje de cada individuo, una tarea inabordable en un contexto de miles de estudiantes (Sunar et al., 2020).

Como ya se ha analizado en el capítulo anterior, la identificación de perfiles de estudiantes es uno de los objetivos principales de las LA aplicadas al campo de los MOOC, junto con la retroalimentación inteligente; la presentación de recomendaciones; la intervención temprana, o la adaptación y personalización del proceso de aprendizaje. Para Chatti et al. (2012), el objetivo debería ser la creación de un perfil de aprendizaje policontextual que permita la elaboración de una imagen detallada de las actividades del alumno en un contexto de aprendizaje más amplio, además de la concentración de todas las acciones educativas de los estudiantes en los distintos sistemas distribuidos, con el propósito de obtener resultados analíticos más precisos y propuestas más efectivas para los alumnos.

Aunque el fenómeno de los MOOC es relativamente reciente, en los últimos años han surgido numerosos estudios que clasifican a los participantes de MOOC en función de rasgos distintivos comunes. El objetivo de este capítulo es profundizar en estos tipos de clasificaciones y en sus respectivas características. Para ello, se ha dividido en cuatro grandes grupos las taxonomías de participantes de MOOC identificadas en la literatura científica recopilada hasta la fecha. Estos grupos de perfiles, los cuales serán analizados en sendos apartados de este capítulo, son: perfiles basados en la motivación y la intención; perfiles basados en la interacción con los elementos del curso; perfiles basados en interacciones sociales, y perfiles temporales.

5.2. PERFILES DE PARTICIPANTES DE MOOC BASADOS EN LA MOTIVACIÓN Y LA INTENCIÓN

Los motivos por los que un estudiante decide inscribirse en un MOOC son diversos y tienen una relación directa con la probabilidad de completar el curso

con éxito (Luik et al. 2019). De hecho, algunos autores como Reich (2014) consideran que una de las razones por las que las tasas de abandono en los MOOC son tan altas es que no se tiene en cuenta la intención del estudiante a la hora de inscribirse en el curso, ya que muchos estudiantes no tienen la pretensión inicial de completar el MOOC, sino que se inscriben para probarlos o para estudiar únicamente una de sus secciones.

Estrechamente relacionada con la intencionalidad, la motivación es otro de los factores que se han identificado como claves en el éxito de los estudiantes de MOOC. La masividad característica de los MOOC y sus elevadas tasas de abandono hacen que la motivación sea determinante en la culminación por parte de los alumnos del recorrido didáctico de estos cursos. El estudio realizado por Semenova (2020) demuestra que la motivación de los estudiantes tiene una relación significativa con la finalización de un MOOC, de tal modo que los participantes con motivación intrínseca (que toman el curso por interés o curiosidad); los participantes con la intención de adquirir habilidades que podrían ser útiles para obtener un nuevo trabajo, o los participantes con motivación externa tienen más probabilidades de completar un MOOC y obtener un certificado que el resto.

La motivación de los participantes y su conceptualización del propósito del MOOC influyen en su capacidad de autorregulación y su autonomía, y en la percepción del propio proceso de aprendizaje (Littlejohn et al., 2016). Una alta autonomía se correlaciona con una alta probabilidad de alcanzar un nivel moderado de competencia al final del MOOC, ya que una de las ventajas de los entornos de aprendizaje MOOC es la flexibilidad y la oportunidad de elegir libremente cuándo acceder a los materiales de aprendizaje (Durksen et al. 2016). Por otra parte, la ausencia de requisitos para la participación en un MOOC; el que el equipo docente no insista en la participación obligatoria, en la realización de pruebas o en la comunicación con otros alumnos, y el que los participantes no tengan información sobre la evolución del resto de estudiantes son otros de los factores que pueden incrementar la importancia de la motivación como factor de éxito en un MOOC (Semenova, 2020).

Son varios los estudios que han empleado técnicas de LA para analizar y clasificar a los estudiantes de MOOC en función de su grado de motivación a la hora de afrontar el curso. Uno de estos estudios es el de Reich (2014), que fue replicado posteriormente por Ho et al. (2015) con similares resultados. La investigación de Reich utilizó un cuestionario previo para categorizar, en función de sus motivaciones e intenciones, a los participantes de nueve cursos de HarvardX alojados en la plataforma edX. Los MOOC tuvieron un total de 290.606 inscritos, de los cuales 79.525 (el 27%) respondieron a la encuesta. De entre estos, el 42% de los estudiantes con al menos una acción en el curso completaron la encuesta, mientras que el porcentaje entre los estudiantes con una calificación distinta de cero ascendió al 68%. A partir de los resultados de la encuesta, se dividió a los participantes en cuatro categorías:

1. **Ojeadores:** su objetivo era navegar por los materiales, pero no planeaban completar ninguna de las actividades del curso. Fueron el 3% de los participantes.
2. **Indecisos:** no habían decidido si iban a completar alguna de las actividades del curso. Fueron el 15% de los participantes.
3. **Auditores:** planeaban completar algunas de las actividades del curso, pero no obtener un certificado. Fueron el 26% de los participantes.
4. **Finalizadores:** planeaban completar suficientes actividades del curso para obtener un certificado. Fueron el 56% de los participantes.

El estudio muestra que el porcentaje de estudiantes que obtuvieron el certificado al terminar el curso fue más alto entre los estudiantes que completaron la encuesta (16,5% frente al 5,9% general). Respecto a los encuestados, las tasas de certificación entre aquellos que tenían la intención de completar el curso fueron más altas que en los estudiantes con otras intenciones y que el resto de participantes del curso como promedio.

Otra taxonomía relevante basada en la motivación y en los objetivos de los estudiantes de MOOC es la de Xu y Yang (2016) quienes, a partir de los datos de diez cursos alojados en la plataforma edX, construyeron un modelo estadístico de predicción de calificaciones basado en tres categorías:

1. **Prueba del curso:** estos alumnos simplemente comprueban el contenido del curso. Puede que visualicen uno o dos vídeos, pero no envían ninguna tarea.
2. **Visualización de vídeos:** alumnos con una elevada actividad de reproducción de vídeos cuyo objetivo es adquirir los conocimientos del curso, en algunos casos como complemento a sus cursos universitarios. Los alumnos de este grupo puede que seleccionen los vídeos que visualizan, las tareas que entregan y, si es posible, realicen la prueba final.
3. **Obtención de la certificación:** estudiantes que completan el curso y realizan los exámenes finales con el objetivo de obtener una certificación.

Por último, Maya-Jariego et al. (2020) aplicaron técnicas de agrupamiento del tipo de k-medias a las encuestas previas realizadas a 1768 de los participantes de seis MOOC ofertados en 2014. La investigación formó parte del proyecto *MOOCKnowledge*, una iniciativa del Centro Común de Investigación de la Comisión Europea que tiene como objetivo construir una base de datos sobre las motivaciones, intenciones y experiencias de aprendizaje de los cursos en línea ofrecidos por múltiples proveedores. Los MOOC seleccionados para el estudio versaban sobre temas diversos (historia, TIC en educación, análisis de datos, emprendimiento, psicología) y fueron organizados por distintas instituciones: uno de los MOOC pertenecía a una universidad israelí; otro, a un proyecto internacional con financiación europea; tres cursos fueron desarrollados por universidades españolas a través de una gran plataforma de MOOC, y el último fue un curso de una universidad holandesa utilizando una plataforma propia. El estudio identificó tres perfiles de motivación de los participantes en los MOOC:

1. Los participantes con **interés bajo** presentaron niveles medios-bajos de motivación e intención de completar el curso, y mostraron una menor influencia de factores de motivación intrínsecos.
2. Los participantes **autorreferenciales** declararon que su intención de persistir en el curso estaba más motivada por factores internos que por factores externos. Estos participantes atribuyeron consecuencias menos

negativas a la participación en el MOOC, y se sintieron menos influenciados por familiares y amigos.

3. Los participantes con un **compromiso alto** exhibieron niveles elevados de motivación extrínseca, motivación intrínseca e intención de persistencia. Opinaron que completar un MOOC tiene resultados positivos y percibieron en su entorno una valoración positiva a la participación en este tipo de cursos.

5.3. PERFILES DE PARTICIPANTES DE MOOC BASADOS EN LA INTERACCIÓN CON LOS ELEMENTOS DEL CURSO

Desde estudios muy pequeños hasta estudios a gran escala, en la literatura científica se han definido numerosas tipologías de perfiles de usuarios de MOOC que clasifican a los participantes en función del nivel de interacción con los elementos del curso. Sunar et al. (2020) dividen estos elementos en cuatro grandes bloques:

- **Vídeos:** gran parte de los MOOC y, particularmente, los xMOOC se basan en el visionado de vídeos de conferencias o distintos tipos de contenido audiovisual con fines didácticos.
- **Foros de discusión y otras herramientas sociales:** los MOOC suelen contar con un foro de discusión donde el equipo docente atiende las dudas de los participantes y donde estos pueden interactuar entre ellos. Los cMOOC, en especial, suelen emplear otras herramientas sociales como wikis o redes sociales.
- **Entrega de tareas:** la forma de evaluación de los MOOC suele basarse en cuestionarios intermedios y pruebas finales, aunque otras actividades como proyectos o tareas evaluadas por pares también son frecuentes.
- **Progreso a través de la estructura del curso:** los MOOC suelen dividirse en módulos, cada uno con contenidos didácticos basados en vídeos y otros tipos de materiales, dentro de los cuales algunas plataformas presentan de forma gráfica el progreso del estudiante.

FIGURA 20. PRINCIPALES ELEMENTOS DE UN MOOC. ELABORACIÓN PROPIA.



Los primeros estudios sobre las tipologías de participantes de MOOC se basaron en clasificaciones manuales y racionales (Hill, 2013; Ho et al., 2014; Milligan et al., 2013), y gradualmente fueron evolucionando hacia esquemas de clasificación más robustos y sistemáticos basados en técnicas de LA. De manera adicional, las tipologías anteriores a 2016 tendían a basarse en el supuesto de que los alumnos deben entregar todas las tareas y asistir al MOOC hasta el final, lo cual se utilizaba como un indicador clave en la delimitación de los elementos de las respectivas taxonomías. Paulatinamente, los estudios y las correspondientes tipologías fueron integrando un nivel más amplio de actividades en las variables de clasificación, tales como intentos en los cuestionarios; actividad en foros de discusión, o acceso a recursos, e incluyendo perfiles de participantes con intereses particulares más específicos (Li y Baker, 2018; Poellhuber et al., 2019).

Los esquemas de clasificación de participantes de MOOC basados en la interacción con los elementos del curso pueden dividirse en tres tipos: clasificaciones cuantitativas son las que dividen a los participantes a través de

mediciones del nivel de compromiso e interacción con el curso; clasificaciones cualitativas son aquellas que describen el tipo de interacción o la preferencia por ciertos elementos del MOOC, y las clasificaciones mixtas combinan aspectos cuantitativos y cualitativos en la definición de las taxonomías (Poellhuber et al., 2019). Estos tres tipos de esquemas de clasificación se refieren, por lo tanto, a la forma en la que los distintos estudios determinan los participantes que pertenecen a cada categoría, no a las técnicas empleadas para definir cada uno de los grupos que conforman la categoría.

5.3.1 CLASIFICACIONES CUALITATIVAS

Una de las primeras clasificaciones de participantes de MOOC, que sirvió como base para tipologías posteriores, fue la de Milligan et al. (2013), quienes entrevistaron a 29 de los 2.300 participantes del MOOC *Change11*, un cMOOC de 35 semanas sobre tecnología y enseñanza organizado por George Siemens; Stephen Downes, y Dave Cormier, desde septiembre de 2011 a mayo de 2012. A partir de los resultados de las entrevistas, los autores establecieron tres perfiles principales de participantes de cMOOC:

1. **Activos:** aquellos que siguieron el curso y lo complementaron con la creación de blogs y cuentas en Twitter, en las que discutían activa y regularmente el curso, y mediante las cuales contactaban con otros participantes. Los participantes activos mostraron altos niveles de motivación, lo cual les permitió avanzar en el curso y superar ciertos desafíos que supusieron una barrera para otros estudiantes.
2. **Merodeadores:** son participantes que siguieron activamente el curso, pero no se involucraron activamente con otros estudiantes. Esta es la categoría más compleja, ya que se identifican tres subcategorías en función de los distintos tipos de comportamiento: (1) aquellos que no se relacionaban con ninguna red, porque no estaban interesados en involucrarse con otros para aprender; (2) aquellos que se involucraron con redes externas al curso, pero no con las redes internas, y (3) los que participaron silenciosamente en redes internas, pero, debido a falta de confianza, no contribuyeron al curso de ninguna manera.

3. **Pasivos:** son participantes que, aunque se sintieron insatisfechos y frustrados con el MOOC, continuaron hasta el final.

Una de las clasificaciones de participantes de MOOC más influyentes y citadas en la literatura científica es la de Hill (2013), la cual, a partir de datos obtenidos de los 46.000 participantes de un xMOOC de la plataforma Coursera, amplía y matiza la taxonomía inicial de Milligan et al. (2013), llegando a identificar cinco perfiles de usuarios:

1. **Ausentes:** son aquellos participantes que se registran en el curso, pero que no llegan a acceder a él.
2. **Observadores:** son aquellos que inician sesión y acceden al contenido del curso, pero no realizan ningún tipo de evaluación más allá de los cuestionarios emergentes incrustados en los vídeos.
3. **Drop-ins:** son aquellos estudiantes que realizan alguna actividad relacionada con un tema específico del curso, por ejemplo, ver vídeos o leer los foros, pero no intentan completar el curso completo. Algunos de estos estudiantes utilizan los MOOC de manera informal para encontrar el contenido que les ayude a cumplir los objetivos del curso en otros lugares.
4. **Pasivos:** son estudiantes que conciben el curso como contenido para ser consumido, de tal modo que visualizan los vídeos; responden a los cuestionarios, y leen los foros de discusión, pero generalmente no se involucran en las tareas.
5. **Activos:** son los estudiantes que tienen la intención de participar en el MOOC plenamente y por ello contribuyen en foros de discusión; entregan la mayoría de las tareas, y responden a todos los cuestionarios y evaluaciones.

Koller et al. (2013) emplearon técnicas estadísticas y de visualización para analizar los datos generados a partir de 86 MOOC de la plataforma Coursera. Los autores dividieron a los participantes en dos categorías principales: (1) navegadores, aquellos estudiantes que bien se registran en un MOOC, pero nunca llegan a acceder a él, o bien navegan durante una semana o dos antes de abandonar el curso, y (2) estudiantes comprometidos, los cuales a su vez los

autores dividen en tres subcategorías no delimitadas de forma precisa, por lo que pueden llegar a solaparse:

1. **Participantes pasivos:** aquellos que interactúan con el MOOC predominantemente visualizando los vídeos de conferencias. Estos estudiantes tienen una participación limitada en los foros del curso y, por lo general, intentan completar pocas tareas y cuestionarios.
2. **Contribuidores:** aquellos estudiantes cuyo medio específico de interacción es a través de la generación de nuevo contenido, por ejemplo, participando en discusiones en foros o contribuyendo con subtítulos en idiomas extranjeros.
3. **Participantes activos:** aquellos que se involucran con el contenido del curso y completan tareas; cuestionarios; exámenes, y actividades de evaluación por pares.

A partir de los datos generados en seis xMOOC ofrecidos en Coursera (tres ediciones de *Machine Learning* con 64.536, 60.092 y 112.897 participantes respectivamente, y tres ediciones de *Probabilistic Graphical Models* con 30.385, 34.693 y 25.930 participantes respectivamente), Anderson et al. (2014) identificaron, mediante métodos estadísticos y de visualización, una nueva taxonomía basada en la tendencia por parte de los participantes a realizar ciertas actividades, en particular el visionado de clases y la entrega de tareas. Esta taxonomía está compuesta de cinco perfiles:

1. Los **espectadores** apenas realizan actividad alguna en el curso.
2. Los **visualizadores** visualizan las clases, pero apenas entregan tareas.
3. Los **coleccionistas** descargan los vídeos de las clases, pero apenas entregan tareas. A diferencia de los espectadores, el que descarguen los vídeos no implica que finalmente los lleguen a ver.
4. Los **solucionadores** entregan tareas y apenas visualizan las clases.
5. Los **versátiles** equilibran el visionado de las clases y las tareas.

Kovanovic et al. (2016) analizaron los datos de 26.025 estudiantes que repitieron inscripción en alguna de las 28 ediciones de 11 MOOC diferentes ofrecidos por la Universidad de Edimburgo en la plataforma Coursera y, tras

aplicar técnicas de agrupamiento mediante el algoritmo k-medias, identificaron cinco perfiles de participantes:

1. **Solo inscrito:** alumnos que se inscribieron en los cursos, pero no realizaron ningún tipo de actividad.
2. **Bajo compromiso:** alumnos con poca actividad en los cursos.
3. **Vídeos:** alumnos que se centraron en el visionado de los vídeos.
4. **Vídeos y cuestionarios:** alumnos que vieron los vídeos, y completaron las pruebas y los cuestionarios.
5. **Sociales:** alumnos que se centraron en las discusiones en línea y no vieron la totalidad de los vídeos.

La tabla 5 muestra un resumen de las distintas clasificaciones cualitativas. Aunque las categorías definidas por cada tipología no son exactamente coincidentes, en la tabla se intentan representar las categorías que guardan mayores similitudes mediante códigos de colores y mediante correspondencia en el nivel de la fila.

TABLA 5. CLASIFICACIONES CUALITATIVAS BASADAS EN ELEMENTOS DE INTERACCIÓN. ELABORACIÓN PROPIA.

	Milligan et al. (2013)	Hill (2013)	Koller et al. (2013)	Anderson et al. (2014)	Kovanovic et al. (2016)
	Clasificación		Ausentes	Solo inscripción	
Pasivos		Observadores	Navegadores	Espectadores	Bajo compromiso
Merodeadores		Drop-ins	Pasivos	Coleccionistas	Vídeos
		Pasivos		Visualizadores	
Activos		Activos	Contribuidores	Solucionadores	Sociales
	Activos	Activos	Versátiles	Vídeos y cuestionarios	
MOOC	cMOOC	1 xMOOC Coursera	86 xMOOC Coursera	6 xMOOC Coursera	28 ediciones de 11 xMOOC Coursera (Universidad Edimburgo)
Participantes	29	46.000 aprox.	Sin especificar	328.533	26.025
Elementos	Conferencias en línea, <i>newsletter</i> y redes sociales	Inicios sesión, vídeos, foros, tareas y cuestionarios	Vídeos, tareas, cuestionarios, pruebas, actividades P2P y foros	Tareas, visualización y descarga de vídeos	Vídeos, tareas, cuestionarios, wikis y foros
Técnica	Encuesta y entrevista	Categorías predefinidas	Estadística y visualización	Estadística y visualización	Agrupamiento (k-medias)

5.3.2 CLASIFICACIONES CUANTITATIVAS

Ho et al. (2014) utilizaron métricas de la actividad de 841.687 participantes basadas en los datos extraídos de 17 MOOC pertenecientes al curso 2012-2013, y organizados en la plataforma edX por las universidades de Harvard y MIT para identificar cuatro perfiles de estudiantes:

1. **Solo inscripción:** participantes que se inscriben al curso, pero que nunca acceden a la plataforma.
2. **Solo visualización:** participantes que acceden al material didáctico y a menos de la mitad de los capítulos disponibles, y que no llegan a obtener un certificado.
3. **Solo exploración:** participantes que acceden a más de la mitad de los capítulos disponibles en el material didáctico, pero no llegan a obtener un certificado.
4. **Certificado:** Personas inscritas que obtienen un certificado del curso.

Cisel et al. (2015) clasificaron a los 57.963 participantes de tres ediciones de un MOOC de cinco semanas sobre emprendimiento y de cuatro ediciones de un MOOC de cuatro semanas sobre gestión de proyectos en función de su nivel de participación, para posteriormente aplicar un tratamiento estadístico y de regresión logística a los datos obtenidos. Ambos cursos fueron desarrollados mediante una plataforma basada en el LMS de código abierto *Canvas de Instructure*⁴⁹. En la clasificación, a los estudiantes que obtuvieron un certificado al final del MOOC se les denominó **finalizadores**; a los que presentaron al menos un cuestionario o tarea, pero no completaron el curso se les denominó **desconectados**, y a los que no enviaron ningún cuestionario ni tarea se les etiquetó como **auditores** si habían visto al menos un 10% de los vídeos del curso y **espectadores** si el porcentaje era menor.

Ferguson y Clow (2015) aplicaron el algoritmo k-medias a cuatro MOOC organizados por la *Open University* y alojados en la plataforma *FutureLearn*, con

⁴⁹ <https://www.instructure.com>

un total agregado de 34.203 participantes. Los autores identificaron siete grupos distintos de participantes:

1. **Samplers:** son usuarios que visualizaron algunos vídeos.
2. **Inicio fuerte:** estos alumnos completaron la primera evaluación del curso, pero luego su compromiso se redujo drásticamente, con muy poca actividad posterior.
3. **Recurrentes:** estos participantes completaron la evaluación en la primera y segunda semanas, pero luego abandonaron el curso.
4. **Desertores a mitad de camino:** estos estudiantes completaron tres o cuatro evaluaciones, pero abandonaron cerca de la mitad del curso.
5. **Cerca de acabar:** estos usuarios completaron las evaluaciones de forma regular, pero abandonaron justo antes de finalizar el curso.
6. **Finalizadores tardíos:** este grupo incluye a los alumnos que completaron la evaluación final y la mayoría del resto de evaluaciones, pero bien las enviaron tarde o les faltó por entregar alguna.
7. **Finalizadores aplicados:** este grupo está formado por alumnos que completaron todas las tareas del curso y participaron activamente durante todo el proceso.

Tseng et al. (2016) también emplearon técnicas de agrupamiento para analizar tres MOOC de la universidad taiwanesa Yuan Ze, *C# Programming*; *Internationalization Strategy*, y *Computer-aided Design and Manufacture*, con un total conjunto de 1489 participantes. A partir de los datos obtenidos, los autores identificaron tres perfiles distintos:

1. **Estudiantes activos:** aquellos con el mayor grado de actividad en el visionado de vídeos y en la entrega de tareas.
2. **Estudiantes pasivos:** aquellos con un reducido grado de actividad (una tercera parte de visualización de vídeos y la mitad de entrega de tareas que los estudiantes activos).
3. **Espectadores:** participantes con niveles mínimos de actividad y de progreso en el curso.

Rodrigues et al. (2016) también aplicaron técnicas de agrupamiento mediante el algoritmo k-medias a los datos de un LMOOC conectivista de

gramática portuguesa de la plataforma Openredu⁵⁰ que se ofertó en 2013 y contó con 5.100 participantes. A través del estudio, los autores identificaron tres perfiles de participantes:

1. **Comprometidos:** este grupo demostró altos niveles de interacción a través de foros de discusión, continuidad en la realización de actividades y baja variabilidad en las notas medias.
2. **Esporádicos:** este grupo mostró una gran variabilidad en el compromiso con el curso, con periodos significativos de ausencia.
3. **Desconectados:** fue el grupo con menor rendimiento y participación.

Arora et al. (2017) aplicaron técnicas de agrupamiento mediante k-medias a los datos de los 641.138 alumnos registrados durante el año académico 2013-2014 en cualquiera de los 16 MOOC de MITx o HarvardX ofrecidos en la plataforma edX durante aquel curso. El resultado fue la identificación de cinco grupos distintos de participantes:

1. **Desinteresados:** fueron los participantes menos activos, ya que apenas accedieron al contenido y obtuvieron un rendimiento insignificante.
2. **Ocasionales:** fueron participantes poco activos, con un acceso al contenido limitado y un rendimiento bajo.
3. **Productivos:** son participantes que mostraron un gran rendimiento, con niveles moderados de esfuerzo e interacción.
4. **Exploradores:** son participantes cuya interacción fue moderada y cuyo rendimiento fue bastante bajo, pero que se mantuvieron comprometidos con el curso incluso después de su conclusión.
5. **Triunfadores:** son participantes muy comprometidos e interactivos, ya que exploraron la mayor cantidad de material del curso y obtuvieron tasas altas de desempeño.

La tabla 6 muestra un resumen de las clasificaciones cuantitativas analizadas. De nuevo, se intentan representar las categorías con mayor similitud mediante códigos de colores y correspondencia en el nivel de fila.

⁵⁰ <http://openredu.org>

TABLA 6. CLASIFICACIONES CUANTITATIVAS BASADAS EN ELEMENTOS DE INTERACCIÓN. ELABORACIÓN PROPIA.

	Ho et al. (2014)	Cisel et al. (2015)	Ferguson y Clow (2015)	Tseng et al. (2016)	Rodrigues et al. (2016)	Arora et al. (2017)
Clasificación	Solo inscripción					
	Solo visualización	Espectadores	Samplers	Espectadores	Desconectados	Desinteresados
			Inicio Fuerte			
	Solo exploración	Desconectados	Audidores	Pasivos	Esporádicos	Ocasionales
			Recurrentes			
	Certificado	Finalizadores	Desertores mitad de camino	Activos	Comprometidos	Triunfadores
Cerca de acabar						
		Finalizadores tardíos				
MOOC	17 xMOOC HarvardX y MITx	3 ediciones XMOOC y 4 ediciones xMOOC Canvas/Instructure	4 xMOOC FutureLearn (Open University)	3 xMOOC Yuan Ze Taiwan	cLMOOC Openredu	16 xMOOC edX (MITx y HarvardX)
Participantes	841.687	57.963	34.203	1.489	5.100	641.138
Elementos	Registros, certificados, recursos	Videos, cuestionarios, tareas, certificado	Contenido y evaluaciones	Login, videos, tareas	Foros, actividades, calificaciones, ausencias	Videos, interacciones en el curso, capítulos, foros, calificaciones
Técnica	Categorías predefinidas	Regresión logística	Agrupamiento (k-medias)	Agrupamiento	Agrupamiento (k-medias)	Agrupamiento (k-medias)

5.3.3 CLASIFICACIONES MIXTAS

Kizilcec et al. (2013) aplicaron técnicas de agrupamiento mediante el algoritmo k-medias a tres MOOC distintos organizados por la Universidad de Stanford a través de la plataforma edX: *Computer Science 101*, un MOOC con 46.096 participantes y contenido de nivel de secundaria; *Algorithms: Design and Analysis*, con 26.887 participantes y contenido de nivel universitario, y *Probabilistic Graphical Models*, un curso con 21.108 participantes de nivel de posgrado. Los autores identificaron cuatro perfiles distintos en función del nivel de compromiso de los estudiantes con el MOOC y su grado de participación:

1. **Finalizadores:** estudiantes que, independientemente del resultado, completaron la mayoría de las tareas ofrecidas en el curso.
2. **Asistentes:** estudiantes que apenas realizaron tareas, pero que siguieron el curso durante la mayor parte de su duración, visualizando los vídeos.
3. **Desconectados:** estudiantes que realizaron las tareas al comienzo del curso, pero que en algún momento (generalmente en el primer tercio del curso) se desvincularon del MOOC.
4. **Sampling:** estudiantes que ven vídeos, normalmente solo uno, durante uno o dos periodos de evaluación.

Alario-Hoyos et al. (2014) se basaron en las taxonomías cualitativas de Milligan et al. (2013) y de Hill (2013) para su análisis de los perfiles de participantes del curso *Digital Education of the Future*, un cMOOC sobre tecnologías educativas desplegado en la plataforma MiríadaX que contó con 5595 inscritos. El MOOC se desarrolló entre el 5 de febrero y el 25 de abril de 2013, y tuvo una duración de 9 semanas, dedicándose tres semanas a cada uno de los tres módulos y optándose por un tipo de evaluación continua. Los autores dividen a los participantes del estudio en tres grupos generales: merodeadores, aquellos alumnos que se registran pero que consultan un número reducido de recursos; estudiantes que participan en el curso, pero no llegan a completarlo, y participantes que completan el curso de principio a fin. A partir de estos tres bloques, los autores identifican siete perfiles de participantes:

1. Los **ausentes** son un tipo de merodeadores que se matriculan en el curso, pero no realizan ninguna actividad (ni ven vídeos ni contestan a pruebas).
2. Los **observadores** son otro tipo de merodeadores quienes, aunque se registran en el curso y ven algunos vídeos, no intentan realizar ninguna actividad de evaluación.
3. Los **drop-ins** pertenecen a la categoría de participantes que no completan el curso. Estos estudiantes completaron la primera actividad de evaluación sumativa, pero no intentaron el cuestionario final del curso.
4. Los **latecomers** también pertenecen a la categoría de participantes que no completan el curso, en este caso por haberse unido al MOOC después de su comienzo y haberse perdido la primera actividad de evaluación sumativa.
5. Los **drop-in latecomers** son una categoría híbrida de las dos anteriores, ya que se unieron al curso una vez iniciado y por lo tanto se perdieron la primera actividad de evaluación sumativa; realizaron al menos una de las actividades intermedias; pero abandonaron el MOOC antes de realizar el cuestionario final.
6. Los **participantes no comprometidos** son aquellos que completan el curso sin realizar actividades que requieran una carga de trabajo importante. De este modo, los estudiantes de este grupo realizaron la primera y última prueba de opción múltiple; algunos cuestionarios sumativos intermedios; pero no participaron en ninguna de las tres actividades de corrección por pares.
7. Los **participantes comprometidos** son aquellos que siguieron el curso de principio a fin y completaron actividades de todas las tipologías, tanto cuestionarios acumulativos como actividades de corrección por pares.

Coffrin et al. (2014) aplicaron análisis de regresión a los datos generados por dos MOOC desarrollados en la Universidad de Melbourne y publicados en la plataforma Coursera: *Principles of Macroeconomics* y *Discrete Optimization*, con 32.598 y 22.731 participantes respectivamente, e identificaron tres perfiles de participantes:

1. **Auditores:** estudiantes que visualizaron los vídeos, pero no participaron en ninguna de las actividades de evaluación.
2. **Activos:** estudiantes que participaron en las actividades de evaluación.
3. **Cualificados:** estudiantes que visualizaron los vídeos y superaron una cierta nota en las actividades de evaluación.

Sharma et al. (2015) predefinieron una estructura jerárquica de perfiles de participantes a la que posteriormente aplicaron un análisis estadístico utilizando los datos de cuatro MOOC de programación de la plataforma Coursera. De este modo identificaron dos supergrupos divididos en tres y cinco subgrupos, respectivamente:

- **Activos:** son los participantes que entregan las tareas evaluables obligatorias para superar el curso. Se dividen en:
 - **Excepcionales:** obtienen mejores puntuaciones en las primeras entregas de las actividades, las cuales entregan con mayor anticipación, y visitan los foros con mayor frecuencia.
 - **Normales:** obtienen mejores puntuaciones que los suspensos, ya que repiten las entregas de las actividades con mayor frecuencia con el objetivo de mejorar.
 - **Suspensos:** muestran los menores niveles de acceso a vídeos y el mayor retraso en la entrega de las tareas.
- **Visualizadores:** son los participantes que no entregan las tareas evaluables obligatorias para superar el curso. Se dividen en:
 - **Activos:** son participantes que completan las actividades de evaluación no obligatorias.
 - **Pasivos:** son participantes que no completan las actividades de evaluación. Presentan un periodo de actividad inferior al de los usuarios activos.
 - **Usuarios wiki:** presentan un periodo de actividad mínimo, ya que utilizan el curso como herramienta puntual de consulta.
 - **Desertores:** son participantes que en algún momento abandonan el curso.
 - **Finalizadores:** son participantes que visualizan la mayor parte de los vídeos, pero no entregan las tareas obligatorias.

Chen et al. (2015) también aplicaron técnicas de agrupamiento a los datos generados por los 10.475 participantes (de 320.092 registrados) que interactuaron con seis MOOC de la plataforma Coursera, e identificaron cinco perfiles distintos:

1. **Versátiles:** quienes participaron activamente en todas las secciones (foros, vídeos, wikis y cuestionarios).
2. **Menos versátiles:** quienes participaron en todas las secciones con menor frecuencia que los versátiles.
3. **Each-bit:** quienes mantuvieron algún nivel de actividad, aunque fuese mínimo, en todas las secciones del MOOC.
4. **Realizadores de pruebas:** quienes se centraron en responder a los cuestionarios.
5. **Ocasionales:** quienes accedieron a los vídeos, los foros y la wiki, pero no completaron cuestionarios.

Coleman et al. (2015) aplicaron un modelo probabilístico no supervisado denominado *Latent Dirichlet Allocation* (LDA en adelante) a los datos obtenidos a partir del MOOC 8.02x: *Electricidad y magnetismo*, un curso de MITx ofrecido en la plataforma edX en la primavera de 2013 en el que se matricularon 43.758 personas. Los autores identificaron tres perfiles de participantes:

1. **Shopping:** son participantes que solo visualizaron los vídeos del curso antes de abandonarlo.
2. **Desconectados:** son estudiantes que participaron activamente en el curso, pero que se desvincularon a mitad de camino.
3. **Finalizadores:** participantes que mostraron un compromiso significativo con la mayoría del material del curso, involucrándose con el material tan pronto como era publicado.

Kahan et al. (2017) aplicaron técnicas de agrupamiento a los datos obtenidos a partir de un MOOC de biología de la plataforma Coursera en el que participaron más de 21.899 estudiantes. Para el trazado de los perfiles, los autores definieron nueve variables relativas a la visualización de vídeos; la participación en los foros, y la realización de pruebas de evaluación. Las categorías de participantes identificadas fueron las siguientes:

1. **Degustadores:** fueron los usuarios con menor puntuación para cada una de las variables, lo que evidencia una baja actividad en el curso.
2. **Downloaders:** presentaron una baja actividad en el curso, pero descargaron los vídeos.
3. **Desconectados:** estos participantes visualizaron aproximadamente el 70% de los vídeos y respondieron al 65% de las preguntas asociadas; completaron alrededor del 40% de los cuestionarios; accedieron a los foros de discusión un promedio de 5 veces, y ninguno presentó el examen final.
4. **Comprometidos fuera de línea:** son participantes con altos niveles de participación en el curso, pero que mostraron una tendencia a descargar los vídeos en lugar de verlos en línea.
5. **Comprometidos en línea:** son participantes que también puntuaron alto en todas las variables del curso y tendieron a visualizar los vídeos en línea.
6. **Comprometidos sociales moderados:** estos participantes también demostraron altos niveles de participación en el curso y, además, su actividad en los foros de discusión fue superior a la de los grupos anteriores.
7. **Comprometidos sociales:** al igual que los tres grupos anteriores, estos participantes participaron activamente en el curso y fueron especialmente activos en los foros de discusión.

Khalil y Ebner (2017) aplicaron LA al MOOC de 10 semanas *Social Aspects of Information Technology*, un curso perteneciente a la plataforma de MOOC austriaca iMooX⁵¹. Aunque este curso era obligatorio para los 459 estudiantes universitarios que se matricularon, también se abrió al público general, de tal modo que se inscribieron 379 alumnos adicionales. Los autores emplearon técnicas de agrupamiento mediante el algoritmo k-medias utilizando cuatro variables: (1) frecuencia de lectura o la cantidad de veces que un usuario hizo clic en publicaciones en el foro; (2) frecuencia de escritura o el número de publicaciones escritas en los foros; (3) número de vídeos visualizados, y (4) la

⁵¹ <http://www.imoox.at>

suma de los intentos realizados en las pruebas de evaluación. A partir de los resultados, los autores identificaron cuatro perfiles:

1. **Desertores:** un grupo con tasas de actividad bajas para todas las variables analizadas.
2. **Estudiantes perfectos:** un grupo con tasas elevadas de frecuencia de lectura y visualización de vídeos.
3. **Sorteando el sistema:** a este grupo pertenecen usuarios con elevados índices de frecuencia de escritura y lectura; tasas de intentos de evaluación excepcionalmente altas; pero un promedio notablemente bajo de visualización de vídeos.
4. **Sociales:** son los únicos participantes que escribieron en los foros.

Li y Baker (2018) analizaron los datos de tres MOOC de la plataforma Coursera, *Álgebra*, *Cálculo I* y *Cálculo II*, mediante análisis estadístico de regresión y de robustez, e identificaron cuatro perfiles de participantes:

1. **Versátiles:** visualizaron e interactuaron en mayor proporción con las conferencias haciendo pausas; buscando hacia atrás y ralentizándolas; además de entregar un mayor porcentaje de cuestionarios, y obtener la calificación de curso más alta.
2. **Realizadores de pruebas:** estos participantes interactuaron más con los cuestionarios que con las conferencias en vídeo, a pesar de lo cual obtuvieron calificaciones de curso mucho más altas que el resto de los perfiles.
3. **Auditores:** estos alumnos mostraron un alto compromiso con las conferencias y un bajo compromiso con los cuestionarios, y en consecuencia obtuvieron calificaciones bajas en el curso.
4. **Desconectados:** aquellos con menor actividad y más bajo rendimiento de todo el curso.

Poellhuber et al. (2019) también emplearon técnicas clásicas de categorización estadística, como el agrupamiento; el análisis de correspondencia múltiple, y el análisis de componentes principales, con los datos obtenidos de los 4.850 participantes de un MOOC en lengua francesa titulado *Problemas y políticas económicas*, ofertado a través del LMS Sakai

dentro de la iniciativa EDUlib. El resultado del estudio fue la identificación de seis tipos de perfiles de usuarios con distintos niveles de participación:

1. **Fantasmas:** son aquellos usuarios sin apenas actividad a partir de la segunda semana del curso.
2. **Navegadores:** el nivel de actividad de este grupo es muy bajo, aunque algo superior que el de los usuarios fantasma. Ningún miembro de este grupo llegó a enviar prueba alguna, aunque un pequeño porcentaje sí accedió a los recursos del MOOC y participó en los foros.
3. **Auto-evaluadores:** este grupo se dedica principalmente a la realización de cuestionarios y pruebas, obviando cualquier otro tipo de actividad.
4. **Lectores serios:** este grupo es más activo en el acceso a los materiales del curso que el grupo anterior y también presenta unas altas tasas de intentos en las pruebas. Sin embargo, su participación en foros es casi nula.
5. **Activos independientes:** este perfil presenta un mayor nivel de compromiso y actividad que los perfiles precedentes tanto en la realización de cuestionarios como en el consumo de materiales, pero no participó en los foros de discusión.
6. **Activos sociales:** también muestran altos niveles de compromiso con el curso, con un nivel de actividad ligeramente superior al del grupo anterior. Sin embargo, en lo que se distinguen del resto de grupos es en su participación en foros de discusión.

La tabla 7 muestra un resumen de las distintas clasificaciones mixtas analizadas en este apartado. También para esta tabla, aunque las categorías definidas por cada tipología no son exactamente coincidentes, se intentan representar las categorías que guardan mayores similitudes mediante códigos de colores y mediante correspondencia en el nivel de la fila.

TABLA 7. CLASIFICACIONES MIXTAS BASADAS EN ELEMENTOS DE INTERACCIÓN. ELABORACIÓN PROPIA.

	Kizilcec et al. (2013)	Alario-Hoyos et al. (2014)	Coffrin et al. (2014)	Sharma et al. (2015)	Chen et al. (2015)	Coleman et al. (2015)	Kahan et al. (2017)	Khalil y Ebner (2017)	Li y Baker (2018)	Poellhuber et al. (2019)	
Clasificación	Ausentes	Observadores	Auditores	Activos	Suspensos	Ocasionales	Shopping	Degustadores	Desertores	Desconectados	Fantasmas
	Samplig	Drop-ins			Normales			Downloader	Sociales	Auditores	Navegadores
	Asistentes	Latecomers	Activos	Visualizadores	Excepcionales	Realizadores pruebas	Desconectados	Desconectados	Sorteando el sistema	Realizadores pruebas	Autoevaluadores
	Desconectados	Drop-in latecomers			Wiki						Comprometidos fuera línea
	Finalizadores	No comprometidos	Cualificados	Pasivos	Desertores	Each-bit	Finalizadores	Comprometidos en línea	Estudiantes perfectos	Versátiles	Activos independientes
		Finalizadores		Activos	Menos versátiles	Comprometidos sociales moderados		Activos sociales			
				Finalizadores	Versátiles	Comprometidos sociales					
MOOC	3 xMOOC (Un. Stanford)	1 cMOOC MiradaX	2 xMOOC Coursera (Un. Melbourne)	4 xMOOC Coursera	6 xMOOC Coursera	1 xMOOC edX (MITx)	1 xMOOC Coursera	1 xMOOC iMooX (Un. Gratz)	3 xMOOC Coursera	1 xMOOC LMS Sakai	
Particip.	94.091	5.595	55.329	Sin especificar	10.475	43.758	21.899	838	160.402	4.850	
Elementos	Vídeos, tareas y cuestionarios	Cuestionarios, tareas P2P, prueba final	Vídeos, cuestionarios, tareas P2P, calificaciones	Vídeos, pruebas no obligatorias, tareas obligatorias, retraso entrega, número intentos	Vídeos, cuestionarios, foro, wiki	Vídeos, texto, problemas, simulaciones	Vídeos, foros, cuestionarios, tareas, calificaciones, certificados	Frecuencia de lectura y escritura, vídeos, intentos pruebas	Vídeos, cuestionarios, calificaciones	Vídeos, foros, pruebas, pdfs, cuestionarios, visitas	
Técnica	Agrupamiento (k-medias)	Categorías predefinidas	Análisis de regresión	Categorías predefinidas	Agrupamiento	LDA	Agrupamiento	Agrupamiento (k-medias)	Análisis de regresión y robustez	Agrupamiento, análisis correspondencia múltiple, análisis componentes principales	

5.4. PERFILES DE PARTICIPANTES DE MOOC BASADOS EN INTERACCIONES SOCIALES

Los foros de discusión, uno de los elementos característicos de los MOOC, figuran entre los rasgos habitualmente analizados por los diferentes esquemas de clasificación de participantes basados en la interacción con el curso. De este modo, gran parte de las taxonomías enumeradas en el apartado anterior tienen en cuenta la participación en los foros y otras redes como uno de los parámetros para establecer las clasificaciones (Milligan et al., 2013; Hill, 2013; Chen et al., 2015; Sharma et al., 2015; Rodrigues et al., 2016;), e incluso algunas de ellas tienen una categoría distintiva para los participantes que limitan sus interacciones a los foros del curso (Koller et al., 2013; Kovanovic et al., 2016; Kahan et al., 2017; Khalil y Ebner, 2017; Poellhuber et al., 2019). Sin embargo, existe una serie de investigaciones cuyo objeto de estudio se centra exclusivamente en las interacciones de los participantes de los MOOC en foros y otros medios sociales, y que trazan un perfilado de los estudiantes en función de su grado de interacción en estos medios.

Uno de los primeros estudios en identificar perfiles de participantes en función de su interacción en los foros fue el de Grünewald et al. (2013), quienes analizaron los datos de dos xMOOC desarrollados por openHPI. Estos dos MOOC compartieron especificaciones técnicas y estructura, ya que se dividieron en seis unidades semanales con clases en vídeo, materiales de lectura, cuestionarios, tareas para la obtención de un certificado y foros de discusión. El primer curso, *In-Memory Data Management*, con 13.126 alumnos inscritos, se ofreció en inglés y trató un tema avanzado de tecnología de bases de datos, mientras que el segundo curso, *Internetworking mit TCP/IP*, con 9.891 estudiantes registrados, se dirigió a un público no especializado de habla alemana y ofreció una introducción a la tecnología de redes. Los autores analizaron los datos del primer curso e identificaron cinco perfiles de usuarios basados en su interacción con los foros:

- **Inactivos:** aquellos que nunca visitaron el foro.
- **Pasivos:** aquellos usuarios que solo consumieron información y no participaron en discusiones centradas en el contenido. Estos

estudiantes, como mucho, llegaron a presentarse en el hilo correspondiente o a responder a problemas técnicos con un «yo también».

- **Reacting:** son los participantes que complementaron aspectos de cuestiones ya planteadas o propusieron respuestas parciales a una pregunta, pero que nunca ofrecieron respuestas completas que pudieran dar por resuelta una discusión, principalmente debido a su inseguridad.
- **Acting:** son los participantes que plantearon cuestiones a partir de los materiales del curso o de experiencias propias. Estos estudiantes lideraron las discusiones y ofrecieron percepciones nuevas respecto al contenido del curso.
- **Supervisores/sustentadores:** estos estudiantes, al igual que los *acting*, plantearon y respondieron preguntas, y lideraron discusiones. Sin embargo, los supervisores/sustentadores tuvieron la capacidad de obtener una visión global de la mayor parte de las discusiones y del progreso del curso, y fueron capaces de sintetizar el conocimiento adquirido y de enlazarlo con cuestiones de naturaleza similar.

Posteriormente, los autores aplicaron la taxonomía al segundo curso (*Internetworking mit TCP/IP*) y añadieron dos nuevos perfiles:

- Los **expertos en un campo** son usuarios que se mostraron muy activos durante las semanas en las que se estudió un tema en el que estaban especializados. Durante este periodo de tiempo, estos participantes actuaron como una fuente confiable de información, dieron respuestas de alta calidad a las preguntas de la comunidad antes de que el equipo docente pudiera reaccionar y señalaron posibles errores en el material didáctico.
- Los **contribuidores** son una variante del grupo **acting**. Estos participantes introdujeron nuevas formas de contenido de aprendizaje en la plataforma (herramientas, podcasts, soluciones alternativas a tareas) y ampliaron el uso del foro de discusión mediante enlaces en Internet.

Gillani et al. (2014) emplearon un modelo probabilístico denominado *Bayesian Non-negative Matrix Factorization* en el análisis textual de los comentarios en los foros de los 90.000 participantes de un MOOC de estrategia empresarial ofrecido en la plataforma Coursera en la primavera de 2013. Los autores focalizaron su estudio en el análisis de dos subforos distintos, un subforo de casos prácticos, diseñado para que los participantes practicasen las herramientas estudiadas en el curso, y otro de proyecto final, más orientado hacia los aspectos sociales; instrumentales, y prácticos del proceso de aprendizaje, en el que los participantes discutían sus proyectos antes de presentarlos.

A partir del primer subforo, los autores identificaron los siguientes perfiles de usuarios de foro de MOOC:

- **Dinamizadores comprometidos:** los participantes de este grupo son los que más mensajes publicaron y leyeron, y los que consiguieron aprobar el curso en un mayor porcentaje.
- **Iniciadores de discusiones:** los participantes de este grupo se caracterizaron por utilizar sus mensajes para iniciar discusiones.
- **Estrategas:** el perfil de este grupo es similar al de los dinamizadores, con la diferencia de que, a pesar de realizar una mayor proporción de declaraciones argumentativas, las publicaciones de los estrategas reflejaban en menor proporción haber alcanzado algún tipo de aprendizaje, lo cual sugiere que los estudiantes de este grupo utilizaron el subforo únicamente cuando lo necesitaban para lograr sus objetivos de aprendizaje.
- **Individualistas:** este grupo se caracteriza por una gran proporción de declaraciones argumentativas, por una menor actividad en el foro y por una mayor tendencia a no presentar el proyecto final.

A partir del segundo subforo, los autores identificaron otros cinco perfiles de usuarios de foro de MOOC:

- **Buscadores de ayuda instrumental:** a pesar de que los participantes de este grupo iniciaron una elevada cantidad de discusiones, mediante las cuales discutieron principalmente el proyecto final, las

tasas de fracaso entre los miembros de este grupo, tanto en el MOOC como en el proyecto final, son más elevadas que en el resto de los grupos.

- **Evaluadores cuidadosos:** los participantes de este grupo presentan la mayor proporción de mensajes iniciadores de discusiones, pero, a diferencia de los buscadores de ayuda, el foco de sus publicaciones era el proceso de revisión por pares. Esto indica que los participantes de este grupo usaron los foros para resolver cuestiones sobre la revisión por pares y solo volvían a publicar si era necesario.
- **Constructores de comunidades:** los participantes de este grupo publicaron una proporción elevada de mensajes en respuesta a otros, los cuales fueron en promedio de una mayor longitud. A diferencia de los otros grupos, el enfoque de sus discusiones se extendió tanto a los proyectos finales como a la revisión por pares. Una elevada proporción de alumnos pertenecientes a este grupo no presentaron el proyecto final, lo cual sugiere que estaban más interesados en el intercambio de ideas que en aprobar el curso.
- **Triunfadores enfocados:** los participantes de este grupo se distinguieron por tener una mayor proporción de publicaciones con mensajes argumentativos que evidenciaban alguna forma de aprendizaje o que mostraban emociones.
- **Buscadores de ayuda en el proyecto:** los participantes de este grupo se distinguieron por una alta proporción de actos de diálogo imperativo y por propiciar la organización de encuentros virtuales, lo cual sugiere que los miembros de esta categoría buscaban apoyo y oportunidades para colaborar en el proyecto final.

Taylor et al. (2014) emplearon un modelo basado en la regresión logística para predecir el momento de abandono de los estudiantes de *Circuitos y electrónica*, un MOOC ofrecido por edX en otoño de 2012. El estudio dividió a los participantes en cuatro grupos, en función de su actividad colaborativa en el foro y la wiki a lo largo del curso:

1. **Colaborador pasivo:** alumnos que consultaron el foro o la wiki, pero nunca contribuyeron activamente.

2. **Colaborador de wiki:** alumnos que generaron contenido en la wiki, pero nunca publicaron activamente en el foro.
3. **Colaborador del foro:** alumnos que publicaron activamente en el foro, pero no en la wiki.
4. **Colaborador total:** alumnos que generaron contenido en la wiki y publicaron en el foro.

Wang et al. (2016) emplearon análisis de regresión combinado con LDA para analizar el comportamiento cognitivo de los 491 participantes que completaron el MOOC *Introducción a la psicología como ciencia*, ofrecido a través de la plataforma Coursera. A través de los mensajes publicados en los foros del MOOC, el estudio identificó tres perfiles de participantes, cuya mayor complejidad en los comportamientos cognitivos se correlacionó con una mayor profundidad de aprendizaje:

1. **Pensamiento de orden superior:** estudiantes que contribuyeron al menos con una publicación constructiva o interactiva durante el curso.
2. **Atentos:** estudiantes que contribuyeron con al menos una publicación activa durante el curso, pero ninguna publicación constructiva o interactiva.
3. **Fuera de tema:** estudiantes que no contribuyeron con ninguna discusión sobre el tema durante el curso.

Sunar et al (2020) emplearon métodos estadísticos para analizar las interacciones sociales de los 2927 participantes del MOOC *Desarrollando tu proyecto de investigación*, organizado por la Universidad de Southampton entre el 15 de septiembre y el 9 de noviembre de 2014 a través de la plataforma *FutureLearn*. El estudio clasificó los patrones de comportamiento social de los participantes en función de las cadenas de comportamiento construidas mediante los diferentes tipos de acciones sociales en que incurrieron los estudiantes a lo largo de las ocho semanas de duración del curso. Los perfiles identificados fueron los siguientes:

1. **Simple:** alumnos que realizaron contribuciones individuales poco frecuentes. Sus aportaciones se limitaron a escribir un solo

comentario, a interactuar de forma participativa respondiendo a un solo comentario y/o a interactuar de forma pasiva siguiendo a una persona.

2. **Moderadamente frecuente:** alumnos que construyeron cadenas mediante acciones sociales simples, pero muy raramente escribieron varios comentarios; siguieron a varias personas, y/o escribieron varias respuestas a los comentarios.
3. **Frecuente:** estudiantes que realizaron contribuciones e interacciones intensas al escribir múltiples mensajes; responder a varios comentarios, y/o seguir a más de una persona.
4. **Frecuente y persistente:** estudiantes que construyeron cadenas mediante interacciones persistentes y reflexivas realizadas de forma frecuente. Los participantes de este grupo tuvieron interacciones recurrentes con un subconjunto de sus compañeros; respondieron repetidamente a los comentarios de un mismo compañero, y/o interactuaron con sus compañeros de aprendizaje después de seguirlos o ser seguidos.

5.4 PERFILES TEMPORALES DE PARTICIPANTES

La mayoría de las taxonomías analizadas en este capítulo describen un grado cuantitativo o cualitativo de interacción de los estudiantes con ciertos elementos del MOOC, lo cual aporta información valiosa en cuanto al grado de vinculación y compromiso de los participantes con los MOOC, y sobre el tipo de contenido preferido. Sin embargo, son pocas las clasificaciones que estudian algún aspecto que revele cómo los participantes de MOOC gestionan el tiempo que dedican al curso. De este modo, un usuario que complete el curso en una sola sesión intensiva, realizando todas las actividades y participando en los foros, sería categorizado de la misma forma que un usuario que programe la realización del curso dividiéndolo en sesiones más cortas y extendidas en el tiempo, aunque claramente se trate de perfiles de estudiantes con características y preferencias de aprendizaje distintas.

Entre las clasificaciones ya analizadas que estudian algún aspecto de la gestión temporal del curso están las taxonomías de Kizilcec et al. (2013); Alario-

Hoyos et al. (2014); Coleman et al. (2015); Ferguson y Clow (2015), y Poellhuber et al. (2019). En estos trabajos se realiza algún tipo de análisis longitudinal de la actividad de los estudiantes en el MOOC durante aquellos periodos en los que se publica contenido nuevo o se espera que se complete alguna actividad. Por su parte, el estudio de Rodrigues et al. (2016) tiene en cuenta la continuidad en la realización de las actividades y los periodos de ausencia para definir su clasificación, y la investigación de Arora et al. (2017) describe como rasgo distintivo de los exploradores el que accedan al MOOC después de su finalización.

Adicionalmente, existen estudios que clasifican a los participantes de los MOOC en función de cómo interactúan con el curso a lo largo del tiempo. Uno de ellos es el trabajo de Halawa et al. (2014), quienes, como parte de un estudio cuyo objetivo era el diseño de un modelo de predicción del momento en el que los participantes de MOOC entran en peligro de abandono, utilizaron métodos de visualización para identificar cuatro patrones de persistencia de los estudiantes del MOOC basados en su frecuencia de interacción con el curso:

1. **Persistencia continua:** estudiantes que visitan el curso cada pocos días y que tienden a pasar varios días en cada unidad.
2. **Persistencia continua con ausencias prolongadas:** estudiantes que suelen visitar el curso cada pocos días y que, tras una o más ausencias superiores a 10 días, continúan desde donde se detuvieron anteriormente.
3. **Persistencia en ráfagas:** estudiantes que visitan el curso ocasionalmente. Por lo general, cada día que lo visitan prueban contenido de diferentes unidades.
4. **Desertores:** estudiantes que comienzan con un patrón de persistencia continua o en ráfagas, pero que desaparecen totalmente antes del final del curso.

Por su parte, del Peral (2019) aplicó técnicas estadísticas a los datos obtenidos de la tercera edición del LMOOC IP, ofertado en la plataforma UNED Abierta y en el que se inscribieron un total de 8326 participantes. Como resultado, se identificaron cinco perfiles de usuario de LMOOC basados en patrones temporales:

1. **Pre-consumidores:** usuarios que nunca acceden a los contenidos principales del MOOC. En este grupo se distinguen tres subgrupos:
 - El primer subgrupo, «solo inscritos», está formado por alumnos que se inscriben pero que no llegan a acceder al curso en ningún momento.
 - El segundo subgrupo, «solo a foros», está formado por alumnos que no acceden a ninguno de los módulos del curso (ni a los módulos iniciales, ni a los módulos principales, ni a las pruebas de evaluación), pero sí a los foros.
 - El tercer subgrupo está integrado por alumnos que interactuaron con el curso (accedieron a los módulos iniciales y/o realizaron la prueba previa), pero no llegaron a acceder a ninguno de los tres bloques de contenido principal del MOOC.
2. **Usuarios intensivos:** son usuarios que acceden a los contenidos principales del curso en el transcurso de una o dos sesiones. Son, como puede deducirse, los usuarios que dedican un mayor tiempo a cada sesión, pero un menor tiempo total al MOOC, además de realizar un menor número de accesos a los foros del curso.
3. **Usuarios rutinarios:** usuarios que acceden a los contenidos principales del curso en el transcurso de más de dos sesiones y menos de quince, siempre comprendidas dentro de la misma franja horaria. El tiempo total; el número de sesiones, y el número de accesos a los foros de estos usuarios es mayor que en los usuarios intensivos, pero menor que en los flexibles y los exhaustivos. La duración de las sesiones de estos usuarios es ligeramente superior que en los usuarios rutinarios (37,33 minutos de media frente a 32,3), aunque sustancialmente inferior que en los usuarios intensivos (76,27 minutos). Cabe destacar que los usuarios rutinarios son los que obtienen una mayor puntuación media en las pruebas de evaluación.
4. **Usuarios flexibles:** usuarios que dedican al curso entre dos y quince sesiones que no pertenecen a la misma franja horaria. El tiempo total; el número de sesiones, y el número de acceso a los foros es mayor que en los usuarios intensivos y rutinarios, pero menor que en los usuarios exhaustivos.

5. **Usuarios exhaustivos:** usuarios que dedican al curso más de quince sesiones. Son los usuarios que más tiempo invierten en el curso, que más interactúan con los foros, pero que presentan las sesiones más cortas.

Como ya se mencionó en el capítulo introductorio, la hipótesis general que se plantea en este trabajo de investigación es la posibilidad de identificar en un nuevo LMOOC esta última taxonomía de perfiles de usuario de LMOOC basada en patrones temporales de uso descrita por del Peral. En el siguiente capítulo se procederá a detallar la metodología empleada en el proceso de análisis de los datos obtenidos a partir de este segundo LMOOC, el curso PA, y se explicará el algoritmo seleccionado para la identificación de los perfiles que se acaban de describir.

CAPÍTULO 6: DISEÑO DE LA INVESTIGACIÓN

6.1. INTRODUCCIÓN

El objetivo de los capítulos previos ha sido establecer un marco teórico que sirva de andamiaje a esta tesis doctoral y explicar detalladamente los desarrollos; las teorías; las herramientas; las definiciones, y los elementos conceptuales que han sido objeto de estudio en la presente investigación. Por otro lado, los capítulos pertenecientes a la segunda parte de esta tesis están dedicados a la descripción de las distintas fases del proceso de investigación. Así pues, el presente capítulo está dedicado a la primera de estas fases, el procedimiento metodológico empleado en el diseño de la investigación; el siguiente capítulo, el séptimo, analizará los resultados obtenidos a partir de dicho proceso y, por último, el capítulo octavo expondrá las conclusiones derivadas de los resultados.

Como se estableció en el capítulo inicial, el objetivo de este estudio es la utilización de técnicas de LA con el objetivo de identificar perfiles de participantes en el LMOOC PA. Debido a que el volumen de datos obtenidos a partir de este LMOOC es ingente, como se analizará más adelante, resulta imprescindible procesar estos datos, y transformarlos en parámetros numéricos y cuantitativos que posibiliten la realización de un análisis científico. En consecuencia, la metodología seleccionada para esta investigación se enmarca en el método cuantitativo.

Una investigación cuantitativa es una estrategia de investigación que aplica un enfoque deductivo a los datos obtenidos a partir de fenómenos observables. Para ello, las investigaciones cuantitativas recurren a métodos y técnicas empíricos y objetivos, con el propósito de probar y comprender las relaciones entre las distintas variables involucradas en la producción de un fenómeno (Bryman, 2012). El objetivo de la investigación cuantitativa es, por lo tanto, el desarrollo y la aplicación de modelos matemáticos; teorías, e hipótesis relacionadas con los fenómenos. La investigación cuantitativa asume una realidad estable, de la cual se obtienen datos sólidos y replicables, a partir de los cuales es posible obtener conclusiones generalizables (Cook y Reichardt,

1979). Es por este motivo por el que el proceso y las técnicas de medición seleccionados en una investigación cuantitativa son fundamentales, ya que proporcionan una conexión entre la observación empírica y la expresión matemática de las relaciones cuantitativas. (Given, 2008).

Aunque algunos proyectos de LA emplean una metodología cualitativa, la inmensa mayoría de los proyectos de LA siguen una metodología cuantitativa. Como se describió en el capítulo cuarto, los pasos a seguir en un proyecto de LA son los siguientes: (1) recopilación de datos y pre-procesado; (2) análisis y acción, y (3) post-procesado (Romero y Ventura, 2007; Chatti et al., 2012; Calvet Liñán y Juan Pérez, 2015; Romero y Ventura, 2020). De este modo, los siguientes apartados de este capítulo detallarán la metodología, las técnicas y las herramientas empleadas en la implementación de cada uno de estos pasos. Así pues, la siguiente sección comenzará por un análisis de la estructura y las características de los LMOOC IP y PA, los cuales son, respectivamente, el LMOOC a partir del que se definió la taxonomía original y el nuevo LMOOC en el que se busca identificar los mismos perfiles. Posteriormente, dentro de la misma sección, se detallarán las técnicas y las herramientas empleadas en la fase de obtención y pre-procesado de los datos. En el tercer apartado se describirá la metodología empleada en la fase de análisis de los datos, junto con el algoritmo y las herramientas elegidos para efectuar la identificación de los perfiles. Finalmente, el último apartado se centrará en la explicación de la fase de post-procesado.

6.2. RECOPIACIÓN DE DATOS Y PRE-PROCESADO

Los proyectos de LA se basan en la recopilación y extracción de información del *big data*, y tienen como objetivo la optimización de los procesos y los resultados de las organizaciones a partir de una mejor comprensión de las actividades personales (Bienkowski et al., 2012; Sclater, 2017).

Los datos empleados en esta tesis proceden del LMOOC PA, cuyas características difieren del LMOOC IP, el curso del que se tomaron los datos para la taxonomía definida por del Peral (2019). Por este motivo, se considera necesario realizar un análisis de ambos LMOOC y describir tanto su estructura como el tipo de estudiantes que se matricularon en cada curso. Posteriormente,

se describirá la primera fase del proyecto de LA, la cual está compuesta de dos pasos: recopilación de datos y pre-procesado de datos.

6.2.1. FUENTES DE DATOS

LMOOC INGLÉS PROFESIONAL/PROFESSIONAL ENGLISH

La investigación de del Peral (2019) analizó los datos obtenidos a partir de la tercera edición del LMOOC IP, ofertado simultáneamente en los portales web de UNED Abierta y de MiriadaX, y albergados en Open edX⁵², plataforma que proporciona soporte técnico para acoger cursos procedentes de universidades e instituciones de todo el mundo. El LMOOC IP, en el que se inscribieron un total de 8326 participantes, tuvo una duración de tres semanas, del 18 de abril al 9 de mayo de 2017. Con el fin de facilitar un aprendizaje flexible y personalizado, los participantes tuvieron acceso a todos los contenidos desde el principio. El curso quedó habilitado hasta el 22 de mayo, fecha hasta la cual los usuarios ya registrados podían acceder al foro y completar actividades, pero sin apoyo por parte del equipo docente ni la posibilidad de realizarse nuevas inscripciones.

Una particularidad de este LMOOC es que se integró como extensión y complemento de la asignatura *Inglés para fines profesionales*, que se imparte en el primer curso del grado en Turismo de la UNED, y se recompensó a los estudiantes que obtuvieron un certificado con 1 crédito ECTS. El curso combinó por tanto un aprendizaje formal y académico, como es la enseñanza universitaria, con otro informal y no reglado, como son los MOOC, implementando de este modo una metodología pedagógica innovadora (Bárcena et al., 2016; García y Cuevas, 2018).

El LMOOC estuvo dividido en cuatro secciones a las que se accedía a través de pestañas localizadas en la barra superior del curso. La primera sección, «Contenidos», dirigía a una página que ofrecía información básica sobre el MOOC, con datos tales como la identificación del equipo docente; los objetivos del curso, o los requisitos previos.

⁵² <https://openedx.org/>

La segunda sección, «Curso», proporcionaba acceso al contenido fundamental del MOOC y se presentaba en forma de página web navegable, con botones de avance y retroceso, y con una barra de menú lateral que permitía el acceso a todos los módulos del curso. El primer bloque del menú lateral estaba formado por una guía inicial dividida en tres apartados, la cual ofrecía información práctica sobre el uso de los foros, las actividades de autoevaluación y el visionado efectivo de los vídeos.

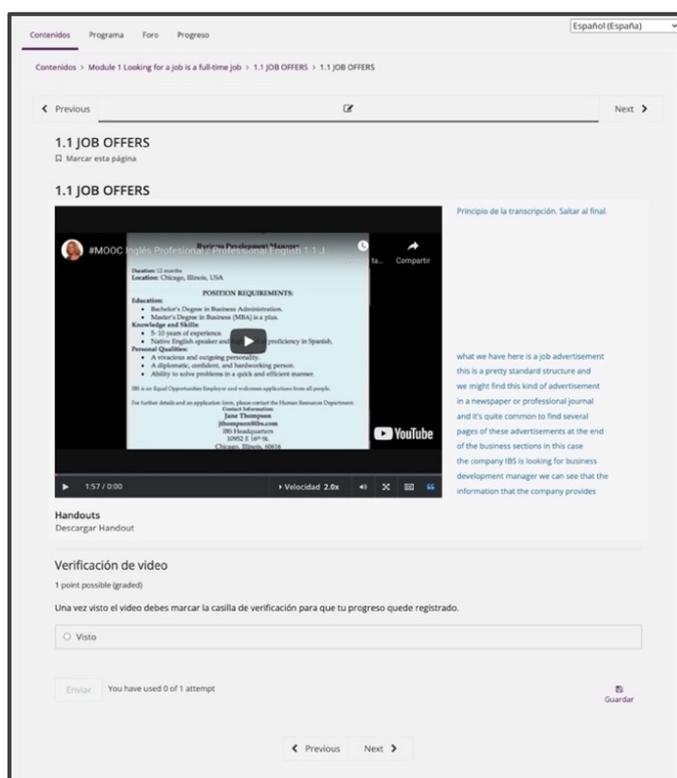
El segundo bloque del menú lateral dirigía a una prueba de conocimientos previos compuesta por 20 preguntas, cuyo objetivo era garantizar que el nivel del alumno se correspondía con el nivel B1 del MCERL.

El tercer bloque del menú lateral estaba denominado «Bloque 0» o bloque de Introducción. En el primer apartado del bloque se incluía un vídeo de 5 minutos y medio de duración con información relativa a la organización y el desarrollo del curso. El segundo apartado dirigía a la sección del foro dedicada a preguntas e información generales. El tercer apartado introducía una actividad especial voluntaria basada en la descarga y utilización de una aplicación de podcasts⁵³, la cual no se tomaba en cuenta para el cálculo de la calificación final del curso.

Los siguientes tres bloques del menú lateral dirigían al contenido fundamental del curso, organizado en tres módulos de estructura paralela. Cada bloque comenzaba con una página en la que se enunciaban los objetivos de aprendizaje del módulo, seguida de cuatro vídeos distintos con una casilla de confirmación que el estudiante debía marcar tras la visualización, como puede observarse en la imagen 1. El hilo conductor de los vídeos de los tres módulos fue la vida de un profesional británico en una empresa multinacional norteamericana. Como apoyo al visionado de los vídeos, se proporcionó un enlace que permitía descargar las transcripciones correspondientes (Bárcena et al., 2016).

⁵³ Debido a que dicha aplicación es un módulo ajeno al MOOC y, por motivo de limitaciones técnicas, no ha sido posible el acceso a los *logs* de la aplicación, la interacción con dicha aplicación no se ha tenido en cuenta para el análisis de los datos de uso de los usuarios y el consiguiente trazado de los perfiles.

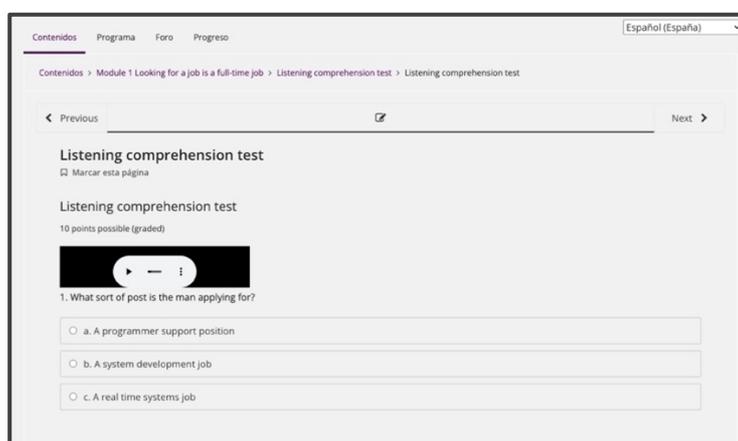
IMAGEN 1. VÍDEO DEL PRIMER MÓDULO DEL LMOOC IP.



Nota. Imagen tomada del LMOOC *Inglés Profesional / Professional English*. Copyright © 2015-2021 UNED Abierta.

En las siguientes páginas de cada bloque, se ofrecía una prueba de evaluación formada por 20 preguntas; un test de comprensión oral compuesto de un archivo de audio y 10 preguntas relativas al mismo, y un último apartado que dirigía a la sección del foro correspondiente al respectivo módulo.

IMAGEN 2. TEST DE COMPRENSIÓN ORAL DEL PRIMER MÓDULO DEL LMOOC IP.



Nota. Imagen tomada del LMOOC *Inglés Profesional / Professional English*. Copyright © 2015-2021 UNED Abierta.

Dentro de cada módulo, la suma de las duraciones de los cuatro vídeos era de 7 minutos y 22 segundos para el primer módulo; 11 minutos y 16 segundos para el segundo, y 6 minutos y 31 segundos para el tercero. La duración de los archivos de audio para el test de comprensión oral de cada módulo era de 2 minutos y 39 segundos para los dos primeros módulos, y 2 minutos y 50 segundos para el tercero.

La última sección del menú lateral contenía una prueba final de evaluación compuesta de 20 preguntas. La tabla 8 muestra un resumen del conjunto de actividades incluidas en el curso.

TABLA 8. RESUMEN DE LAS ACTIVIDADES DEL LMOOC IP. ELABORACIÓN PROPIA.

Módulo 1	Objetivos	4 vídeos (7' 22'')	Test 20 preguntas	Test oral (2' 39'') 10 preguntas	Foro módulo 1
Módulo 2	Objetivos	4 vídeos (11' 16'')	Test 20 preguntas	Test oral (2' 39'') 10 preguntas	Foro módulo 2
Módulo 3	Objetivos	4 vídeos (6' 31'')	Test 20 preguntas	Test oral (2' 50'') 10 preguntas	Foro módulo 3
Test final (20 preguntas)					

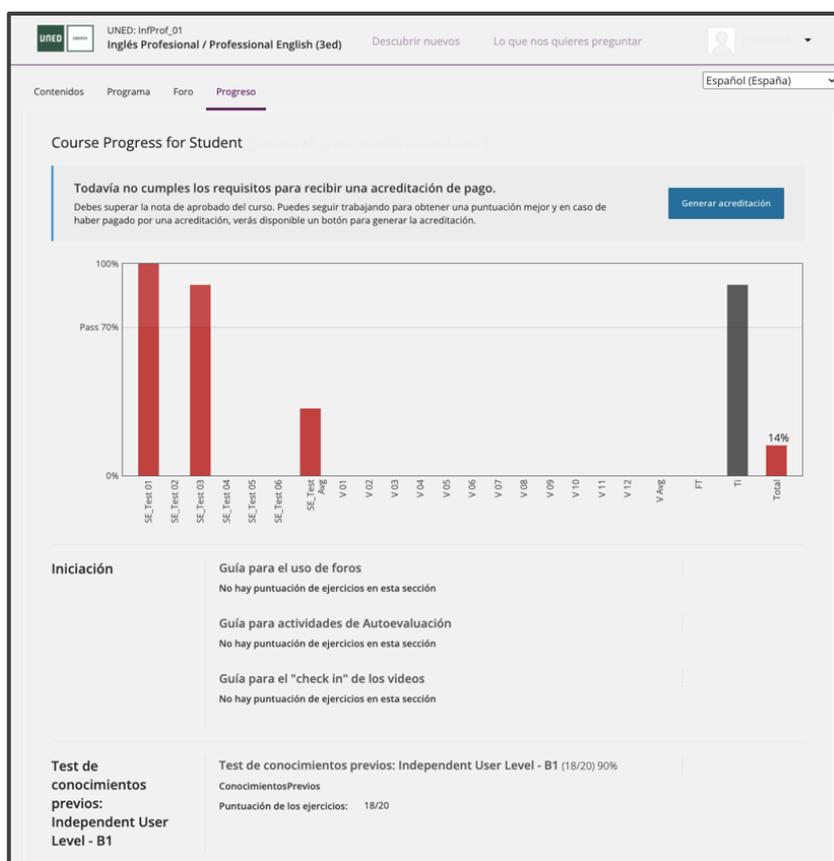
La tercera sección, «Programa», consistía en una tabla de contenido de los tres módulos principales del curso, y un breve resumen de los métodos de evaluación empleados y de las tareas propuestas en el LMOOC.

La cuarta pestaña, «Foro», proporcionaba acceso a cuatro foros moderados por los miembros del equipo docente: un foro general dedicado a dudas generales del curso y dudas de carácter técnico, y tres foros dedicados a los módulos del curso. Dichos foros fueron atendidos por los facilitadores durante el periodo oficial de desarrollo del curso, y en ellos cada semana se trataron dos aspectos gramaticales y se comentaron algunos de los errores más frecuentes a través de preguntas de reflexión.

Por último, la pestaña «Progreso» mostraba el avance individual del estudiante en el MOOC y dividía la información de dos modos distintos, como bien puede observarse en la imagen 4. Por un lado, un gráfico de barras ofrecía información estadística de progreso para cada una de las actividades del curso. Por otro lado, los distintos módulos y submódulos eran presentados en forma de

lista con hipervínculos que mostraban la puntuación asignada a cada una de las tareas de evaluación. Adicionalmente, en la parte superior de esta pestaña se mostraba un botón que dirigía a una página que, si se cumplían los requisitos de evaluación, permitía generar una acreditación de superación del curso.

IMAGEN 3. SECCIÓN DE PROGRESO DEL LMOOC IP.



Nota. Imagen tomada del LMOOC *Inglés Profesional / Professional English*. Copyright © 2015-2021 UNED Abierta.

DESCRIPCIÓN DEL LMOOC PUERTAS ABIERTAS I

El LMOOC PA también fue ofrecido a través del portal de UNED Abierta mediante la plataforma Open edX. El curso, cuyo número total de inscritos fue de 2252, tuvo una duración seis semanas, del 15 de enero al 10 de marzo de 2019. El curso fue diseñado y desarrollado por el grupo de investigación ATLAS⁵⁴

⁵⁴ ATLAS (Applying Technology to Languages)
http://portal.uned.es/portal/page?_pageid=93,8842771&_dad=portal&_schema=PORTAL.

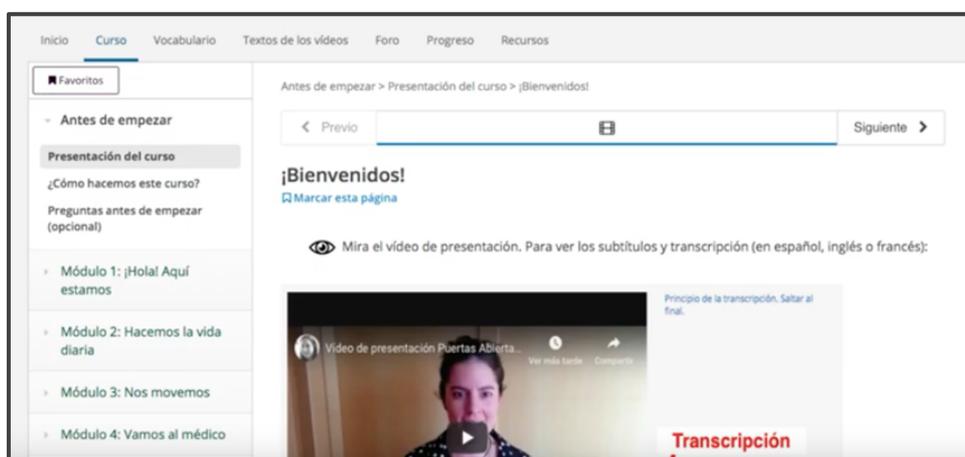
de la UNED, dentro del marco del proyecto europeo MOONLITE⁵⁵ (*Massive Open Online courses eNhancing LInguistic and Transversal skills for social inclusion and Employability*). Este proyecto se ejecutó entre 2016 y 2019, con la colaboración de 16 entidades de apoyo y recepción de personas migrantes y refugiadas, las cuales contribuyeron al análisis de las necesidades formativas; vivenciales; vitales, y tecnológicas de los participantes. Por consiguiente, el LMOOC PA estuvo dirigido a personas desplazadas, particularmente a migrantes y refugiados, y su principal objetivo fue el fomento de las habilidades lingüísticas; interculturales, y transversales como medio de alcanzar la inclusión social y la empleabilidad entre los participantes pertenecientes a estos colectivos. De hecho, una de las características de este LMOOC fue su apertura a todos los niveles, ya que no se exigía ningún requisito académico o administrativo para acceder al curso, y todo el contenido fue gratuito, incluido el certificado de realización, algo poco frecuente entre los MOOC (Castrillo y Sedano, 2021; Sanz Gil, 2021).

De manera adicional, ciertas características particulares en los niveles técnico, lingüístico y pedagógico hicieron que el LMOOC PA se distinguiera de otros MOOC. Desde el punto de vista técnico, en el diseño del curso se buscó un nivel elevado de usabilidad y de accesibilidad, mediante la implementación de un entorno de navegación lo más sencillo posible (Sanz Gil, 2021).

El curso estaba dividido en seis secciones a las que se accedía a través de pestañas localizadas en la barra superior de la página. La sección con el contenido principal del LMOOC, denominada «Curso», se presentaba en forma de página web navegable con botones de avance y retroceso, e incluía una barra de menú lateral que permitía el acceso a cada uno de los apartados de todos los módulos del curso, como puede observarse en la imagen 5.

⁵⁵ Proyecto europeo Erasmus + MOONLITE, Key Action 203 – Strategic Partnerships for Higher Education, ref. 2016-1-ES01-KA203-02573.

IMAGEN 4. EJEMPLO DE NAVEGACIÓN EN UNA PÁGINA DEL LMOOC PUERTAS ABIERTAS I.



Nota. Imagen tomada del LMOOC *Puertas Abiertas: Curso de español para necesidades inmediatas I*. Copyright © 2015-2021 UNED Abierta.

Del resto de secciones, una de ellas, «Textos de los vídeos», contenía enlaces a ficheros descargables con las transcripciones a varios idiomas de las grabaciones del curso, como puede observarse en la imagen 6; otra de ellas, «Vocabulario», presentaba un glosario multilingüe de términos, como puede observarse en la imagen 7, y otra ofrecía enlaces a diccionarios visuales, y a diccionarios y traductores en diversas lenguas.

IMAGEN 5. TRANSCRIPCIONES MULTILINGÜES DE LOS TEXTOS DE LOS VÍDEOS DEL CURSO.



Nota. Imagen tomada del LMOOC *Puertas Abiertas: Curso de español para necesidades inmediatas I*. Copyright © 2015-2021 UNED Abierta.

IMAGEN 6. GLOSARIO MULTILINGÜE EN LA SECCIÓN VOCABULARIO.

UNED MOONLITE_L_001
Puertas Abiertas: Curso de español para necesidades inmediatas (I) Descubrir nuevos Lo que nos quieres preguntar

Contenidos **Vocabulario** Textos de los vídeos Foro Progreso Recursos

Aquí puedes ver y descargar el [vocabulario del curso](#)

Vocabulario	Explicación y ejemplo	English	Français	العربية
(Estar) a dos calles	La distancia es de dos calles. <i>Ejemplo: Mi casa está a dos calles del parque.</i>	Two streets away	À deux rues	على بُعد شارعين
Administrativo (el)	Persona que trabaja en una oficina con documentos (papeles) o con el ordenador. <i>Ejemplo: El administrativo hace mi certificado de empadronamiento.</i>	Clerk	Commis	موظف إداري
(Tener) alergia	Hay algo que me sienta mal. <i>Ejemplo: No puedo tomar penicilina porque tengo alergia.</i>	I have an allergy	J'ai une allergie	لدي حساسية
Antibiótico (el)	Medicina para curar algunas enfermedades. <i>Ejemplo: El antibiótico es bueno para la faringitis.</i>	Antibiotic	Antibiotique	مضاد حيوي

Nota. Imagen tomada del LMOOC Puertas Abiertas: Curso de español para necesidades inmediatas I. Copyright © 2015-2021 UNED Abierta.

Algunas características técnicas adicionales que facilitaban a los participantes la navegación y la comprensión del contenido del curso fueron una guía de uso; mecanismos de soporte técnico; recursos audiovisuales de baja resolución adaptados para dispositivos móviles y, por extensión, accesibles desde una conexión inestable o de ancho de banda reducido; materiales descargables en diversos formatos (vídeo, audio, imagen, PDF), como puede verse en la imagen 8, y la adaptación de los elementos gráficos del LMOOC a personas con dificultades visuales (Sanz Gil, 2021).

IMAGEN 7. MATERIALES DESCARGABLES CORRESPONDIENTES AL MÓDULO 1.

Materiales Módulo 1
 Marcar esta página

Para estudiar en casa sin conexión a internet,

- Los [contenidos del Módulo 1](#).
- Las [soluciones de las actividades](#).
- Los dos vídeos del módulo: [Vídeo 1](#) y [Vídeo 2](#).

Nota. Imagen tomada del LMOOC Puertas Abiertas: Curso de español para necesidades inmediatas I. Copyright © 2015-2021 UNED Abierta.

Desde el punto de vista lingüístico e intercultural, PA destacó por la sencillez de expresión y por adaptar tanto los vídeos como los audios a una velocidad comprensible para los participantes de nivel inicial. El curso también ofreció un glosario de términos, y subtítulos traducidos a lenguas vehiculares identificadas como mayoritarias entre la población migrante, tales como el inglés; el francés, y el árabe, como puede observarse en las imágenes 6 y 7. De manera adicional, el curso prestó atención a la diversidad de identidades y procedencias en los materiales y en las actividades propuestas, a través de las cuales se fomentó la conciencia y la comunicación interculturales entre los participantes (Castrillo y Sedano, 2021).

Finalmente, a nivel pedagógico, los contenidos del LMOOC fueron de carácter eminentemente práctico, ya que fomentaron los intercambios colaborativos y en línea en lengua española en los formatos oral y escrito. De hecho, los contenidos del curso estuvieron orientados a las necesidades de los participantes, de tal modo que se procuró facilitar el aprendizaje del español necesario para desenvolverse en los aspectos más elementales de la vida cotidiana, a la vez que se aprendían las convenciones culturales inherentes a las situaciones comunicativas más frecuentes (Castrillo y Sedano, 2021; Sanz Gil, 2021).

Como resultado, los temas propuestos en cada uno de los módulos estuvieron dirigidos, entre otros, a ayudar a los participantes a realizar trámites cotidianos, desenvolverse por la ciudad, hacer planes, buscar vivienda, ir al médico o defender sus derechos, como puede verse en la siguiente tabla de materias del contenido del curso:

TABLA 9. TABLA DE CONTENIDO DEL LMOOC PUERTAS ABIERTAS I

Puertas Abiertas I	
Antes de empezar	Presentación del curso
	¿Cómo hacemos este curso?
	Preguntas antes de empezar
Módulo 1: ¡Hola! Aquí estamos	1.1. Los sonidos del español
	1.2. Nos comunicamos
	1.3. En la oficina de empadronamiento. Hacemos trámites
	1.4. Materiales Módulo 1
	1.5. Actividad final Módulo 1

Módulo 2: Hacemos la vida diaria	2.1. Actividades diarias
	2.2. En una cafetería después de clase
	2.3. En la calle con un amigo
	2.4. Materiales Módulo 2
	2.5. Actividad final Módulo 2
Módulo 3: Nos movemos	3.1. El barrio de Nadia
	3.2. ¿Cómo se va?
	3.3. Vamos en tren
	3.4. Materiales Módulo 3
	3.5. Actividad final Módulo 3
Módulo 4: Vamos al médico	4.1. En la consulta del médico de familia
	4.2. En la farmacia
	4.3. En el pediatra
	4.4. Materiales Módulo 4
	4.5. Actividad final Módulo 4

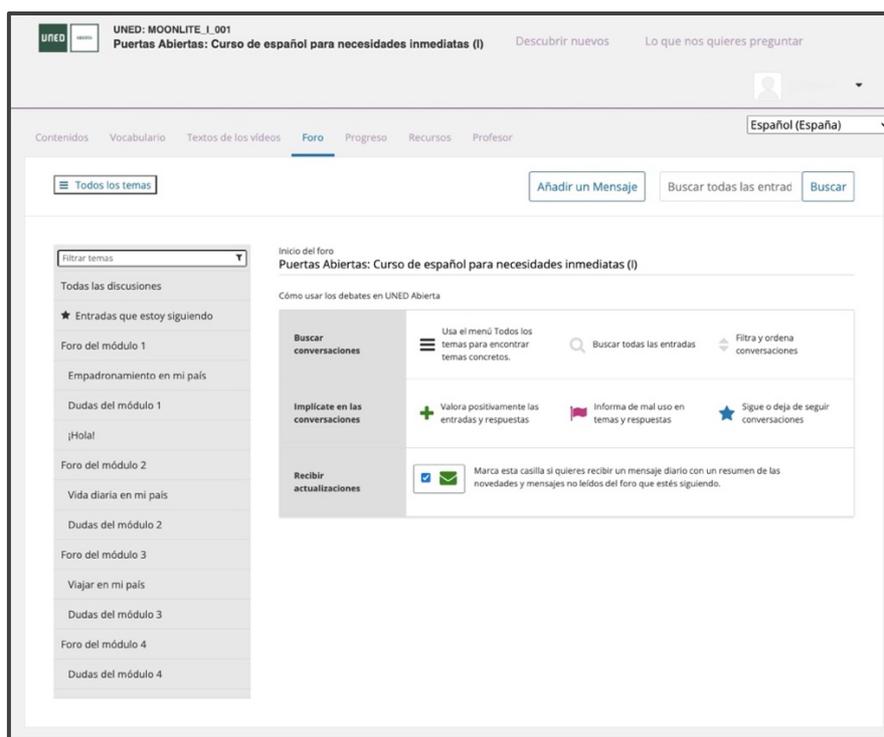
De manera adicional, el diseño del LMOOC fomentó el aprendizaje autónomo mediante la inclusión de temas que eran obligatorios para la obtención del certificado y otros temas que fueron opcionales, además de presentar un formato modular; flexible, y progresivo basado en el andamiaje y la retroalimentación positiva por parte de los facilitadores (Castrillo y Sedano, 2021; Sanz Gil, 2021).

Se estimaron un total de 25 horas para completar el curso, con la siguiente propuesta de calendario para las seis semanas de duración del LMOOC: una primera semana destinada a familiarizarse con la plataforma y los materiales; una semana para cada uno de los cuatro módulos de contenido, y una semana final de revisión o para completar tareas pendientes (Castrillo y Sedano, 2021; Sanz Gil, 2021).

La primera sección del curso consistía en un módulo de introducción con vídeo tutoriales que explicaban cómo orientarse por el LMOOC y cómo realizar cada una de las actividades que formaban parte del curso (ver vídeos, escuchar audios, descargar materiales, participar en los foros, etc.). Posteriormente, cada uno de los módulos del contenido principal del curso se componía de actividades de aplicación práctica en contextos de integración y empleabilidad, basadas en numerosos textos; imágenes; infografías; audios; vídeos; actividades de auto evaluación, y P2P. Al final de cada módulo se ofrecía un enlace al foro de debate correspondiente, en el que los participantes exponían sus experiencias y puntos de vista, y explicaban el contraste entre los

aspectos cotidianos de su país y cultura de origen, y los de su país de acogida. Una pestaña en la barra superior, bajo el nombre «Foro», también proporcionaba acceso a los cuatro foros de los módulos, y a un foro general dedicado a dudas globales sobre el curso y a otras cuestiones de carácter técnico (Castrillo y Sedano, 2021; Sanz Gil, 2021).

IMAGEN 8. SECCIÓN DE FOROS DE LMOOC PA.



Nota. Imagen tomada del LMOOC *Puertas Abiertas: Curso de español para necesidades inmediatas I*. Copyright © 2015-2021 UNED Abierta.

Por último, al igual que en el LMOOC IP, había una pestaña de «Progreso» que mostraba el avance del estudiante en el curso dividiendo la información en un gráfico de barras y en una lista de las secciones de los módulos. También desde esta sección se accedía a una página que permitía generar la acreditación gratuita de superación del curso, en el caso de que se cumplieren los requisitos de evaluación requeridos.

6.2.2. RECOPIACIÓN DE DATOS

Según Chatti et al. (2012), las fuentes de datos para proyectos de LA pueden proceder de entornos de aprendizaje distribuidos o, en el caso del

presente estudio, de sistemas educativos centralizados. La actividad de los usuarios de los dos LMOOC analizados en este capítulo se almacenó en el servidor de la plataforma edX mediante una serie de ficheros de *log*, de tal forma que todas las acciones de los participantes dentro de los cursos quedaron registradas en estos archivos. Un archivo de *log* es un archivo de texto o un archivo XML que se utiliza en la mayoría de los servidores informáticos para registrar el proceso automático de producción, en una fecha y hora determinadas, de documentación relativa a eventos; comportamientos, y condiciones relevantes dentro de un sistema informático (Waters et al., 2004; DeLaRosa, 2021). Con respecto a los ficheros de *log* de los LMOOC, cada una de las acciones realizadas por los usuarios dentro del curso se almacenó en un objeto de tipo JSON, un formato de intercambio de datos derivado del lenguaje de programación JavaScript. Este formato es utilizado habitualmente en las transacciones informáticas entre un navegador y un servidor, o entre dos servidores. JSON utiliza una representación textual definida por un reducido conjunto de reglas que rigen la forma en la que se estructuran los datos. Las especificaciones del formato JSON establecen que los datos se pueden estructurar, bien como una colección de pares de nombre y valor, bien como una lista ordenada de valores (Smith, 2015). A continuación, se muestra un registro JSON obtenido de los *logs* del LMOOC PA⁵⁶:

```
{ "username": "xxxxxxx", "event_source": "browser", "name": "pause_video",
  "accept_language": "es-ES,es;q=0.9", "time": "2019-04-02T17:07:45.247514+00:00", "agent":
  "Mozilla/5.0 (Linux; Android 4.4.4; SM-J100H) AppleWebKit/537.36 (KHTML, like Gecko)
  Chrome/71.0.3578.99 Mobile Safari/537.36", "page":
  "https://iedra.uned.es/courses/course-
  v1:UNED+MOONLITE_II_001+2018/courseware/00ccd40329d3475291423755eaed58a1/a3f
  d35cb5b0043d9a649b73c37d118c6/", "host": "iedra.uned.es", "session":
  "e72a13fb8f86f7764fb69316f28dff86", "referer": "https://iedra.uned.es/courses/course-
  v1:UNED+MOONLITE_II_001+2018/courseware/00ccd40329d3475291423755eaed58a1/a3f
  d35cb5b0043d9a649b73c37d118c6/", "context": {"user_id": 217604, "org_id": "UNED",
  "course_id": "course-v1:UNED+MOONLITE_II_001+2018", "path": "/event"}, "ip":
  "xxx.xxx.xxx.xxx", "event": "{ \"code\": \"mDRNRQdwBCI\", \"id\":
```

⁵⁶ Tanto el nombre de registro del usuario como su dirección IP han sido reemplazados dentro del objeto JSON por cadenas del carácter «x».

```
\\"6d094f620e9149b28d90a49c1665153b\\", \\"currentTime\\": 98.63874900572205}",  
"event_type": "pause_video"}
```

Como puede observarse, el objeto JSON contiene una serie de campos relativos a la interacción del usuario con alguno de los contenidos del curso. En este caso específico, la acción realizada es la de pausar la reproducción de un vídeo. Algunos de los campos incluidos en el JSON son: el nombre del usuario, la hora de la transacción, el navegador desde el que se realiza la conexión, la URL a la que se accede, la dirección IP y el tipo de acción realizada. Durante el transcurso del LMOOC PA, se generaron un total de 394 archivos de *log*, con un tamaño acumulado de 4,51GB.

Algunos LMS, entre ellos la plataforma edX, integran herramientas que permiten la generación de informes simples a partir de los datos educativos (Romero y Ventura, 2007). De este modo, a partir de la información almacenada en la base de datos de la plataforma edX tras la finalización del LMOOC PA, fue posible generar un fichero CSV con información relativa al desempeño de cada participante del curso. Los archivos CSV (del inglés, *comma-separated values*) son un tipo de documento no estandarizado que permite representar datos en forma de tabla. De este modo, las distintas columnas que conforman la tabla se separan mediante un carácter delimitador (suele ser la coma, aunque en formatos alternativos puede usarse el punto y coma, o la tabulación) y las filas se separan mediante saltos de línea (Shafranovich, 2005). Dentro del fichero CSV generado para PA, en la primera línea se representa el nombre de las columnas que forman la tabla, y cada una de las siguientes filas contiene, para cada uno de los participantes del curso, el conjunto de valores de los campos definidos en las columnas. A continuación, se ofrece el contenido de la primera fila del fichero CSV:

```
Student ID,Email,Username,Grade,Actividad 1: 1.1. Los sonidos del español,Actividad 2:  
1.2. Nos comunicamos,Actividad 3: 1.3. En la oficina de empadronamiento. Hacemos  
trámites,Actividad 4: 2.1. Actividades diarias,Actividad 5: 2.2. En una cafetería después de  
clase,Actividad 6: 2.3. En la calle con un amigo,Actividad 7: 3.1. El barrio de  
Nadia,Actividad 8: 3.2. ¿Cómo se va?,Actividad 9: 3.3. Vamos en tren,Actividad 10: 4.1.  
En la consulta del médico de familia,Actividad 11: 4.2. En la farmacia,Actividad 12: 4.3. En  
el pediatra,Actividad (Avg),Actividad final 1: 1.5. Actividad final Módulo 1,Actividad final  
2: 2.5. Actividad final Módulo 2,Actividad final 3: 3.5. Actividad final Módulo 3,Actividad
```

final 4: 4.5. Actividad final Módulo 4,Actividad final (Avg),Enrollment Track,Verification Status,Certificate Eligible,Certificate Delivered,Certificate Type

Como se puede observar, además de los tres campos que permiten identificar a cada usuario («Student ID», «Email» y «Username»), el fichero incluye la nota del alumno en cada una de las pruebas del LMOOC, además de las medias de los cuestionarios intermedios y de las actividades finales. El fichero CSV también contiene campos que indican si el alumno tiene derecho a recibir la acreditación del curso y si dicha acreditación ha sido expedida.

Las fuentes de los datos en los que se basa esta investigación pertenecen, por lo tanto, a dos de las categorías de datos educativos que define Sclater (2017). De este modo, el fichero CSV contiene datos académicos de cada estudiante que son generados durante el proceso de aprendizaje, mientras que los *logs* registran los datos relativos a la propia actividad de aprendizaje.

6.2.3. PRE-PROCESADO DE DATOS

Una vez recopilados los datos, la transformación o pre-procesado de los datos es un elemento esencial de las LA. Como se ha analizado en el apartado anterior, los datos obtenidos a partir del LMOOC PA presentan diferentes formatos, por lo que se requiere de una transformación a un formato específico que se pueda utilizar como entrada para un método particular de LA (Chatti et al., 2012; Waterman y Bruening, 2014; Sclater, 2017). Como se analizó en el capítulo cuarto, las técnicas empleadas en la fase de pre-procesado de los datos suelen ser técnicas procedentes de la EDM, tales como la limpieza de datos; la integración de datos; la transformación de datos; la reducción de datos; el modelado de datos, o la identificación de usuarios y sesiones (Liu, 2006; Romero y Ventura, 2007; Chatti et al., 2012; Calvet Liñán y Juan Pérez, 2015).

En el presente estudio, la mayor parte de la fase de pre-procesado de datos ha sido implementada mediante el diseño de un programa informático *ad hoc* en el lenguaje de programación Java, cuyo código está disponible en el Anexo I. Este programa ha sido escrito mediante el *software* de Eclipse⁵⁷, un

⁵⁷ <https://www.eclipse.org/>

entorno de desarrollo especialmente diseñado para proyectos Java. Java es un lenguaje de programación orientado a objetos que fue lanzado por *Sun Microsystems* en 1995 y que continúa siendo uno de los lenguajes más populares del mundo, particularmente en el ámbito de las aplicaciones web cliente-servidor. Como en otros lenguajes orientados a objetos, la unidad fundamental del código Java son las clases, las cuales pueden definirse como componentes de aplicaciones que contienen código y datos ejecutables. Las clases de Java compiladas, las cuales son cargadas dinámicamente en tiempo de ejecución a medida que las necesita una aplicación, se distribuyen en un formato binario universal que contiene el código de bytes de Java, lo que hace de Java un lenguaje seguro; rápido, y portable, capaz de ser ejecutado en cualquier ordenador y sistema operativo (Niemeyer y Leuck, 2013; Schildt, 2021). La plataforma Java proporciona un conjunto integral de bibliotecas propias de clases que permiten implementar gran parte de las funcionalidades y operaciones que se necesitan dentro del código de un programa informático. Sin embargo, Java también permite utilizar paquetes externos de clases con el objetivo de acceder a funcionalidades adicionales. De este modo, para la lectura de los objetos JSON contenidos en los *logs*, se ha utilizado una librería externa llamada *org.json*⁵⁸.

Dentro de la fase de pre-procesado de datos, las tareas que ha sido necesario implementar son las siguientes:

LIMPIEZA DE DATOS

Durante la fase de pre-procesado, se identificaron dos tipos distintos de datos inválidos: entradas incompletas y registros falsos. Por un lado, de entre los casi dos millones de objetos JSON almacenados en los ficheros, una treintena de objetos JSON estaban corruptos, ya que les faltaba un fragmento y provocaban un error al ser leídos por el programa. La forma de gestionar estas entradas corruptas fue la siguiente: si el fragmento presente en el fichero de *log* contenía los campos que se necesitaban dentro del programa, el objeto se reparaba manualmente eliminando el campo que había quedado incompleto y añadiendo el carácter de cierre «}»;

⁵⁸ <https://github.com/stleary/JSON-java>

que se utilizaban en el programa no estaba presente o estaba incompleto, el objeto JSON se eliminaba del fichero. Por otro lado, en los *logs* se incluían datos generados por instructores y facilitadores que se registraron en la plataforma para realizar pruebas. Estas entradas fueron excluidas desde el propio código del programa, discriminando los JSON correspondientes a partir del nombre de usuario.

IDENTIFICACIÓN DE SESIONES

El cálculo del tiempo total dedicado por los estudiantes al curso representa un desafío, porque los datos incluidos en los *logs* no registran el momento en el que los estudiantes se vuelven inactivos. Cada vez que un estudiante interactúa con uno de los elementos de una página del curso, se registra un objeto JSON en los *logs*, por lo que *a priori* es posible hacer una estimación del tiempo dedicado al curso a partir de la diferencia de tiempo entre los distintos eventos. Sin embargo, no es posible determinar con precisión el tiempo dedicado a la última página a la que accede un estudiante.

Dentro de la informática, se define una sesión como un único conjunto continuo de actividades atribuible a un navegador o a un usuario con cookies, resultando en una o más descargas de texto y/o gráficos extraídos de un sitio web (Interactive Advertising Bureau, 2013). Aunque cada sitio web puede configurar el valor del umbral de inactividad tras el cual se considera que una sesión ha concluido, el valor más generalizado es el de 30 minutos (Mullaney, 2014; Ho et al., 2015; Google, 2022). Esto significa que, por lo general, si un usuario no realiza ninguna actividad en un sitio web durante más de 30 minutos, la sesión del usuario se considera finalizada y se ignora el tiempo transcurrido desde la última acción realizada por el usuario.

En consecuencia, a la hora de calcular la duración de las sesiones de los usuarios del LMOOC PA, se ha tomado 30 minutos como el valor del umbral de inactividad, de tal modo que cualquier actividad realizada por un usuario transcurridos 30 minutos desde el último evento registrado en los *logs* se considera parte de una nueva sesión. Este es también el valor que del Peral (2019) utilizó para el cálculo de las sesiones de los usuarios del LMOOC IP. Aunque la elección de este valor puede producir una sobreestimación del

tiempo que dedican al curso las personas que están distraídas con otras tareas, para el LMOOC PA este valor resulta razonable, ya que la duración de cada uno de los vídeos del curso es de alrededor de un minuto y las actividades de evaluación están diseñadas para ser de corta duración. Adicionalmente, ante la posibilidad de que un estudiante hubiese podido dedicar mucho tiempo a redactar una respuesta en el foro antes de pulsar el botón de publicar, se ha comprobado que la mayoría de los mensajes en el foro son de una sola oración y en ningún caso se encontraron párrafos cuya redacción pudiera llevar más de 30 minutos.

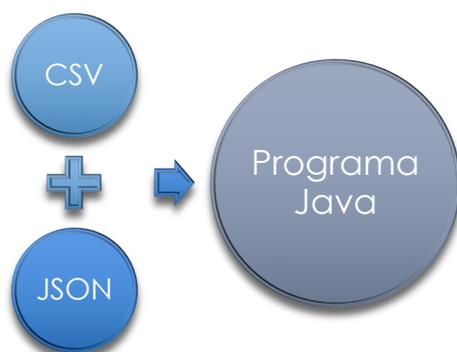
La forma que tiene el programa de calcular las sesiones es por lo tanto la siguiente:

1. Para cada uno de los usuarios, se almacenan en una lista todos los eventos registrados en los *logs*.
2. Se ordenan los eventos de la lista en orden cronológico.
3. A partir de los eventos, se crea para cada usuario una lista de sesiones, y se almacenan los instantes de inicio y finalización de las sesiones.

INTEGRACIÓN DE DATOS

Una de las tareas que implementa el programa Java es la lectura de los dos tipos de ficheros generados por la plataforma edX, los ficheros de *log* y el archivo CSV, y la integración de la información almacenada en ellos en una serie de listas de objetos. De esta forma, el programa Java procesa la información de los *logs* y crea una lista de usuarios a la que, a su vez, asocia una lista de sesiones. Posteriormente, el programa lee el fichero CSV y, para cada uno de los usuarios incluidos en el fichero, accede al listado de usuarios generado a partir de los *logs* y añade a cada participante la siguiente información contenida en el fichero CSV: notas de cada actividad, notas medias, notas finales, si el usuario es elegible para la acreditación y si el usuario solicitó la acreditación. De esta forma, en el momento de la ejecución el programa integra y almacena la información relativa a cada uno de los participantes que se considera relevante y que se encuentra almacenada en los dos tipos de ficheros disponibles.

FIGURA 21. INTEGRACIÓN DE LOS DATOS PROCEDENTES DEL LMOOC. ELABORACIÓN PROPIA.



REDUCCIÓN DE DATOS

Una de las condiciones que establece del Peral (2019) para clasificar a los participantes de LMOOC es la tendencia a concentrar las sesiones de estudio en las mismas franjas horarias, de modo que en su trabajo se establece una categoría específica para ello: los participantes rutinarios. Desafortunadamente, no existe un método evidente que permita obtener, a partir de la lista de sesiones que el programa Java calcula para cada usuario, una variable que determine el grado de concentración en franjas horarias de las sesiones de estudio de un participante. De esta forma, resulta complicado calcular medidas de dispersión a partir de los datos relativos a los instantes de inicio de las sesiones de los participantes. Por ejemplo, para un usuario que en un día determinado inicie una sesión a las 23:59 y en otro día diferente inicie sesión a las 00:01, se debería considerar una diferencia horaria de dos minutos en vez de 11 horas y 58 minutos.

Por consiguiente, para poder obtener una variable que mida la concentración horaria de las sesiones de los participantes del LMOOC, se ha recurrido a la medida de dispersión denominada mediana desviación absoluta (MDA en adelante), una medida de dispersión robusta que presenta una menor sensibilidad ante valores extremos (Leys et al., 2013). La MDA se define como la mediana de las desviaciones absolutas de la mediana de los datos (Huber, 1981), de tal modo que:

$$MDA_n = \text{med} \{ |x_i - M_n| \},$$

donde

$$M_n = \text{med} \{x_i\}$$

De este modo, para el cálculo de la MDA, el programa almacena en una lista ordenada el valor en segundos de todas las horas de inicio de cada una de las sesiones de un usuario y calcula la mediana de esa lista. A partir de ese valor, y para cada uno de los usuarios, el programa calcula y almacena la MDA.

MODELADO DE DATOS

Los datos generados por el programa Java durante la fase de pre-procesado se han volcado en una tabla formada por cuatro columnas y 1777 filas, donde la primera fila contiene el nombre identificador de cada una de las columnas y el resto de las filas contienen, para cada uno de los participantes del curso, los valores de las distintas variables (previos a la operación de escalado) obtenidos en la fase de pre-procesado. El contenido de las columnas es el siguiente:

- La primera columna contiene el nombre de cada participante.
- La segunda columna contiene la nota final de cada participante.
- La tercera columna contiene la MDA de las horas de inicio de las sesiones de cada participante.
- La cuarta columna contiene el número total de sesiones de cada participante.

Siguiendo esta estructura de tabla, el programa Java genera un archivo CSV en el que cada uno de los valores dentro de una fila aparece separado por una coma y cada nueva fila aparece indicada por un retorno de línea.

TRANSFORMACIÓN DE DATOS

Antes de poder aplicar a las variables obtenidas el algoritmo k-medias, es necesario realizar una transformación de los datos. El objetivo es minimizar el sesgo de aquellas características cuya contribución numérica sería mayor en los cálculos de las distintas agrupaciones. Este paso implica la transformación de las variables a un rango común, de tal modo que los valores de las variables numéricas más grandes no puedan dominar a los valores de las variables numéricas más pequeñas (Singh y Singh, 2020). La transformación de las variables suele realizarse mediante una de estas dos operaciones: la normalización y el escalado.

La normalización se emplea en algoritmos que presuponen una distribución normal o gaussiana de los datos, tales como los t-tests, el análisis de la varianza, la regresión lineal, el LDA o el clasificador bayesiano ingenuo. Por otra parte, el escalado implica la transformación de los datos para que se ajusten a una escala específica y se emplea en algoritmos basados en cálculos relativos a la separación de los datos, tales como máquinas de vectores de soporte; KNN, o k-medias (Tatman, 2018).

A partir de las variables obtenidas en la fase de pre-procesado, si se optara por la normalización de los datos, una diferencia de un segundo en la medida de distribución de las franjas horarias tendría el mismo peso en la formación de los distintos grupos que una diferencia de uno en el número de sesiones. Esto, evidentemente, produciría agrupamientos no deseados. En consecuencia, los valores de las variables se han sometido a un proceso de transformación mediante escalado, utilizando para tal efecto las funciones incluidas en el programa estadístico R, del cual se hablará con más detalle en la siguiente sección.

FIGURA 22. OPERACIONES REALIZADAS EN LA FASE DE PRE-PROCESADO. ELABORACIÓN PROPIA.



6.3. ANÁLISIS Y ACCIÓN

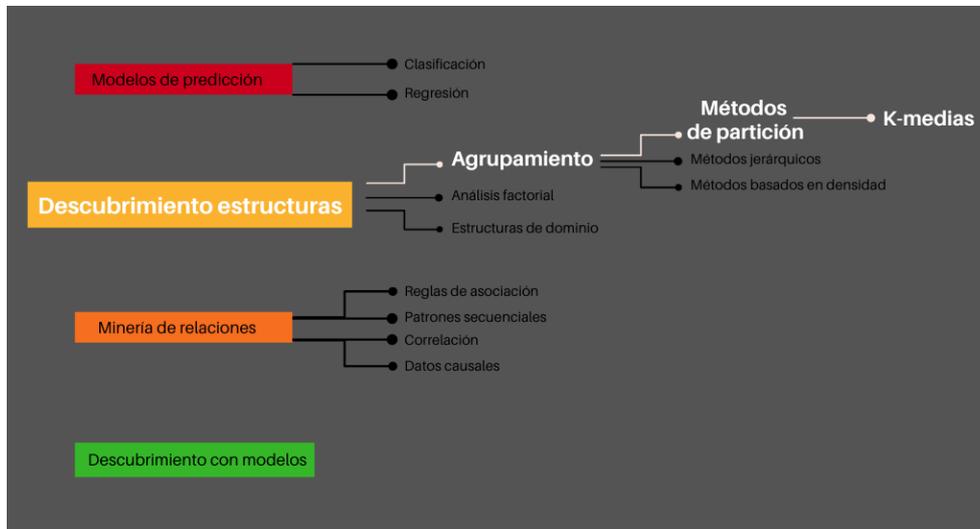
Una vez completados la recopilación y el pre-procesado de los datos, el siguiente paso es realizar una exploración de los datos mediante la aplicación de técnicas y algoritmos de LA, con el objetivo de identificar patrones recurrentes en los datos. Para ello, de entre las numerosas técnicas de LA descritas en el capítulo cuarto, se ha seleccionado el algoritmo k-medias, un algoritmo de descubrimiento de estructuras catalogado como una de las técnicas de la minería de datos. Como se explicó en dicho capítulo, las técnicas de minería de datos son el método más comúnmente utilizado para analizar e interpretar los datos de registro de los alumnos, ya que su objetivo es el descubrimiento de patrones a partir de datos educativos brutos e ininteligibles (Liu, 2006; Khalil y Ebner, 2016; Romero et al., 2016).

Un tipo de algoritmos dentro de las técnicas de minería de datos son los algoritmos de descubrimiento de estructuras, los cuales intentan encontrar una estructura en los datos sin partir de suposiciones previas (Baker e Inventado, 2014). Este tipo de algoritmos es el adecuado para el análisis de los datos obtenidos a partir del LMOOC, ya que el objetivo de esta investigación es identificar el número de perfiles existente en el curso sin recurrir a un etiquetado previo de los datos, como sí se hace en otros algoritmos de minería de datos, por ejemplo, en los algoritmos de predicción. Dentro de los algoritmos de descubrimiento de estructuras, se ha seleccionado el algoritmo k-medias, un método de partición perteneciente al grupo de algoritmos de agrupamiento o *clustering*.

Como se analizó en el capítulo cuarto, los algoritmos de agrupamiento organizan los objetos de datos en grupos o clústeres, de modo que los objetos dentro de un clúster sean similares entre sí y a la vez diferentes de los objetos de otros clústeres. Dentro de los algoritmos de agrupamiento, los métodos de partición comienzan con una partición en clústeres aleatoria, la cual se intenta mejorar trasladando objetos de un grupo a otro mediante una técnica de reubicación iterativa. El algoritmo k-medias, en particular, está basado en un método heurístico en el que cada grupo está representado por el valor medio

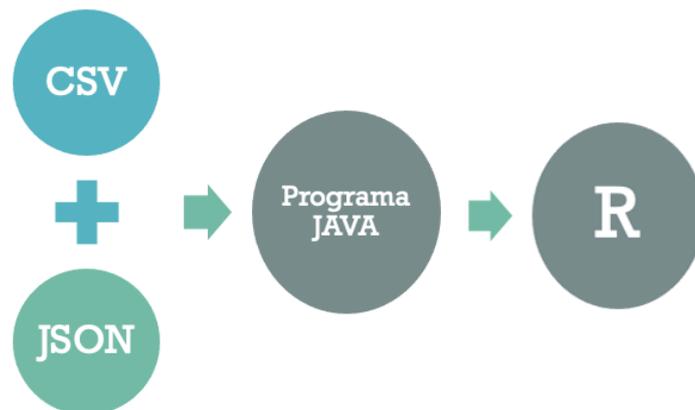
de los objetos pertenecientes al grupo (MacKay, 2005; Chatti et al., 2012; Baker e Inventado, 2014; Aldowah, 2019).

FIGURA 23. POSICIÓN JERARQUIZADA DEL ALGORITMO K-MEDIAS DENTRO DE LOS PRINCIPALES MÉTODOS DE LA MINERÍA DE DATOS. ELABORACIÓN PROPIA



La herramienta empleada para aplicar a los datos de salida del programa Java el algoritmo k-medias fue R. R es el nombre de, por un lado, un lenguaje de programación y, por otro, un entorno para la generación y el tratamiento de análisis y gráficos estadísticos. R se puede considerar como una implementación alternativa en código abierto de S, el lenguaje y el entorno que fueron desarrollados por John Chambers para *Bell Laboratories*. R es, por lo tanto, una plataforma estadística integral que ofrece todo tipo de técnicas de análisis de datos, entre ellas modelado lineal y no lineal; pruebas estadísticas clásicas; análisis de series temporales; clasificación, y agrupamiento. De manera adicional, R posee potentes funcionalidades gráficas que permiten la visualización de datos complejos con calidad de publicación, y con la inclusión de símbolos matemáticos y fórmulas. R también permite la importación y exportación de datos de una amplia variedad de fuentes, incluidos archivos de texto; sistemas de administración de bases de datos; paquetes estadísticos, y repositorios de datos especializados. R está disponible como *software* libre bajo los términos de la Licencia Pública General GNU de la *Free Software Foundation*. De este modo, existen versiones gratuitas de este *software* para una amplia gama de plataformas, incluidas Windows, Unix y Mac OS X (Kabacoff, 2011; R, s.f.).

FIGURA 24. FLUJO DE LOS DATOS DE ENTRADA DEL PROGRAMA R. ELABORACIÓN PROPIA.



Una de las ventajas de R es que permite la programación de nuevos métodos estadísticos de una manera fácil y directa, de tal modo que los usuarios pueden crear sus propios paquetes de funciones y hacerlos disponibles al resto de la comunidad de usuarios. Aunque el programa R cuenta con funciones nativas para calcular el algoritmo k-medias, se ha optado por la instalación de *factoextra*⁵⁹, un paquete estadístico externo que presenta varias ventajas. Por un lado, este paquete permite calcular el agrupamiento de particiones mediante la función *eclust()*, una función consistente en una sola línea que simplifica el flujo del análisis del agrupamiento. Por otro lado, al contrario que las funciones de partición estándar de R, las cuales requieren que el usuario especifique el número óptimo de clústeres, la función *eclust()* calcula automáticamente el número óptimo de agrupaciones para el algoritmo k-medias. Este paquete también incluye funcionalidades gráficas para representar los grupos y para presentar la información de silueta (Alboukadel, 2018).

Por consiguiente, para calcular el número óptimo de clústeres o agrupaciones mediante el algoritmo k-medias, se ha ejecutado el siguiente comando en R:

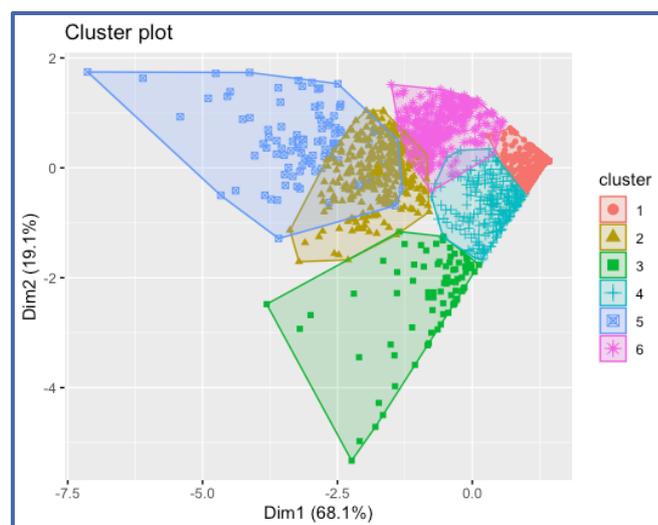
```
eclust(df, "kmeans", nstart = 25)
```

⁵⁹ <https://cran.r-project.org/web/packages/factoextra/index.html>

donde *df* es la tabla de salida del programa Java con las variables escaladas y *nstart* es un parámetro que indica el número de configuraciones iniciales con las que se inicia el algoritmo, siendo 25 el valor comúnmente recomendado (Alboukadel, 2018; Finnstats, 2021).

En el gráfico 1 se muestran las agrupaciones o clústeres óptimos calculados por el algoritmo k-medias, cuyo número, en el caso de los datos obtenidos del LMOOC PA, es de seis. En el gráfico se representan los valores de las tres variables de entrada para cada una de las muestras, es decir, para cada uno de los participantes en el LMOOC. De este modo, se dibujan con una combinación de color y de signos distintivos (círculos, triángulos, cuadrados, etc.) los elementos pertenecientes a cada uno de los grupos. Ya que el número de variables es de tres, el gráfico ofrece una reducción bidimensional de la distribución de los elementos que configuran cada grupo.

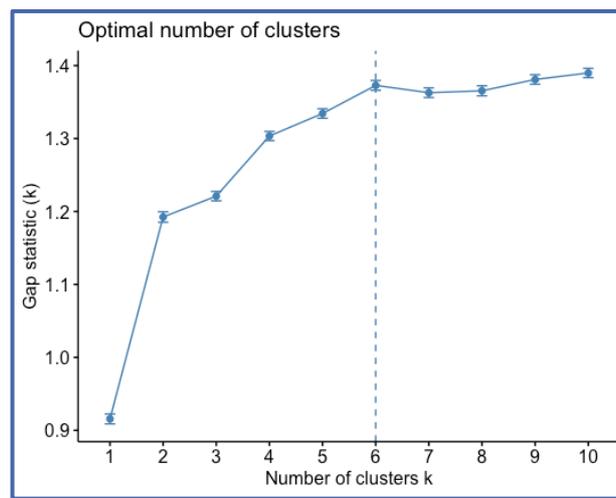
GRÁFICO 1. CLÚSTERES OBTENIDOS A PARTIR DE LA FUNCIÓN ECLUST() EN R PARA LOS DATOS DEL LMOOC PA.



Una vez ejecutado el algoritmo, existen varias funciones que permiten evaluar la calidad del número de agrupaciones obtenido. Una de ellas es *gap statistic*, un método que permite aproximar el número idóneo de grupos *k* en métodos de agrupamiento no supervisado. Para ello, se evalúa una métrica de error con respecto al número *k* obtenido, que en el caso de *gap statistic* es la suma de cuadrados dentro del grupo. De este modo, el error tenderá a disminuir constantemente a medida que aumenta el número de *k* (Tibshirani et al., 2001). El gráfico 2 presenta el resultado de ejecutar esta operación en R mediante la

función `fviz_gap_stat()` del paquete `factoextra`. En el eje vertical de este gráfico se representa el valor de la métrica de error, mientras que en el eje horizontal se representa el número de agrupamientos. Como puede observarse, el error va aumentando hasta llegar al valor seis del eje horizontal, valor a partir del cual desciende. Por lo tanto, puede concluirse que seis agrupamientos o clústeres es el valor óptimo para este conjunto de datos.

GRÁFICO 2. RESULTADO DE LA FUNCIÓN GAP STATISTIC EN R PARA LOS DATOS DEL LMOOC PA.



Otro método para evaluar la calidad del número de agrupaciones obtenido es el coeficiente de silueta, el cual se calcula de la siguiente manera (Fonseca, 2019):

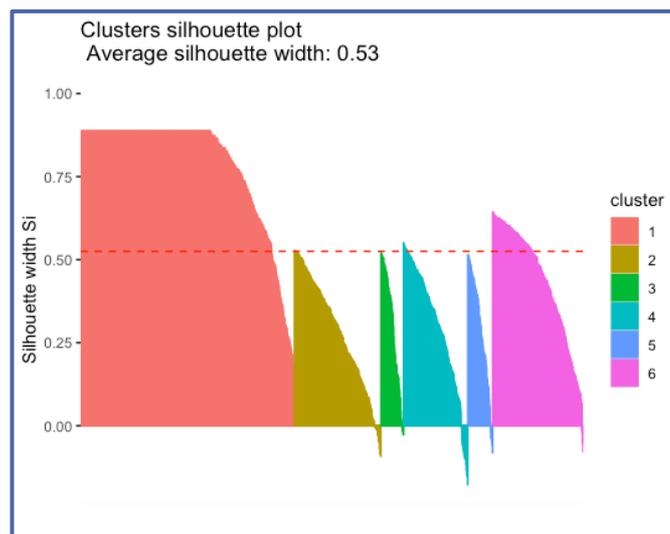
1. Para cada observación i , se calcula la disimilitud promedio (D_i) entre i y todos los demás puntos dentro del mismo grupo al que pertenece i .
2. Posteriormente, se realiza el mismo cálculo de disimilitud entre i y todos los demás grupos y se obtiene el valor más bajo entre ellos (C_i). Es decir, se encuentra la disimilitud entre i y el siguiente grupo más cercano a i .
3. El ancho de la silueta (S_i) es la diferencia entre C_i y D_i ($C_i - D_i$) dividida por el mayor de esos dos valores ($\max(D_i, C_i)$):

$$S_i = (C_i - D_i) / \max. (C_i, D_i)$$

4. Por último, el valor de S_i , o el ancho de la silueta, indica la idoneidad de la agrupación, de tal modo que:
- $S_i > 0$ significa que la observación está bien agrupada. De hecho, cuanto más cerca está de 1, más idónea es su ubicación en el clúster.
 - $S_i < 0$ significa que la observación se colocó en el grupo incorrecto.
 - $S_i = 0$ significa que la observación está entre dos grupos distintos.

Para el cálculo del coeficiente de silueta se ha utilizado la función *fviz_silhouette()* del paquete *factoextra*, cuyo resultado se muestra en el gráfico 3. Este gráfico agrupa los valores de coeficiente de silueta de cada uno de los elementos de un clúster y los representa en orden descendente, asignando a cada clúster un color distinto. De esta forma, los elementos más a la izquierda de cada agrupación son los que presentan un valor mayor del coeficiente de silueta, mientras que los valores más a la derecha son los que presentan un valor menor, llegando a ser negativo para un número ínfimo de elementos de los clústeres segundo a quinto.

GRÁFICO 3. RESULTADO DEL CÁLCULO DEL COEFICIENTE DE SILUETA EN R PARA LOS DATOS DEL LMOOC PA.



Un valor negativo indica que la muestra se ha agrupado en el clúster incorrecto, fenómeno que también puede observarse en el gráfico 1, en el que

puede apreciarse el solapamiento de varios de los clústeres. Sin embargo, como puede observarse en el gráfico 3, la gran mayoría de las observaciones presenta un valor superior a uno, lo cual indica que han sido agrupadas correctamente. Además, la media de todos los valores es de 0,53, lo que de nuevo apunta a que un resultado de seis agrupamientos para este conjunto de datos es un resultado correcto.

Sin embargo, este resultado de seis agrupaciones de participantes a partir de los datos del LMOOC PA difiere de los cinco grupos identificados en el LMOOC IP por del Peral (2019). Por lo tanto, se considera necesario implementar el siguiente paso dentro de la metodología de un proyecto de LA: el post-procesado de los datos.

6.4. POST-PROCESADO

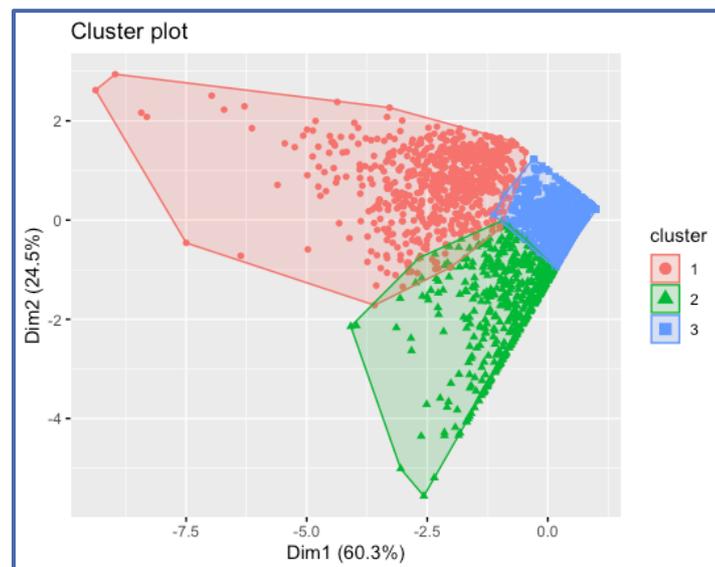
Dentro de un proyecto de LA, la fase de post-procesado se implementa en los casos en los que los resultados no son concluyentes debido, entre otros motivos, a deficiencias en la calidad de los datos o en la metodología. Esta fase puede requerir la compilación de nuevos datos a partir de fuentes adicionales; refinar el conjunto de datos; determinar nuevos atributos requeridos para la nueva iteración; identificar nuevos indicadores o métricas; modificar las variables de análisis, o elegir un nuevo método de análisis. Esto implica, por lo tanto, la modificación del diseño del estudio y la repetición del análisis (Chatti et al., 2012; Calvet Liñán y Juan Pérez, 2015).

En el caso del presente estudio, el tratamiento de los datos obtenidos a partir del LMOOC PA durante la fase de análisis ha sido más riguroso que el tratamiento de los datos del LMOOC IP a partir del cual se obtuvo la taxonomía original. Los grupos identificados en el trabajo de del Peral (2019) fueron determinados de antemano y posteriormente se utilizaron técnicas de LA para confirmar que, efectivamente, estos grupos estaban presentes en los datos. Pero esto no significa que el número óptimo de grupos fuese de cinco, por lo que, podría darse la circunstancia de que existan grupos adicionales que hayan sido pasados por alto. Por consiguiente, en la fase de post-procesado se ha procedido a repetir los dos primeros pasos del análisis, tal y como han sido

descritos en los apartados anteriores, a partir de los datos originales del LMOOC IP. Se han aplicado, por lo tanto, la fase de recopilación y pre-procesado de datos, y la fase de análisis mediante el algoritmo k-medias a los datos originales extraídos del LMOOC IP.

En el gráfico 4 se representan, esta vez para el LMOOC IP, los valores de las tres variables de entrada para cada uno de los participantes. De nuevo, el gráfico ofrece una reducción bidimensional de la distribución de los elementos que configuran cada grupo. El número de agrupaciones obtenidas como resultado del algoritmo k-medias para el LMOOC IP, como puede observarse en el gráfico, es de tres.

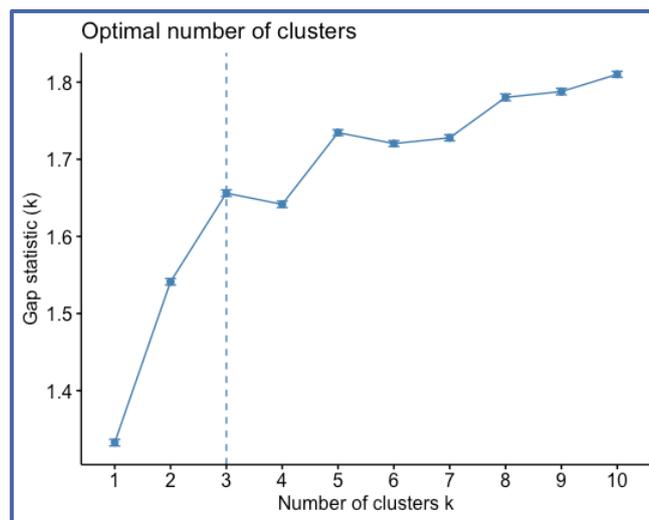
GRÁFICO 4. CLÚSTERES OBTENIDOS A PARTIR DE LA FUNCIÓN *ECLUST()* EN R PARA LOS DATOS DEL LMOOC IP.



Al ejecutar la función *gap statistic*, se puede apreciar cómo, aunque el número óptimo de clústeres es en principio de tres, cinco clústeres sería también una agrupación apropiada para el algoritmo k-medias. En el gráfico 5 se presenta el resultado de ejecutar esta operación en R mediante la función *fviz_gap_stat()* del paquete *factoextra*. Al igual que en el gráfico 2, en el eje vertical se representa el valor de la métrica de error, mientras que en el eje horizontal se representa el número de agrupamientos. Como puede observarse, el valor del error aumenta de una a dos agrupaciones y de dos a tres agrupaciones, para después descender de tres a cuatro clústeres. Este es el

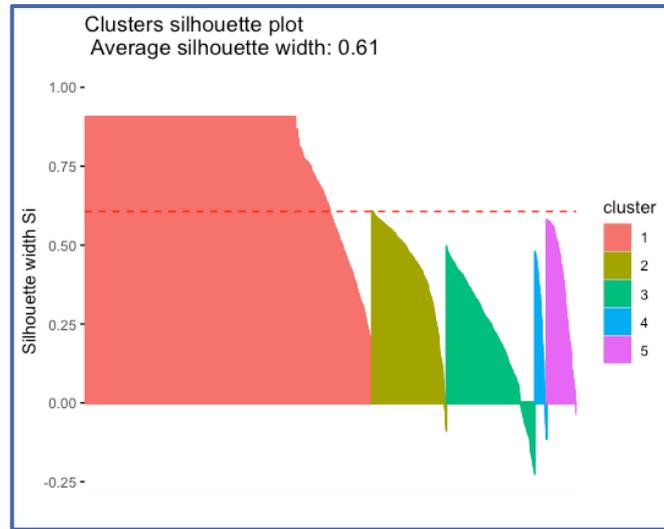
motivo por el cual la función determina que el resultado óptimo es el de tres grupos, ya que tres es el primer valor para el que se produce un descenso de la métrica del error. Sin embargo, como puede observarse en el gráfico, este valor de error vuelve a aumentar de cuatro a cinco agrupaciones y desciende de cinco a seis, por lo que puede concluirse que cinco agrupamientos o clústeres es también un valor posible y adecuado para este conjunto de datos.

GRÁFICO 5. RESULTADO DE LA FUNCIÓN GAP STATISTIC EN R PARA LOS DATOS DEL LMOOC IP.



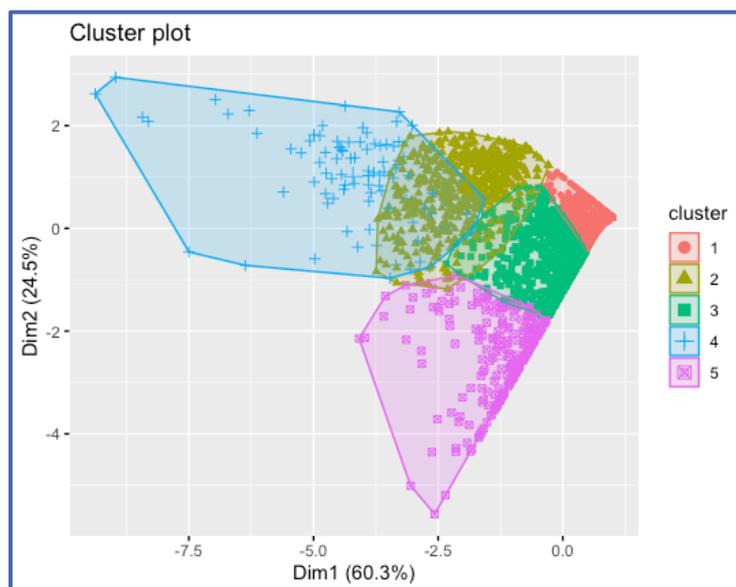
De hecho, al calcular el coeficiente de silueta para un valor de cinco clústeres mediante la función *fviz_silhouette()*, como se muestra en el gráfico 6, puede comprobarse que la gran mayoría de observaciones presentan un valor mayor a cero, lo cual indica que se han agrupado correctamente. Solo un pequeño grupo de muestras, particularmente en el grupo 3, presentan un valor negativo, lo cual indica que la muestra ha sido agrupada en el clúster incorrecto. También puede observarse que la media de los valores de todas las observaciones es de 0,61, un valor incluso mayor que el obtenido para las seis agrupaciones del LMOOC PA. Por lo tanto, puede concluirse que la calidad de los clústeres obtenidos es adecuada, por lo que podría considerarse que la existencia de cinco agrupaciones en el LMOOC IP para este conjunto de datos es un resultado correcto.

GRÁFICO 6. RESULTADO DEL CÁLCULO DEL COEFICIENTE DE SILUETA EN R CON UN VALOR DE CINCO CLÚSTERES PARA LOS DATOS DEL LMOOC IP.



A partir del algoritmo k-medias, la distribución de los participantes del LMOOC IP en cinco grupos sería la que se muestra en el gráfico 7, en el que se representan en dos ejes de coordenadas los elementos que configuran cada uno de estos cinco grupos.

GRÁFICO 7. CLÚSTERES OBTENIDOS A PARTIR DE LA FUNCIÓN *ECLUST()* CON VALOR 5 EN R PARA LOS DATOS DEL LMOOC IP.



En conclusión, en este capítulo se han descrito los pasos seguidos en la implementación de la metodología del presente estudio, los cuales se corresponden con las fases de un proyecto de LA. En primer lugar, se han descrito los dos LMOOC de los que se han recopilado los datos de la investigación, dos cursos con duración; características, y público objetivo muy distintos, y se han descrito una serie de operaciones que resultan fundamentales en la fase de pre-procesado de un proyecto de LA. Dentro de estas operaciones, la identificación de sesiones y la limpieza; integración; reducción, y modelado de los datos se han implementado mediante la creación de un programa en código Java, mientras que la transformación de los datos se ha realizado con el *software* estadístico R. Mediante este paquete estadístico se ha llevado a cabo también la segunda fase del proyecto, la fase de análisis y acción, en la que se ha aplicado el algoritmo k-medias a las tres variables obtenidas a partir del pre-procesado de los datos del LMOOC PA. Tras la ejecución del algoritmo, se ha determinado que el número óptimo de agrupaciones para dicho conjunto de datos es seis. Ya que este valor no coincide con el número de perfiles de del Peral, en la fase de post-procesado se han repetido los pasos previos, esta vez para los datos recopilados del LMOOC IP, y se ha considerado que, aunque el resultado inicial del algoritmo para las muestras de este segundo LMOOC es el de tres agrupaciones, puede concluirse que cinco perfiles, el mismo número que en la taxonomía inicial, es también un resultado adecuado.

Los resultados obtenidos a partir de la aplicación del algoritmo k-medias a ambos LMOOC van a estudiarse con más detalle en el siguiente capítulo. En este capítulo también se analizará cada una de las agrupaciones o clústeres determinados por el algoritmo, y se ofrecerá una explicación al hecho de que el número de perfiles resultantes para cada uno de los dos LMOOC no coincida.

CAPÍTULO 7: ANÁLISIS DE RESULTADOS

7.1. INTRODUCCIÓN

Como se estableció en el capítulo inicial, el objetivo general de esta investigación es, por un lado, la extrapolación de la taxonomía de perfiles de usuarios de LMOOC descrita por del Peral (2019) a un nuevo LMOOC y, por otro, el análisis de las posibles diferencias entre perfiles en función de las características distintivas de cada uno de los dos cursos descritos en el capítulo anterior. Para ello, tal y como también se ha descrito en el capítulo anterior, se ha seguido una metodología cuantitativa basada en la implementación de técnicas de LA, en cuyas fases de pre-procesado y análisis se ha aplicado el algoritmo k-medias a los datos generados por el LMOOC PA y, en la fase de post-procesado, también a los generados por el LMOOC IP.

Durante estas tres fases del proyecto de LA, se ha generado una serie de datos, los cuales serán analizados en el presente capítulo con el objetivo de comparar los perfiles temporales identificados en los LMOOC IP y PA. De manera adicional, se van a estudiar las características de los distintos perfiles con el fin de determinar si los rasgos propios de cada uno de los cursos y de sus participantes producen disparidades en el número y en la prevalencia de los grupos de usuarios, y en sus correspondientes niveles de desempeño.

Así pues, en la primera sección de este capítulo se procederá a analizar las agrupaciones identificadas tras la ejecución del algoritmo k-medias a partir de los datos de ambos LMOOC y a explicar el motivo de la discrepancia en el número de clústeres mediante la comparación de los grupos obtenidos. Finalmente, las distintas agrupaciones o perfiles de participantes se irán analizando en cada una de las cuatro secciones subsiguientes, en las que también se describirán las diferencias que pudieran emerger de las características particulares de cada uno de los perfiles identificados para los LMOOC PA e IP.

7.2. IDENTIFICACIÓN DEL NÚMERO DE GRUPOS

Como ya se adelantó en el capítulo anterior, el número de clústeres o perfiles obtenidos como resultado de la aplicación del algoritmo k-medias a las

variables extraídas de los datos generados en el LMOOC PA fue de seis. Esta cifra no coincide ni con las cinco agrupaciones obtenidas como resultado de la aplicación del algoritmo k-medias a los datos generados en el LMOOC IP ni con los cinco perfiles identificados en el estudio de del Peral (2019), los cuales, como se detalló en el capítulo cuarto son:

- **Pre-consumidores:** usuarios que apenas acceden al curso y por lo tanto obtienen las notas más bajas.
- **Intensivos:** usuarios con un número muy reducido de sesiones.
- **Rutinarios:** usuarios con una distribución del tiempo que tiende a concentrarse en torno a los mismos momentos del día.
- **Exhaustivos:** participantes que presentan un elevado número de sesiones en el curso.
- **Flexibles:** usuarios que presentan, para un número bastante elevado de sesiones, la mayor variación en cuanto a las horas de conexión.

Aunque es evidente que para el LMOOC PA existe al menos un perfil adicional, es necesario analizar para ambos cursos las características de los clústeres descubiertos mediante el algoritmo k-medias. De este modo se intentará establecer, por un lado, una correlación entre los perfiles obtenidos como resultado de la ejecución del algoritmo en ambos LMOOC y, por otro lado, una correlación de estos mismos perfiles con la taxonomía de del Peral. Esta correlación es necesaria, ya que existe la posibilidad de que alguno de los perfiles identificados no sea compartido por ambos LMOOC y/o no se corresponda con ninguno de los perfiles de la taxonomía de partida. Por este motivo, en los siguientes subapartados se procederá a analizar, primero para el curso PA y luego para IP, las características de las distintas agrupaciones, las cuales se intentarán correlacionar con los perfiles descritos por del Peral. Posteriormente, se analizarán las diferencias resultantes del estudio de todos los perfiles y de su comparación con la taxonomía original.

7.2.1. AGRUPACIONES PARA EL LMOOC PA

El LMOOC PA contó con un total de 1776 participantes inscritos que interactuaron con la plataforma al menos una vez durante la duración del curso. Como se explicó en el capítulo anterior, a partir de las variables generadas

mediante el tratamiento de los datos registrados en la plataforma para cada uno de estos usuarios, el algoritmo k-medias recurre al valor medio de las muestras para efectuar una organización de datos en clústeres, de tal modo que los objetos dentro de un clúster sean similares entre sí y a la vez diferentes de los objetos de otros clústeres (Chatti et al., 2012; Baker e Inventado, 2014; MacKay, 2015; Aldowah, 2019).

Al calcular la agrupación óptima de la muestra, la función *eclust()* del programa R permite visualizar como resultado de salida de la operación los valores para cada una de las variables del elemento central del clúster, esto es, el elemento alrededor del cual se distribuyen el resto de los elementos de la agrupación. En la tabla 10 se representan, para el LMOOC PA, cuáles son los valores escalados que presentan los elementos centrales de cada clúster para las tres variables analizadas. Adicionalmente, se incluye el tamaño poblacional de cada uno de los grupos.

TABLA 10. TAMAÑO DE CADA GRUPO Y VALORES MEDIOS DE LAS VARIABLES PARA LOS DATOS DEL LMOOC PA.

PUERTAS ABIERTAS				
CLÚSTER	TAMAÑO	NOTA	DISPERSIÓN INICIO SESIONES	NÚMERO SESIONES
1	756	-0,84	-0,81	-0,62
2	305	1,21	1,04	0,80
3	80	-0,59	2,35	-0,26
4	227	-0,66	0,33	-0,24
5	88	1,08	1,12	3,21
6	320	1,14	0,20	0,05

Como puede observarse a partir de los datos de la tabla, el primer grupo dentro del LMOOC PA, el más numeroso, está formado por usuarios con un número reducido de sesiones; notas medias muy bajas, y baja dispersión en las horas de inicio de las sesiones de estudio, lo cual se corresponde claramente con el perfil de **pre-consumidores**: usuarios que apenas tienen interacción con los contenidos principales del curso.

Los participantes del segundo grupo presentan, para un número relativamente elevado de sesiones, una variación elevada en cuanto a las horas de conexión, además de las notas más altas en el curso. Este grupo parece corresponderse con el perfil de participantes **flexibles**.

Los grupos tercero y cuarto parecen presentar ciertas características comunes, ya que, en ambos grupos, tanto el número de sesiones como las notas presentan valores notablemente bajos. El grupo de la taxonomía de del Peral que se caracteriza por un número de sesiones reducido es el perfil de participantes intensivos. Dentro de los participantes intensivos se incluían a todos aquellos usuarios que dedicaban una o dos sesiones al curso, independientemente de la duración de las sesiones o del progreso dentro del curso. De hecho, en el estudio realizado por del Peral, solo 52 participantes intensivos, el 1,3% de todos los participantes en el curso, lo completaban con éxito y eran elegibles para el certificado final. De ello se puede deducir que el grupo inicialmente denominado como intensivo está formado mayoritariamente por estudiantes con poca vinculación con el LMOOC, los cuales se conectan una o dos veces al curso y posteriormente lo abandonan. En consecuencia, sería más preciso denominar el grupo de participantes intensivos como participantes **ocasionales**.

Dentro de esta categoría de participantes **ocasionales**, pueden distinguirse dos perfiles distintos, los correspondientes a los grupos tercero y cuarto obtenidos como resultado de la ejecución del algoritmo k-medias con los datos del LMOOC PA. El primero de ellos, además de valores bajos tanto en las notas como en el número de sesiones, presenta una gran dispersión en las horas de inicio de las sesiones, como bien puede observarse en la tabla 10, mientras que el segundo de ellos presenta valores más reducidos de esta variable. Estos dos grupos se analizarán con más detalle en el apartado 7.2.3.

El quinto grupo está formado por los usuarios con el mayor número de sesiones; con una alta variación en las horas de conexión, y con notas altas en el curso, lo cual coincide claramente con el perfil de participantes **exhaustivos**.

Por último, el sexto grupo muestra, para un número significativo de sesiones, una tendencia a concentrar las horas de estudio en las mismas franjas horarias, lo cual se correspondería con el perfil de participantes **rutinarios**.

7.2.2. AGRUPACIONES PARA EL LMOOC IP

El LMOOC IP contó con un total de 3859 participantes inscritos que, al interactuar con alguno de los elementos del curso, generaron una serie de datos a partir de los cuales se extrajeron las tres variables empleadas por el algoritmo k-medias para el cálculo de los distintos agrupamientos. Al igual que en el apartado anterior, en la tabla 11 se representan, esta vez para el LMOOC IP, cuáles son los valores escalados de los elementos centrales de cada uno de los clústeres.

TABLA 11. TAMAÑO DE CADA GRUPO Y VALORES MEDIOS DE LAS VARIABLES PARA LOS DATOS DEL LMOOC IP.

INGLÉS PROFESIONAL				
CLÚSTER	TAMAÑO	NOTA	DISPERSIÓN INICIO SESIONES	NÚMERO SESIONES
1	2253	-0,50	-0,62	-0,44
2	586	2,08	0,45	0,76
3	694	-0,25	0,57	0,17
4	93	1,54	0,92	4,51
5	233	-0,23	2,74	0,6

Como puede observarse en la tabla, el primer grupo, que es de nuevo el más numeroso, está formado por usuarios con un número reducido de sesiones; notas medias muy bajas, y baja dispersión en las horas de inicio de las sesiones de estudio. Este grupo se corresponde, al igual que en el LMOOC PA, con el perfil de **pre-consumidores**, usuarios que apenas tienen interacción con los contenidos principales del curso.

Los participantes del segundo grupo presentan, para un número relativamente elevado de sesiones, una tendencia a concentrar el estudio en las mismas franjas horarias, además de las notas más altas en el curso. Este grupo se correspondería, por lo tanto, con el perfil de participantes **rutinarios**.

Los grupos 3 y 5 de IP muestran similitudes con los grupos 3 y 4 del LMOOC PA, analizados en el subapartado anterior y denominados participantes **ocasionales**, ya que, en ambos grupos, tanto el número de sesiones como las notas presentan valores notablemente bajos. De nuevo, lo que diferencia a ambos es que el grupo 5 presenta una mayor dispersión de las horas de inicio de las sesiones, mientras que el grupo 3 exhibe valores más reducidos para esta variable.

Por lo tanto, dentro de esta categoría de participantes **ocasionales**, pueden distinguirse dos perfiles distintos. El primero de ellos, que se correspondería con el grupo 3 de PA y el grupo 5 de IP, además de valores bajos tanto en las notas como en el número de sesiones, presenta una gran dispersión en las horas de inicio de las sesiones, como bien puede observarse en la tabla 11. El segundo, que se correspondería con el grupo 4 de PA y el grupo 3 de IP, presenta valores más reducidos de esta variable. Estos dos grupos se analizarán con más detalle en el apartado 7.2.3.

Por último, en el cuarto grupo se encuentran los usuarios con mayor número de sesiones; una alta variación en las horas de conexión, y notas elevadas en el curso, lo cual indica claramente una correspondencia con el perfil de participantes **exhaustivos**.

En el siguiente subapartado se procederá a comparar y analizar las distintas agrupaciones identificadas mediante el algoritmo k-medias para ambos LMOOC.

7.2.3. COMPARACIÓN DE LAS AGRUPACIONES PARA LOS LMOOC IP Y PA

En el capítulo anterior se analizó la estructura y las características de los dos LMOOC cuyos datos son la base de este estudio, PA e IP. Por consiguiente, pudo apreciarse claramente cómo, a pesar de que ambos son MOOC de lenguas, tanto el público objetivo como algunos de sus rasgos son diferentes. Las tablas 12 y 13 muestran, para el conjunto íntegro de participantes de cada LMOOC, la media de las tres variables analizadas mediante el algoritmo k-medias: nota media final del curso; diferencia media del tiempo de conexión

entre sesiones, y número de sesiones, además del tiempo total invertido en el curso y el número de accesos a los foros.

TABLA 12. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC PA.

	Tiempo total (min.)	Nota media	Dispersión inicio sesiones (min.)	Número sesiones	Accesos al foro
MEDIA	189,74	0,40	92,87	7,81	17,21
DESV. TÍPICA	251,30	0,44	103,07	10,15	63,24

TABLA 13. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC IP.

	Tiempo total (min.)	Nota media	Dispersión inicio sesiones (min.)	Número sesiones	Accesos al foro
MEDIA	68,02	0,18	68,24	3,39	4,02
DESV. TÍPICA	109,87	0,33	97,61	4,62	12,99

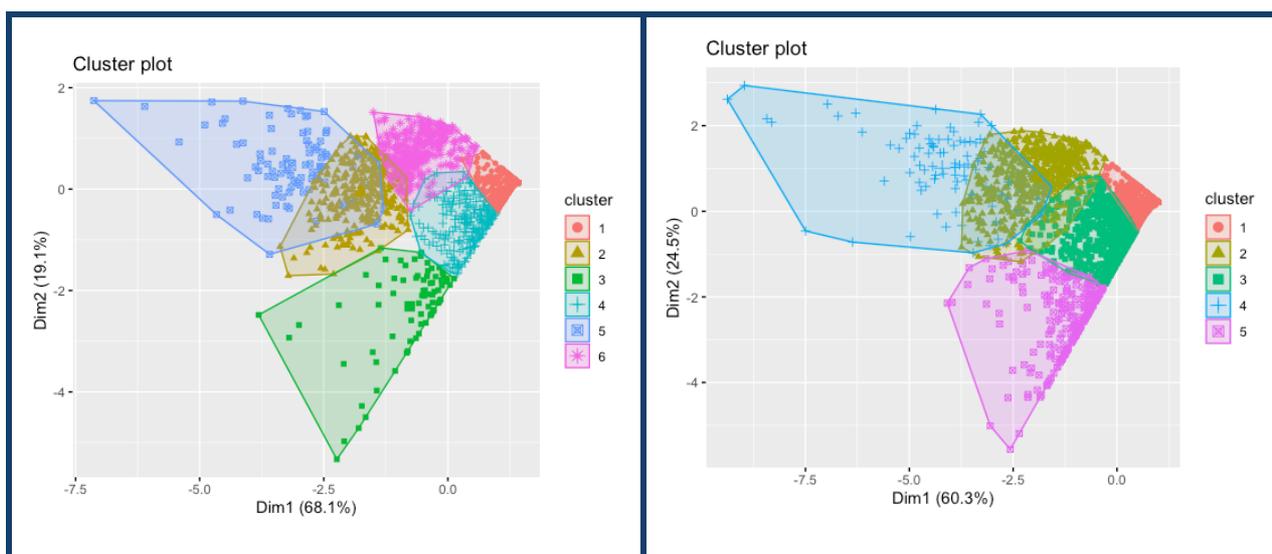
Como puede observarse al comparar las dos tablas, los datos generales del LMOOC PA son mejores que los de IP. Los participantes dedicaron más tiempo de media, emplearon un número mayor de sesiones y accedieron a los foros en mayor proporción que los participantes del curso IP. Esto tiene su reflejo en una nota media superior para PA y en una mayor dispersión de las sesiones, lo cual es de esperar al ser el número de sesiones mayor en PA que en IP.

Estas diferencias pueden deberse, en primer lugar, a la longitud de cada curso. Recordemos que el LMOOC IP estaba dividido en tres módulos distintos y tuvo una duración de tres semanas, mientras que PA contaba con cuatro módulos y una duración de seis semanas. En segundo lugar, el perfil de los participantes de cada curso exhibía notables diferencias. IP estaba dirigido y diseñado, en general, para trabajadores con un nivel B1 de inglés y, en particular, para estudiantes del grado de turismo de la UNED. Los individuos pertenecientes a estos perfiles se caracterizan por tener un tiempo limitado para dedicar al MOOC, ya que deben compatibilizarlo con su empleo y/o con sus estudios. Como resultado, la posibilidad y la necesidad de profundizar en los contenidos del curso son limitadas para este perfil de participantes. Por el contrario, PA estaba diseñado para migrantes, un perfil de estudiantes con una necesidad más acusada de asimilar los contenidos del curso, y con una menor

tasa de ocupación, lo cual facilita la dedicación de más tiempo a la interacción con los contenidos del LMOOC. Todo ello explicaría un mayor promedio de sesiones y de tiempo total para el curso PA.

A pesar de las diferencias existentes en ambos LMOOC, en los subapartados anteriores se ha detectado una coincidencia entre varios de los grupos identificados por el algoritmo k-medias, tanto para PA como para IP, y entre dichos grupos y los de la taxonomía inicial de del Peral (2019). Esta coincidencia puede observarse claramente en el gráfico 8, en el que se yuxtaponen los gráficos que representan los clústeres resultantes de la ejecución del algoritmo k-medias para ambos LMOOC y que ya se analizaron en el capítulo anterior.

GRÁFICO 8. YUXTAPOSICIÓN DE LOS GRÁFICOS QUE REPRESENTAN LOS CLÚSTERES OBTENIDOS A PARTIR DE LA FUNCIÓN `ECLUSO()` EN R PARA LOS DATOS DE LOS LMOOC PA (IZQUIERDA) E IP (DERECHA).



En este gráfico se puede apreciar una clara coincidencia entre los grupos 1 de los dos cursos (en rojo en ambos gráficos), que se corresponderían con el perfil de los **pre-consumidores**, participantes que presentan valores muy bajos para las tres variables: la nota, el número de sesiones y la concentración de las horas de estudio⁶⁰.

⁶⁰ Es necesario aclarar que, si un usuario solo establece una sesión, el valor de dispersión de las horas de conexión será de 0. Por lo tanto, los perfiles que contengan un gran número de usuarios con una sola sesión en los cursos presentarán valores bajos en esta variable.

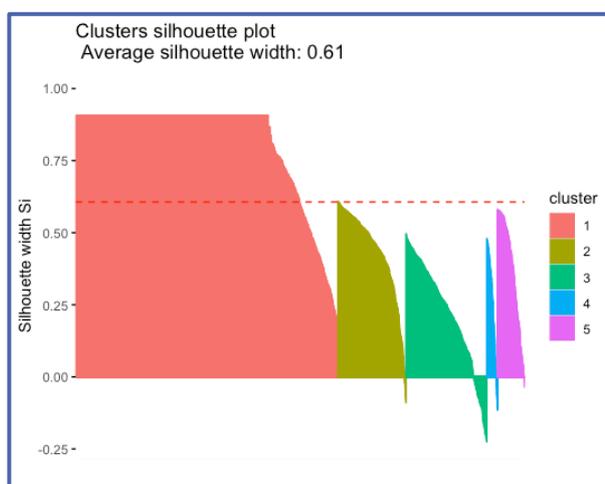
Por otro lado, el grupo 5 del LMOOC PA (en azul oscuro en el gráfico de la izquierda) y el grupo 4 del LMOOC IP (en azul claro en el gráfico de la derecha) también presentan una clara coincidencia, con una forma geométrica; un tamaño; una posición, y una distribución de las muestras muy similares. Ambos grupos coinciden a su vez con el perfil de participantes **exhaustivos**: usuarios con el mayor número de sesiones, con una alta variación en las horas de conexión y con notas altas en el curso.

También se aprecia una gran similitud en la forma; el tamaño; la posición, y la distribución del grupo 3 del LMOOC PA (en verde en el gráfico de la izquierda) y del grupo 5 del LMOOC IP (en morado en el gráfico de la derecha). La característica común de ambos grupos es una gran variación con respecto al inicio de las horas de estudio, y valores bajos de las notas y el número de sesiones. Este agrupamiento corresponde, por lo tanto, con uno de los dos perfiles que en los apartados anteriores se denominaron participantes ocasionales. Estos usuarios serían, más específicamente, participantes **ocasionales flexibles**.

Valores bajos en notas y sesiones es también el rasgo que caracteriza a los grupos adyacentes al de los participantes ocasionales flexibles: el grupo 4 del LMOOC PA (en turquesa en el gráfico de la izquierda) y el grupo 3 del LMOOC IP (en verde en el gráfico de la derecha), en los cuales también se aprecian similitudes en la posición y la distribución. Aunque, como puede observarse en el gráfico, el clúster de IP se expande algo más hacia la esquina izquierda, es precisamente en este extremo superior izquierdo del clúster donde se produce un solapamiento con el grupo 2 (en dorado). Esto apunta a la posibilidad de que esas muestras hayan sido agrupadas por el algoritmo de forma incorrecta. De hecho, como se muestra en el gráfico 9⁶¹, es precisamente el grupo 3 el que presenta un mayor número de muestras con un coeficiente de silueta negativo, lo cual indica que han sido agrupadas por el algoritmo en el clúster equivocado.

⁶¹ Este gráfico ya fue analizado en el capítulo 6.

GRÁFICO 9. RESULTADO DEL CÁLCULO DEL COEFICIENTE DE SILUETA EN R PARA LOS DATOS DEL LMOOC IP.



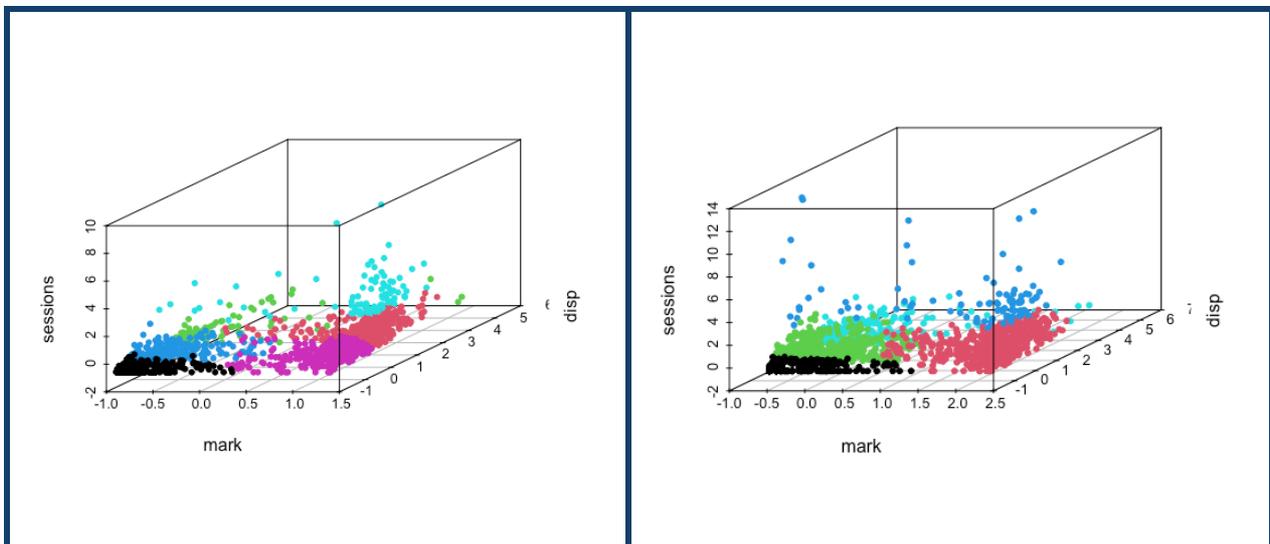
Por consiguiente, si se integraran los elementos solapados de la gráfica dentro del grupo 2, se establecería una coincidencia clara entre el grupo 4 del LMOOC PA y el grupo 3 del LMOOC IP. Estos dos grupos se caracterizan por presentar valores bajos en notas y sesiones, y poca variación con respecto al inicio de las horas de estudio. Se corresponderían, por lo tanto, con el segundo de los dos perfiles que forman el grupo de participantes ocasionales. Estos usuarios serían, más específicamente, participantes **ocasionales rutinarios**.

En resumen, los dos últimos pares de clústeres analizados (grupos 3 y 4 de PA, y grupos 5 y 3 de IP), los cuales se caracterizan por notas bajas y un reducido número de sesiones, forman parte de un supergrupo denominado participantes **ocasionales**, que son aquellos que presentan una baja vinculación con el curso. Este grupo de participantes ocasionales estaría, a su vez, dividido en dos subgrupos, uno con horas de conexión más próximas entre sí (grupo 4 de PA y 3 de IP), denominado **ocasionales rutinarios**, y otro con horarios de conexión más dispares (grupo 3 de PA y 5 de IP), denominado **ocasionales flexibles**.

Por último, es posible apreciar que, para PA, el algoritmo ha dividido en dos grupos, los grupos 2 (en dorado en el gráfico de la izquierda) y 6 (en rosa en el gráfico de la izquierda), las muestras que para IP han sido agrupadas en un solo grupo, el grupo 2 (en dorado en el gráfico de la derecha). Este fenómeno puede identificarse más claramente en el gráfico 10, en el que se ofrece una representación tridimensional de los valores de las variables de ambos cursos.

Como puede observarse en este gráfico, existe una correspondencia muy clara en la forma; el tamaño, y la distribución de todos los clústeres identificados por el algoritmo k-medias para ambos LMOOC, tal y como se acaba de analizar a partir del gráfico 8. La única diferencia sustancial entre los clústeres del gráfico 10 reside en el grupo en rojo del gráfico de la derecha (grupo 2 del LMOOC IP), el cual no tiene correspondencia unívoca en el gráfico de la izquierda, sino que se aprecia claramente dividido en dos subgrupos, rosa y rojo (grupos 2 y 6 del LMOOC PA), con una distribución muestral y forma muy similares.

GRÁFICO 10. REPRESENTACIÓN TRIDIMENSIONAL DE LOS VALORES DE LAS MUESTRAS DE LOS LMOOC PA (IZQUIERDA) E IP (DERECHA).



En consecuencia, en vista de la distribución de los valores muestrales observada en las gráficas 8 y 10, es correcto concluir que los grupos 2 y 6 del LMOOC PA pueden considerarse como dos subgrupos distintos del grupo 2 del LMOOC IP. Este supergrupo, al que se denominará participantes **constantes**, estaría formado por participantes con un alto nivel de desempeño: obtienen notas altas y dedican al curso un número elevado de sesiones. La diferencia entre los dos subgrupos que lo forman es la manera de organizar estas sesiones. Por un lado, uno de los subgrupos (grupo 6 de PA) tiende a concentrar las sesiones de estudio en los mismos horarios, lo cual se correspondería con el perfil de participantes **rutinarios** de del Peral, mientras que el segundo subgrupo (grupo 2 de PA) manifiesta una mayor dispersión en las horas de inicio de las sesiones de estudio, lo cual coincide con el perfil de participantes **flexibles**. Por

conseguido, estos dos subgrupos se denominarán participantes **constantes rutinarios** y **constantes flexibles**, respectivamente.

En las secciones restantes de este capítulo se analizarán más detalladamente cada uno de los perfiles identificados.

7.3. GRUPO 1: PRE-CONSUMIDORES

Del Peral (2019) define la categoría de pre-consumidores como aquellos usuarios que nunca acceden a los contenidos principales del MOOC. En el caso del LMOOC IP, para el que se acuñó inicialmente la taxonomía, los pre-consumidores fueron participantes que no accedieron a ninguno de los tres módulos en que estuvo dividido el curso. El tiempo que estos usuarios dedican al MOOC es mínimo, por lo que no aportan información sustancial para el trazado de perfiles temporales de uso del MOOC. Dentro de esta categoría han sido agrupados el 42,6% de los participantes del curso PA y el 58,4% del curso IP.

En las tablas 14 y 15 se muestran, para el grupo de pre-consumidores de los LMOOC PA e IP, respectivamente, los valores medios de las variables utilizadas por el algoritmo k-medias: nota; dispersión de las sesiones, y número de sesiones, además de otras dos variables que correlacionan con el número de sesiones: el tiempo total en el curso y el número de accesos a los foros.

Como puede observarse en ambas tablas, en el grupo de pre-consumidores los valores de todas las variables son muy bajos. El número medio de sesiones, variable cuyo valor mínimo posible es de 1, es de 1,53 para PA, y 1,34 para IP, con un tiempo total invertido en el curso de 24,01 minutos y 16,3 respectivamente. Respecto a las notas medias finales, en ambos casos son cercanas a 0, con 0,03 para el caso de PA y 0,02 para IP. Respecto a la tendencia a la variación de las horas de estudio, los valores son también muy bajos, de unos 9 minutos para PA y unos 8 para IP. Esto es esperable, ya que una gran parte de las muestras cuentan con una sola sesión y ello hace que el valor de dispersión de las horas de conexión sea de 0. Por último, el número de accesos a los foros presenta también tasas muy bajas, con 1,22 accesos para PA y 0,73 para IP.

TABLA 14. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC PA DEL GRUPO DE PRE-CONSUMIDORES.

	Tiempo total (min.)	Nota media	Dispersión inicio sesiones (min.)	Número sesiones	Accesos al foro
MEDIA	24,01	0,03	8,98	1,53	1,22
DESV. TÍPICA	42,59	0,08	17,46	1,18	4,65

TABLA 15. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC IP DEL GRUPO DE PRE-CONSUMIDORES.

	Tiempo total (min.)	Nota media	Dispersión inicio sesiones (min.)	Número sesiones	Accesos al foro
MEDIA	16,30	0,02	8,05	1,34	0,73
DESV. TÍPICA	25,86	0,07	17,31	0,72	2,89

El valor tan bajo en todas las variables indica que una gran parte de los integrantes de este grupo son usuarios con un nivel de participación simbólico. Sin embargo, para algunas variables de ambos cursos, tales como el tiempo total en el curso; la nota media, o los accesos a los foros, la desviación típica, como se observa en las tablas 14 y 15, presenta un valor relativamente alto, con valores que llegan a cuadruplicar el valor de la media en el caso del acceso a los foros. Esto apunta a la existencia, dentro de esta agrupación, de un subgrupo de participantes cuya interacción con el LMOOC no se limita a un vistazo general al contenido del curso, sino que parece que comienzan con la intención de completarlo. La prueba de ello es que llegan a realizar algunos de los cuestionarios y a acceder a los foros.

La tabla 16 muestra el porcentaje de alumnos en ambos LMOOC que presentan el valor mínimo posible en algunas de las variables analizadas. De este modo, 3 de cada 4 participantes del curso PA y 9 de cada 10 de los de IP tienen una nota final en el curso de 0, bien porque no han completado ningún cuestionario, bien porque han fallado todas las preguntas de los cuestionarios realizados. Además, el 25% de los participantes del curso PA y el 29% de los de IP tienen una dedicación en el curso de 0 minutos, lo cual significa que solo han efectuado una interacción con la plataforma (un solo acceso a una de las páginas del curso). Puede observarse también que los porcentajes de usuarios con una sola sesión y los que no han accedido a los foros es muy elevado. Como consecuencia de estos datos, ninguno de los participantes de IP y solo 3 de PA

son elegibles para la obtención del certificado. Sin embargo, aunque los valores de estas variables son bajos, no son absolutos. En ambos casos hay participantes que obtienen puntuación en los cuestionarios, que dedican tiempo y sesiones al curso y que acceden a los foros. Hay, como se ha mencionado, incluso participantes de PA que aprueban el curso.

TABLA 16. ESTADÍSTICAS RELATIVAS A INTERACCIÓN MÍNIMA EN LOS CURSOS PARA LOS LMOOC PA E IP

	Nota = 0	Tiempo = 0	Sesiones = 1	Foros = 0	Certificado = no elegible
PA	74,9%	25,1%	74,3%	77%	99,6%
IP	91%	29,3%	68,3%	82,7%	100%

Como se ha explicado anteriormente, el algoritmo k-medias que se ha empleado para clasificar las muestras en grupos está basado en un método heurístico en el que cada grupo está representado por el valor medio de los objetos pertenecientes al grupo (Baker e Inventado, 2014; MacKay, 2015; Aldowah, 2019). Esto implica que los elementos más extrínsecos de la agrupación tendrán valores más desemejantes a aquellos más cercanos al centro. Como se estableció en el capítulo inicial, uno de los principales problemas de los MOOC es las elevadas tasas de abandono, y de participantes que se inscriben y no realizan el curso. Este grupo de participantes es muy numeroso, por lo que tiene un gran peso en el cálculo de los distintos clústeres a partir de las medias. Esto provoca que se agrupen en el clúster de pre-consumidores participantes que tienen una actividad cercana a cero junto con participantes que tienen un nivel de actividad bajo en el curso.

Por lo tanto, el perfil de participantes pertenecientes a este grupo estaría formado por:

1. Usuarios que presentan una actividad cercana a cero en el curso, los cuales del Peral definió como pre-consumidores.
2. Usuarios que comienzan el curso, pero que en una etapa temprana lo abandonan.
3. Usuarios que realizan el curso, pero cuyo desempeño es muy bajo.

7.4. GRUPO 2: PARTICIPANTES EXHAUSTIVOS

Del Peral (2019) define a los participantes exhaustivos como aquellos participantes que dedican más de quince sesiones al curso. Son usuarios que característicamente presentan un número elevado de acceso a los foros y una proporción de tiempo en cada sesión más baja que la media. Como se ha explicado con anterioridad, para la clasificación en perfiles de este trabajo no se han utilizado valores límites en el número de sesiones, sino que se ha recurrido al algoritmo k-medias, el cual ha agrupado dentro de una categoría que puede asemejarse al perfil de usuarios exhaustivos descrito por del Peral al 4,9% de los participantes del curso PA y el 2,4% del curso IP.

En las tablas 17 y 18 se muestran de nuevo, para PA e IP respectivamente, los valores medios de las variables utilizadas por el algoritmo k-medias, junto con el tiempo total en el curso; el número de accesos a los foros, y el porcentaje de alumnos elegibles para el certificado final del curso. El rasgo característico de este grupo es el elevado número de sesiones y, por lo tanto, de tiempo dedicado al curso, lo cual supone una media de unos 800 minutos divididos en 40 sesiones para PA y de 391 minutos en 24 sesiones para IP.

La dispersión horaria de las sesiones también es elevada para ambos cursos: 209 minutos para PA y 158 para IP, lo cual es esperable dado el número tan elevado de sesiones. Además, tanto las notas medias finales (0,87 para PA y 0,78 para IP) como la elegibilidad para el certificado final (90% para PA y 69% para IP) son elevadas, algo también predecible debido al mayor tiempo dedicado al curso por parte de los participantes de este grupo.

Por último, el acceso a los foros también presenta tasas elevadas, como describió del Peral para el grupo de usuarios exhaustivos, donde cabe destacar la media del curso PA, con 105 accesos a los foros.

TABLA 17. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC PA DE LOS PARTICIPANTES EXHAUSTIVOS.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	805,13	0,87	208,7	40,42	104,97	89,8%
DESV. TÍPICA	398,68	0,25	65	13,06	214,28	-

TABLA 18. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC IP DE LOS PARTICIPANTES EXHAUSTIVOS.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	390,71	0,78	158,1	24,26	23,81	68,8%
DESV. TÍPICA	223,19	0,60	69,9	10,58	32,14	-

La conclusión es que las características de ambos grupos, tanto el de PA como el de IP, coinciden claramente con el perfil de participantes exhaustivos descrito por del Peral: usuarios con el mayor número de sesiones y de accesos a los foros en los MOOC, además de notas altas en el curso, y con una alta variación en las horas de conexión. Sin embargo, se aprecia una notable diferencia entre los datos relativos al LMOOC PA y a IP, con medias de dedicación más elevadas para el primero. Estas diferencias pueden deberse, como ya se ha apuntado en el apartado 7.2.3., a una mayor longitud del curso PA y al perfil de estudiantes de este LMOOC, compuesto por alumnado con necesidades formativas más inmediatas y, *a priori*, mayor disponibilidad horaria.

7.5. GRUPO 3: PARTICIPANTES OCASIONALES

Como se ha mencionado anteriormente, el perfil de del Peral (2019) más próximo al supergrupo de participantes ocasionales es el de los usuarios intensivos. Del Peral define a los usuarios intensivos como aquellos participantes que acceden a los contenidos principales del curso en el transcurso de una o dos sesiones. Son usuarios que característicamente dedican una menor cantidad de tiempo al curso e interactúan poco con los foros, pero que presentan una proporción elevada de minutos por sesión de estudio.

Como se ha explicado en el apartado 7.2.1., si se toman los datos originales del LMOOC IP que sirvieron de base para la clasificación de del Peral, realmente son pocos los participantes pertenecientes al grupo de los usuarios intensivos que terminan el curso. Por ese motivo, un nombre más apropiado para este grupo sería el de participantes **ocasionales**: aquellos participantes que en algún momento del curso se desvinculan, razón por la cual presentan un reducido número de sesiones y notas bajas.

El algoritmo k-medias ha identificado, para cada uno de los LMOOC, dos perfiles distintos cuyas características se ajustan a las del supergrupo de los participantes **ocasionales**. Por lo tanto, se establece que el grupo de participantes ocasionales estaría formado a su vez por dos subgrupos: el de participantes **ocasionales rutinarios** y el de participantes **ocasionales flexibles**, los cuales se analizarán en los dos siguientes subapartados.

7.5.1. GRUPO 3A: PARTICIPANTES OCASIONALES RUTINARIOS

Los integrantes del grupo de usuarios ocasionales, entre ellos los participantes ocasionales rutinarios, se caracterizan por sus bajas tasas de vinculación con el curso, lo cual se traduce en un número reducido de sesiones y en notas bajas. Sin embargo, el rasgo que distingue al subgrupo de participantes rutinarios es que tienden a programar las pocas sesiones que dedican al curso en las mismas franjas horarias. Los participantes ocasionales rutinarios suponen un 12,8% de los usuarios del LMOOC PA y un 18% de IP.

En las tablas 19 y 20 se muestran, para PA e IP respectivamente, los valores medios de las variables utilizadas por el algoritmo k-medias, junto con el tiempo total en el curso; el número de accesos a los foros, y el porcentaje de alumnos elegibles para el certificado final del curso. Como puede observarse en las tablas, los valores de la mayoría de las variables son bajos, incluida la dispersión en las horas de conexión, de 127 minutos para PA y 124 para IP, ya que estos participantes tienden a concentrar las sesiones de estudio en las mismas franjas horarias. Este perfil de participantes muestra un rendimiento bajo, aunque no tanto como el de los pre-consumidores. Esto se refleja especialmente en la poca elegibilidad para el certificado y en unas calificaciones sumamente bajas. De hecho, llaman la atención los valores medios tan similares en la nota final del curso, 0,11 y 0,10 para PA e IP respectivamente.

TABLA 19. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC PA DE LOS PARTICIPANTES OCASIONALES RUTINARIOS.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	93,50	0,11	127,1	5,37	17,39	5,3%
DESV. TÍPICA	104,91	0,14	45,5	3,89	76,01	-

TABLA 20. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC IP DE LOS PARTICIPANTES OCASIONALES RUTINARIOS.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	63,28	0,10	124,6	4,17	3,15	1,3%
DESV. TÍPICA	57,29	0,14	50,4	2,51	6,23	-

Sin embargo, aunque el rendimiento de este perfil sea generalmente bajo, hay indicios que hacen pensar que estos usuarios, en su mayoría, comienzan el curso con la intención de completarlo, pero que en cierto punto lo abandonan. Por un lado, el tiempo y el número de sesiones invertidos en el curso presentan valores relativamente bajos (93 minutos y 5 sesiones para PA, y 63 minutos y 4 sesiones para IP), pero asociados a una alta desviación típica, lo que indica que hay participantes con niveles de implicación más elevados. Por otro lado, el acceso a los foros no presenta valores tan bajos como el resto de las variables, especialmente para PA. En general, el nivel de actividad de los foros en ambos LMOOC fue superior durante los primeros días, cuando se animaba a los participantes a presentarse en la nueva lengua. De hecho, con respecto al uso de foros, PA fue un curso más dinámico, ya que los foros fueron un lugar en el que los participantes podían compartir sus experiencias como migrantes. Por esta razón, para todos los perfiles del LMOOC PA, el acceso a los foros presenta tasas superiores a IP. Pues bien, como puede observarse en las tablas 19 y 20, el perfil de usuarios ocasionales rutinarios presenta cierto grado de actividad en los foros, especialmente en PA, con una media de 17 accesos. Además, la desviación típica es de nuevo elevada, notablemente más en PA con un valor de 76, lo que indica que una parte de los participantes fue especialmente activa en los foros. Todo esto apunta a un conato real de finalizar el curso por parte de los participantes de este grupo.

Como en el resto de los perfiles analizados, los valores de las variables en PA son superiores a los de IP. Como ya se ha explicado, la mayor duración de PA y el perfil particular de su alumnado ofrecen una explicación plausible a estas diferencias.

7.5.2. GRUPO 3B: USUARIOS OCASIONALES FLEXIBLES

Al igual que en el caso de los participantes ocasionales rutinarios, los usuarios que pertenecen al grupo de participantes ocasionales flexibles, los cuales suponen un 12,8% y un 18% de los participantes de PA e IP respectivamente, presentan unas tasas bastante bajas de vinculación con el curso. Sin embargo, y al contrario que los rutinarios, estos participantes tienden a mostrar una gran disparidad en los momentos de inicio de las pocas sesiones que dedican al curso.

En las tablas 21 y 22, de nuevo se muestran para PA e IP respectivamente los valores medios de las variables utilizadas por el algoritmo k-medias, junto con el tiempo total en el curso; el número de accesos a los foros, y el porcentaje de alumnos elegibles para el certificado final del curso. Como puede observarse en las tablas, los valores de la nota final (0,14 y para PA y 0,11 para IP) y la elegibilidad para la obtención del certificado (15% para PA y 3,8% para IP) presentan valores muy bajos, lo cual indica un rendimiento insuficiente por parte de este grupo de participantes. La única variable para la que se observa un valor muy elevado (y prácticamente idéntico en ambos cursos, 335 minutos), es la de las horas de conexión al curso. Este hecho es predecible, ya que la característica que define a este perfil es la gran variación en las horas de inicio de las sesiones.

TABLA 21. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC PA DE LOS PARTICIPANTES OCASIONALES FLEXIBLES.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	86,91	0,14	335,4	5,19	15,10	15%
DESV. TÍPICA	108,48	0,22	105,2	4,41	77,86	-

TABLA 22. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC IP DE LOS PARTICIPANTES OCASIONALES FLEXIBLES.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	56,49	0,11	335,8	3,69	2,19	3,8%
DESV. TÍPICA	64,26	0,2	83,8	2,44	4,77	-

Al igual que sucedía en el grupo de ocasionales rutinarios, a pesar de un rendimiento generalmente bajo, hay indicios que apuntan a la intención de completar el curso por parte de los miembros de este grupo. De nuevo, el tiempo y el número de sesiones invertidos en el curso presentan valores relativamente bajos (87 minutos y 5,2 sesiones para PA, y 57 minutos y 3,7 sesiones para IP) asociados a una alta desviación típica que sugiere la existencia de participantes con mayores niveles de actividad. Además, el acceso a los foros no presenta valores tan bajos como otras variables, especialmente para PA, con 15 accesos y una desviación típica de 78, lo cual una vez más apunta a la existencia de miembros especialmente activos en los foros. Todo esto es indicio de que una parte de los participantes de este grupo tenía la intención real de completarlo con éxito.

Al igual que en el grupo anterior, el valor medio del número de sesiones; el tiempo total en el curso; la nota; la elegibilidad, y el acceso a los foros es mayor para PA. Las diferencias del número de sesiones y el tiempo en el curso se pueden explicar, como en el caso del grupo anterior, por la mayor duración del curso PA, mientras que los mejores resultados se pueden atribuir a una mayor disponibilidad y dedicación por parte del perfil de participantes de PA. Adicionalmente, la misma explicación que se dio para el grupo anterior en cuanto al uso de los foros es válida en este caso: un mayor acceso a los foros en PA se debe a un mayor dinamismo en los foros de ese LMOOC.

7.6. GRUPO 4: USUARIOS CONSTANTES

Como se ha explicado en el apartado 2.3, en el que se analizaron los gráficos de distribución de las muestras de ambos LMOOC en los distintos clústeres, para el curso PA el algoritmo k-medias ha dividido en dos grupos un conjunto de muestras que para IP ha agrupado en uno solo. En consecuencia, se consideró que existe un supergrupo, al que se ha denominado usuarios **constantes**, formado por participantes con un alto nivel de desempeño, ya que presentan un valor elevado tanto en las calificaciones como en el número de sesiones.

El grupo de usuarios **constantes** fue identificado por el algoritmo k-medias como un solo clúster del LMOOC IP, el cual reunía al 15,2% del total de los participantes del curso. La tabla 23 representa, para este LMOOC, los valores medios de las variables utilizadas por el algoritmo para agrupar las muestras, junto con el tiempo total en el curso; el número de accesos a los foros, y el porcentaje de alumnos elegibles para el certificado final del curso.

Como puede observarse en la tabla y es esperable del perfil de participantes **constantes**, los valores de todas las variables son elevados, con la excepción de la dispersión en las horas de conexión, cuyo valor es medio. El motivo de este valor intermedio en la variabilidad de los momentos de inicio de las sesiones del LMOOC IP es que este clúster engloba las medias de los subgrupos de participantes constantes rutinarios y flexibles. En consecuencia, el valor medio de esta variable para IP se encuentra comprendido entre los valores de la misma variable para cada uno de los dos subgrupos identificados por el algoritmo para PA.

TABLA 23. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC IP DE LOS PARTICIPANTES CONSTANTES.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	225,87	0,87	112,3	6,92	15,32	88,7%
DESV. TÍPICA	108,06	0,13	75,2	3,53	25,02	-

Como ya se ha establecido con anterioridad, el supergrupo de participantes constantes se divide a su vez en dos subgrupos: uno que tiende a concentrar las sesiones de estudio en los mismos horarios, lo cual se corresponde con la categoría de participantes rutinarios de del Peral, y otro que tiende a una mayor dispersión en las horas de inicio de las sesiones de estudio, lo cual coincide con el perfil de participantes flexibles. Estos dos grupos, los **participantes constantes rutinarios** y los **participantes constantes flexibles**, fueron identificados por el algoritmo k-medias como grupos separados únicamente para el curso PA y se analizarán en los dos siguientes subapartados.

7.6.1. GRUPO 4A: PARTICIPANTES CONSTANTES RUTINARIOS

Los participantes constantes rutinarios presentan altas tasas de vinculación con el curso, con un número elevado de sesiones y calificaciones altas. Sin embargo, el rasgo característico de este subgrupo es que los usuarios que lo integran tienden a iniciar las sesiones que dedican al curso en las mismas franjas horarias. Los participantes de este grupo suponen un 18% de los usuarios del LMOOC PA.

La tabla 24 muestra los valores medios de las variables utilizadas por el algoritmo k-medias para agrupar las muestras de PA, junto con el tiempo total en el curso; el número de accesos a los foros, y el porcentaje de alumnos elegibles para el certificado final del curso. Como puede observarse en la tabla, los valores de todas las variables son elevados, con la excepción de la dispersión en las horas de conexión, cuyo valor, como es de esperar, es muy bajo.

Al comparar los valores de las variables de la tabla 24 (participantes constantes rutinarios de PA), con los de la tabla 23 (supergrupo de participantes constantes de IP), se observa claramente su similitud, ya que, con la excepción de la dispersión, se hallan valores muy próximos para todas las variables, especialmente la nota final, con una diferencia de tan solo tres décimas.

TABLA 24. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC PA DE LOS PARTICIPANTES CONSTANTES RUTINARIOS.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	291,66	0,90	72,4	8,28	19,26	100%
DESV. TÍPICA	148,78	0,13	44,3	4,20	17,12	-

Sin embargo, de nuevo puede apreciarse cómo los valores de PA para las variables de tiempo total en el curso, nota final, número de sesiones, accesos a los foros y elegibilidad para el certificado final (que alcanza el 100% para PA) son ligeramente superiores a los de IP. Estas ligeras diferencias entre PA e IP se podrían explicar de nuevo por la mayor duración del LMOOC PA y por contar con un perfil de estudiantes con mayor dedicación al curso. Y, también de nuevo, estas desemejanzas son menos acusadas que en el perfil de

participantes exhaustivos. Esto se debe a que el de los participantes rutinarios es un perfil más moderado, lo cual contribuye a suavizar las disimilitudes entre ambos cursos.

7.6.2. GRUPO 4B: USUARIOS CONSTANTES FLEXIBLES

Los participantes **constantés flexibles**, los cuales suponen un 17,17% de los participantes del LMOOC PA, exhiben, tal y como lo hacía el grupo de usuarios flexibles descrito por del Peral, una alta vinculación con el curso y una elevada dispersión de las sesiones de estudio. Al igual que en el apartado anterior, para este subgrupo se representa una única tabla, la tabla 25, correspondiente al curso PA, en la que se muestran los valores medios de las variables utilizadas por el algoritmo k-medias, junto con el tiempo total en el curso; el porcentaje de alumnos elegibles para el certificado final del curso, y el número de accesos a los foros.

Como se observa en la tabla, los participantes constantes flexibles presentan valores altos en el tiempo total en el curso, el número de sesiones, la nota final y el número de accesos a los foros. Estos valores elevados producen, como es esperable, una alta tasa de éxito en el curso, con un 99,3% de los integrantes de este grupo elegible para el certificado final. Sin embargo, lo que caracteriza a los participantes de este grupo es la gran diferencia en los momentos de inicio de sesiones, con una media de 200 minutos.

TABLA 25. ESTADÍSTICAS DE INTERACCIÓN CON EL LMOOC PA DE PARTICIPANTES CONSTANTES Y FLEXIBLES.

	Tiempo (min.)	Nota	Dispersión sesiones (min.)	Sesiones	Accesos foro	Elegib. cert.
MEDIA	414,64	0,93	200	15,96	29,81	99,3%
DESV. TÍPICA	185,40	0,13	61,6	5,86	27,76	-

En resumen, en este capítulo se han analizado y cotejado los distintos clústeres identificados para los LMOOC PA e IP, y se han comparado con la taxonomía original de del Peral. De esta forma, al emplear un método más riguroso como son los algoritmos de descubrimiento de estructuras, específicamente el algoritmo k-medias, se ha demostrado una gran

coincidencia entre los clústeres resultantes y los perfiles descritos por del Peral. La siguiente imagen contiene un resumen de los seis perfiles identificados.

FIGURA 25. RELACIÓN DE LOS SEIS PERFILES IDENTIFICADOS A PARTIR DE LOS LMOOC PA E IP MEDIANTE EL ALGORITMO K-MEDIAS. ELABORACIÓN PROPIA.



En el siguiente capítulo se procederá a exponer las conclusiones derivadas del trabajo de investigación.

CAPÍTULO 8: CONCLUSIONES

8.1. INTRODUCCIÓN

Una vez finalizadas la interpretación y el análisis de los datos, este último capítulo está dedicado a la exposición de las conclusiones derivadas del trabajo de investigación.

De este modo, se comenzará con las conclusiones derivadas del análisis del marco teórico y, de las fases de diseño de la investigación y análisis de datos, recogidas en los capítulos dos a cinco de este documento y en los capítulos seis y siete, respectivamente. Seguidamente, se ofrecerá una respuesta a las preguntas de investigación planteadas en el capítulo inicial, tanto a las preguntas generales como a las específicas, y se verificará si la hipótesis inicial ha sido efectivamente demostrada.

Por último, en la sección final de este capítulo, se expondrán las aportaciones de esta tesis al emergente y fructífero campo de estudio de los LMOOC y las LA, además de proponerse líneas futuras de investigación que permitan ampliar y confirmar los descubrimientos realizados en la presente investigación.

8.2. CONCLUSIONES DEL MARCO TEÓRICO Y DE LA FASE DE INVESTIGACIÓN

Dentro de los trabajos de tesis, la fase de establecimiento del estado de la cuestión suele ser una de las más amplias y laboriosas. Sin embargo, esta fase es fundamental para el diseño de la estrategia metodológica de investigación que mejor se adapte a las características del objeto de estudio. En el caso particular de este trabajo, es más que evidente el efecto que una sólida base bibliográfica ha tenido en la selección de las herramientas y los métodos de análisis de datos. La profundización en los enriquecedores estudios y trabajos de investigación de otros autores que están contribuyendo a la expansión del conocimiento sobre este fenómeno relativamente joven como son los MOOC ha sido clave en el diseño de una estrategia que ha permitido obtener

información relevante a partir de los datos masivos generados por los LMOOC. En consecuencia, esta investigación se ha beneficiado enormemente de las herramientas y metodologías tan eficaces y rigurosas que se describen en los capítulos del estado de la cuestión de esta tesis, en particular en el capítulo cuarto.

Por lo tanto, una de las conclusiones de la fase de investigación es que esta debe estar basada en un estudio minucioso de la literatura científica del tema abordado. Si se pretende realizar un proyecto de investigación riguroso y de calidad, un investigador debe haber asimilado los conocimientos de los conceptos; principios; relaciones, e instrumentos sobre los que se asienta su tesis y debe utilizarlos como guía en el diseño de una estrategia metodológica que permita recoger; tratar, y analizar de forma eficaz el espacio muestral sobre el que se basará el estudio.

La metodología empleada en este trabajo de investigación es LA, una de las tendencias de futuro en educación más citadas en los informes internacionales de los últimos años, y la cual permite analizar la cantidad ingente de datos sobre los estudiantes y sus contextos de aprendizaje para así mejorar muchos aspectos de la experiencia educativa (Lang et al., 2017; León Urrutia et al., 2017; Sclater, 2017; Alexander et al., 2019; Ifenthaler et al., 2019; Brown et al., 2020; Romero y Ventura, 2020). Dentro de las LA existen una gran variedad de técnicas y de herramientas, tanto comerciales como de *software* libre, que permiten diseñar proyectos de investigación ambiciosos. Aunque en esta investigación se han utilizado varias de ellas, para la fase de pre-procesado de los datos se decidió desarrollar un programa *ad hoc* en código Java. El motivo fue que, en primer lugar, este autor ya había escrito gran parte del código del proyecto de LA que sirvió de base para la identificación de la taxonomía original de perfiles temporales (del Peral, 2019), por lo que el esfuerzo de adaptar el programa a las características del nuevo LMOOC fue asumible. En segundo lugar, la adaptación del programa evitaba la dependencia de *software* adicional, que en ocasiones conlleva un coste y una curva de aprendizaje que puede llegar a ser bastante acusada. Por último, la creación de un programa a medida permitió cubrir las necesidades exactas de procesado de los datos educativos de los que se disponía, lo cual supuso una

gran ventaja para la obtención de las variables a las que posteriormente se aplicó el algoritmo de descubrimiento de estructuras.

La elección de este algoritmo de descubrimiento de estructuras, el algoritmo k-medias, estuvo motivada de nuevo por lo aprendido durante la fase de revisión bibliográfica, ya que, debido a su fiabilidad y robustez, un gran número de los estudios analizados en el capítulo cinco recurrieron a este algoritmo (Kizilcec et al., 2013; Ferguson y Clow, 2015; Kovanovic et al., 2016; Rodrigues et al., 2016; Arora et al., 2017; Khalil y Ebner, 2017; Maya-Jariego et al., 2020). El algoritmo k-medias es, por lo tanto, un algoritmo que cuenta con un amplio respaldo por parte de la comunidad científica y que resulta idóneo para el descubrimiento de perfiles en datos educativos. Para la ejecución del algoritmo k-medias, se recurrió a *software* libre, más concretamente al programa estadístico R. De nuevo, este es un programa con amplio reconocimiento por la comunidad científica y que cuenta con numerosos manuales de aprendizaje e incluso con sus propios MOOC. La utilización de un programa con amplia documentación ha reducido la curva de aprendizaje, y ha facilitado enormemente el procesado de los datos y la generación de las gráficas que se han analizado en este documento.

En conclusión, puede afirmarse que la elección de la metodología y de las herramientas empleadas en este trabajo de investigación han sido las adecuadas, ya que estas han favorecido, por un lado, el ahorro de tiempo y de costes y, por otro, han facilitado un análisis confiable de la ingente cantidad de datos educativos generados por la plataforma en la que se alojaron los cursos.

El proceso metodológico que se acaba de exponer es evidencia de las dificultades que presenta un campo de estudio como el de los MOOC, ya que a los problemas intrínsecos de cualquier investigación se suman una serie de retos particulares. La inabarcable cantidad de datos generados por los MOOC obligan a utilizar técnicas como LA o la minería de datos para poder obtener información valiosa que permita mejorar la efectividad de tales cursos y así poder reducir las tasas de abandono. Sin embargo, el aprendizaje de estas técnicas y herramientas requiere perfiles e inclinaciones técnicas que pueden suponer un desafío a los investigadores de la rama de la filología. En

consecuencia, otra de las conclusiones derivadas de la fase de análisis de datos es la necesidad de que los investigadores tengan a su disposición distintas iniciativas formativas que les permitan aprovechar toda la potencialidad de LA y de las numerosas herramientas estadísticas disponibles, algunas incluso de forma gratuita, de modo que resulte factible su incorporación a los proyectos de investigación. Ello permitirá el diseño y la implementación de proyectos de investigación ambiciosos, que arrojen luz sobre innovaciones tecnológicas como los MOOC, cursos con capacidad de satisfacer las crecientes necesidades formativas de un número masivo de personas.

8.3. HIPÓTESIS Y PREGUNTAS DE LA INVESTIGACIÓN

Como se adelantó en el capítulo inicial, la presente tesis parte de una hipótesis general y de una serie de preguntas generales y específicas. La hipótesis era la siguiente: «es posible identificar mediante LA la taxonomía de perfiles de usuario de LMOOC basada en patrones temporales de uso descrita por del Peral (2019) para el LMOOC IP en un nuevo LMOOC, PA, pero los niveles de prevalencia y desempeño variarán en función de las características particulares de cada LMOOC y de sus participantes».

En esta sección se dará respuesta a cada una de las preguntas planteadas en la introducción, tanto a las generales como a las específicas, tras lo cual será posible confirmar o refutar la hipótesis inicial.

8.3.1. 1ª PREGUNTA DE LA INVESTIGACIÓN

La primera pregunta de la investigación planteada en el capítulo de introducción fue la siguiente: «¿existe una taxonomía de perfiles de participantes de LMOOC basada en patrones temporales de uso?».

A partir del análisis de datos llevado a cabo en el capítulo anterior, puede responderse afirmativamente a esta pregunta, ya que la ejecución del algoritmo k-medias sobre los datos obtenidos tanto de PA como de IP ha resultado en la identificación de una serie de clústeres que se corresponden con perfiles diferenciados de participantes de LMOOC. Las variables a partir de las cuales se determinaron las distintas agrupaciones fueron tres:

- Nota media final en el curso.
- Número total de sesiones en el curso.
- Medida de dispersión de la hora de inicio de las sesiones.

De estas variables, las dos últimas son variables que entrañan un componente temporal. La medida de dispersión de la hora de inicio de las sesiones es una variable directamente temporal, ya que mide la concentración de las horas de estudio en franjas horarias. Por otra parte, el número total de sesiones es una variable que mide el tiempo dedicado al curso de forma indirecta, ya que todas las sesiones tienen asociadas una duración determinada. En consecuencia, los perfiles conformados a partir de la ejecución del algoritmo k-medias son perfiles de participantes de LMOOC basados en patrones temporales de uso.

Dentro de esta pregunta general, en el capítulo inicial se planteó una serie de preguntas específicas a las cuales se dará respuesta en los siguientes subapartados.

¿CUÁNTOS PERFILES PUEDEN IDENTIFICARSE?

La ejecución del algoritmo k-medias para los datos obtenidos del LMOOC PA produjo un resultado de seis clústeres diferenciados, que no coincidían con los cinco clústeres identificados originalmente por del Peral. Posteriormente, el análisis de los datos de salida del algoritmo permitió una mejor comprensión de las discrepancias entre el número de clústeres definidos en la taxonomía de del Peral y el número de clústeres determinados por el algoritmo. De este modo, dentro de la taxonomía de del Peral, los pre-consumidores; los participantes exhaustivos; los rutinarios, y los flexibles se corresponden claramente con cuatro de los seis perfiles identificados en PA por el algoritmo, ya que los rasgos de estos perfiles concuerdan claramente con la descripción de los perfiles de del Peral. Sin embargo, se ha considerado que los perfiles de participantes rutinarios y participantes flexibles son en realidad subconjuntos de un supergrupo al que se ha denominado participantes constantes. El rasgo que define a este grupo de participantes es su alto nivel de desempeño, ya que presentan un valor elevado tanto en las calificaciones como en el número de sesiones.

El quinto perfil de del Peral, participantes intensivos, ha sido renombrado en este estudio a participantes ocasionales, ya que se ha considerado que el rasgo definitorio de este grupo no es que los estudiantes completen el curso en muy pocas sesiones (solo un porcentaje muy reducido lo hace), sino el poco tiempo dedicado al curso, consecuencia de un abandono temprano. De este modo, los últimos dos perfiles identificados en PA encajan con esta redefinición del perfil original de usuarios intensivos (ahora ocasionales), ya que son usuarios que comienzan el curso con la intención de completarlo, pero que en su mayoría lo abandonan. La diferencia entre estos dos últimos perfiles identificados por el algoritmo es que en uno las horas de inicio de las sesiones tienden a concentrarse en las mismas franjas horarias, mientras que en el otro grupo existe una mayor dispersión. Por lo tanto, se ha considerado que estos dos perfiles de PA son en realidad dos subgrupos del perfil de usuarios intensivos de del Peral, bautizado ahora como participantes ocasionales, a los cuales se ha denominado participantes ocasionales rutinarios y participantes ocasionales flexibles.

Al corroborar que el número de seis clústeres identificados por el algoritmo k-medias para el LMOOC PA no correspondía con el número de perfiles identificados por del Peral, como parte de la fase de post-procesado del proyecto de LA se ejecutó el algoritmo k-medias sobre los datos procedentes del LMOOC IP. El resultado de esta operación fue, en un primer momento, de tres clústeres, aunque se razonó que una solución de cinco clústeres era también apropiada. En principio, podría deducirse que estos cinco perfiles identificados corresponderían con los perfiles descritos por del Peral. Sin embargo, el análisis de los datos de salida del algoritmo permitió indagar en la relación existente entre ambos conjuntos de cinco clústeres.

En primer lugar, se comprobó que los perfiles de pre-consumidores y participantes exhaustivos encontraban una clara correspondencia con dos de los cinco perfiles de IP, y que la descripción que hizo del Peral de estos perfiles concordaba con los rasgos hallados en las agrupaciones análogas de IP.

En segundo lugar, el perfil de participantes intensivos (renombrado a participantes ocasionales) encuentra su equivalente, al igual que para PA, en

dos perfiles distintos de IP. Estos dos perfiles aglutinan a un grupo de estudiantes que, pese a comenzar el curso activamente y con la intención presumible de finalizar el MOOC, en un alto porcentaje terminan por abandonarlo. La diferencia entre ambos perfiles es que en uno las horas de inicio de las sesiones tienden a concentrarse en las mismas franjas horarias, mientras que en el otro grupo existe una mayor dispersión. De nuevo, se ha considerado que estos dos perfiles de IP son, al igual que en el caso de PA, dos subgrupos del perfil de participantes ocasionales: participantes ocasionales rutinarios y participantes ocasionales flexibles.

Por último, los dos últimos perfiles de del Peral, el de los participantes rutinarios y el de participantes flexibles, encuentran su equivalencia en las agrupaciones calculadas por el algoritmo, no en dos grupos análogos, sino en un supergrupo. Se ha denominado a esta categoría participantes constantes, ya que se corresponde con un perfil de estudiantes con un alto nivel de desempeño y un valor elevado tanto en las calificaciones como en el número de sesiones. Dentro de este grupo de participantes constantes, existen por lo tanto dos subgrupos: el de los participantes constantes rutinarios y el de los participantes constantes flexibles. Como ya se ha detallado, el algoritmo k-medias identificó el supergrupo de participantes constantes para el LMOOC IP, mientras que para PA identificó a sus dos subgrupos: participantes constantes rutinarios y participantes constantes flexibles.

La tabla 26 muestra la correspondencia entre los perfiles identificados por del Peral y los grupos obtenidos a partir de la ejecución del algoritmo k-medias, tanto para el LMOOC PA como para el LMOOC IP.

TABLA 26. CORRESPONDENCIA DE LOS PERFILES DE DEL PERAL Y EL ALGORITMO K-MEDIAS PARA PA E IP

DEL PERAL	K-MEDIAS PA	K-MEDIAS IP
Pre-consumidores	Pre-consumidores	Pre-consumidores
Exhaustivos	Exhaustivos	Exhaustivos
Intensivos	Ocasionales rutinarios	Ocasionales rutinarios
	Ocasionales flexibles	Ocasionales flexibles
Rutinarios	Constantes rutinarios	Constantes
Flexibles	Constantes flexibles	

En consecuencia, tal y como se refleja en la tabla, en respuesta a la pregunta sobre el número de perfiles identificados en los LMOOC, puede determinarse la existencia de cuatro perfiles principales, dos de los cuales están divididos a su vez en dos sub-perfiles. Por lo tanto, el número total de perfiles es de seis, divididos de este modo:

- Pre-consumidores.
- Participantes exhaustivos.
- Participantes ocasionales.
 - Participantes ocasionales rutinarios
 - Participantes ocasionales flexibles
- Participantes constantes.
 - Participantes constantes rutinarios
 - Participantes constantes flexibles

¿CUÁLES SON LOS RASGOS IDENTIFICATIVOS DE CADA PERFIL?

Las variables utilizadas por el algoritmo k-medias para la clasificación de los participantes de los LMOOC en distintos perfiles fueron las siguientes:

- Nota media final en el curso.
- Número total de sesiones en el curso.
- Medida de dispersión de la hora de inicio de las sesiones.

Los seis perfiles identificados presentan una serie de características que dependen de los distintos valores de dichas variables. Además, también se ha calculado el valor de otras variables correlacionadas, las cuales aportan información adicional sobre los distintos perfiles. Estas variables son:

- Tiempo total en el curso.
- Número de accesos a los foros.
- Porcentaje de participantes elegibles para el certificado.

De este modo, las características que definen a cada uno de los perfiles son las siguientes:

- **Pre-consumidores.** Como puede observarse en la tabla 27, los pre-consumidores son participantes con niveles mínimos de actividad en el curso; con calificaciones cercanas a 0; un total de minutos dedicados al curso menor de 25, y un promedio de sesiones y de accesos a los foros cercano a 1. La variabilidad en el inicio de las sesiones presenta los valores más bajos entre todos los perfiles, ya que una gran parte de estos participantes cuenta con una sola sesión. El porcentaje de participantes elegibles para el certificado es, evidentemente, también cercano al 0%.

TABLA 27. ESTADÍSTICAS DE PRE-CONSUMIDORES.

	Min.	Sesiones	Min./ses.	Nota	Disp.	Foro	Elegib.
PA	24,01	1,53	15,69	0,03	8,98	1,22	0,4%
IP	16,30	1,34	12,16	0,02	8,05	0,73	0%

- **Participantes exhaustivos.** El rasgo característico de este grupo es que presenta un valor elevado en todas las variables, como puede observarse en la tabla 28. De este modo, el número de sesiones, el tiempo dedicado al curso y el acceso a los foros presentan los valores más elevados de entre todos los perfiles. La dispersión horaria de las sesiones también es elevada, lo cual es predecible, dado el número tan elevado de sesiones. También las notas medias finales y la elegibilidad para el certificado final exhiben valores altos, aunque no los más elevados de entre todos los perfiles, ya que son superados por los participantes constantes, tanto en el LMOOC PA como en IP. Este menor rendimiento de los participantes exhaustivos con respecto a otros perfiles fue adelantado por el trabajo inicial de del Peral (2019) y es confirmado en este estudio. Se trata de fenómeno contraintuitivo, debido a que, a mayor tiempo dedicado al curso, un participante debería exhibir un mejor rendimiento y unas calificaciones más elevadas.

TABLA 28. ESTADÍSTICAS DE PARTICIPANTES EXHAUSTIVOS.

	Min.	Sesiones	Min./ses.	Nota	Disp.	Foro	Elegib.
PA	805,13	40,42	19,92	0,87	208,7	104,97	89,8%
IP	390,71	24,26	16,10	0,78	158,1	23,81	68,8%

Finalmente, destaca el hecho de que los participantes de este grupo tiendan a concentrar el estudio en un número elevado de sesiones de duración corta. Como puede observarse en la tabla 28, la media de minutos por sesión es notablemente inferior a la de los participantes constantes y solo ligeramente superior al resto de perfiles.

- **Participantes ocasionales.** El rasgo definitorio de este grupo de participantes es el escaso rendimiento en el curso, debido principalmente a un abandono temprano del LMOOC. Como puede observarse en las tablas 29 y 30, este grupo de participantes presenta valores bajos en la mayoría de las variables (número de sesiones, tiempo total, minutos por sesión, notas finales, tasas de elegibilidad, acceso a foros), aunque dichos valores son claramente superiores a los del perfil de pre-consumidores. Respecto al valor de la variable de dispersión del inicio de las sesiones, el valor varía en cada uno de los dos sub-grupos que forman esta categoría:
 - **Participantes ocasionales rutinarios.** Tienden a concentrar las pocas sesiones dedicadas al curso dentro de las mismas franjas horarias, por lo que el valor de la variable de dispersión de sesiones es bajo.

TABLA 29. ESTADÍSTICAS DE PARTICIPANTES OCASIONALES RUTINARIOS

	Min.	Sesiones	Min./ses.	Nota	Disp.	Foro	Elegib.
PA	93,50	5,37	17,41	0,11	127,1	17,39	5,3%
IP	63,28	4,17	15,17	0,10	124,6	3,15	1,3%

- **Participantes ocasionales flexibles.** Tienden a concentrar las pocas sesiones dedicadas al curso en distintas franjas

horarias, por lo que el valor de la variable de dispersión de sesiones es elevado.

TABLA 30. ESTADÍSTICAS DE PARTICIPANTES OCASIONALES FLEXIBLES

	Min.	Sesiones	Min./ses.	Nota	Disp.	Foro	Elegib.
PA	86,91	5,19	16,75	0,14	335,4	15,10	15%
IP	56,49	3,69	15,31	0,11	335,8	2,19	3,8%

- **Participantes constantes.** El rasgo que define a este grupo de participantes es su alto nivel de desempeño, ya que presentan valores muy altos en el tiempo total en el curso; el número de sesiones, y el número de accesos a los foros. Estos valores producen, como es esperable, una alta tasa de éxito en el curso y una nota final elevada, con cifras superiores al resto de los perfiles, incluido el perfil de participantes exhaustivos. En la tabla 31 se muestran los valores de las variables para el curso IP, mientras que en las tablas 32 y 33 se muestran, para el curso PA, los valores de los dos subgrupos de los participantes constantes: los rutinarios y los flexibles.

TABLA 31. ESTADÍSTICAS DE PARTICIPANTES CONSTANTES DEL LMOOC IP

	Min.	Sesiones	Min./ses.	Nota	Disp.	Foro	Elegib.
IP	225,87	6,92	32,64	0,87	112,3	15,32	88,7%

Cabe también destacar que la media de minutos por sesión de los participantes de este grupo es claramente superior a la del resto de perfiles, lo cual indica que los participantes de este grupo tienden a concentrar el estudio en sesiones de duración larga. Respecto a la variable de dispersión del inicio de las sesiones, su valor varía en cada uno de los dos sub-grupos que forman esta categoría:

- **Participantes constantes rutinarios.** Tienden a concentrar las sesiones dedicadas al curso dentro de las mismas franjas horarias.

TABLA 32. ESTADÍSTICAS DE PARTICIPANTES CONSTANTES RUTINARIOS DEL LMOOC PA

	Min.	Sesiones	Min./ses.	Nota	Disp.	Foro	Elegib.
PA	291,66	8,28	35,22	0,90	72,4	19,26	100%

- **Participantes constantes flexibles.** Tienden a concentrar las sesiones dedicadas al curso en distintas franjas horarias.

TABLA 33. ESTADÍSTICAS DE PARTICIPANTES CONSTANTES FLEXIBLES DEL LMOOC PA

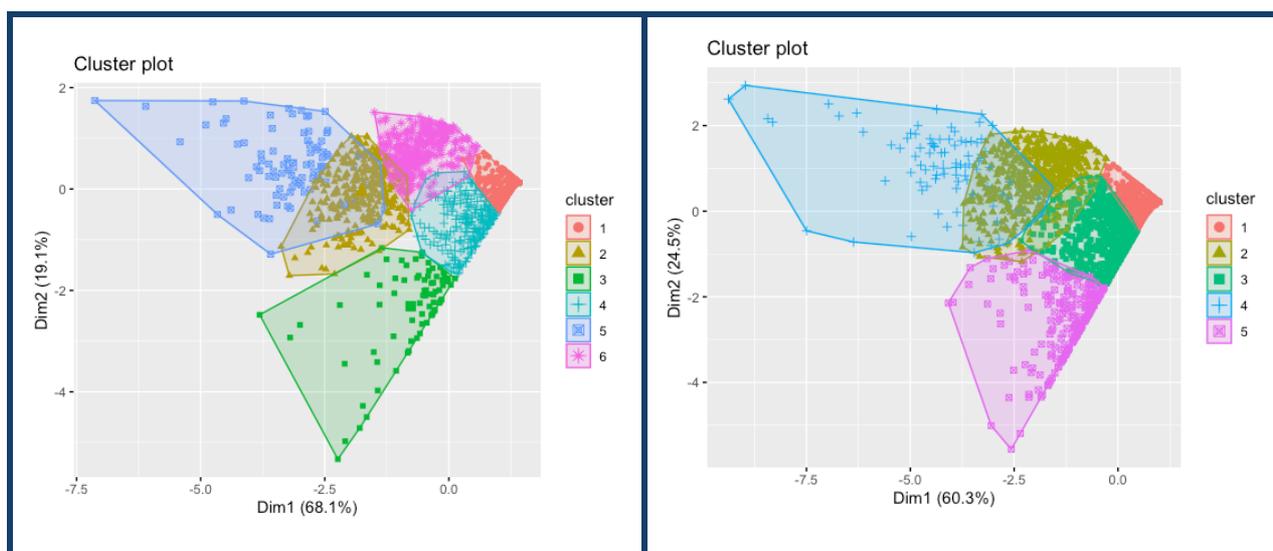
	Min.	Sesiones	Min./ses.	Nota	Disp.	Foro	Elegib.
PA	414,64	15,96	25,98	0,93	200	29,81	99,3%

¿ES DICHA TAXONOMÍA EXTRAPOLABLE A DIFERENTES LMOOC?

Después de que en el capítulo anterior se analizaran los datos obtenidos tras la ejecución del algoritmo k-medias, tanto para el LMOOC PA como para IP, puede afirmarse que la taxonomía de LMOOC basada en patrones temporales de uso previamente descrita es extrapolable a distintos LMOOC de la misma tipología que los dos cursos estudiados. Esta afirmación se basa principalmente en tres hechos claramente contrastados.

En primer lugar, la distribución de muestras obtenidas a partir de la ejecución del algoritmo k-medias exhibe, para todos los clústeres, agrupaciones muy similares en proporción; forma; densidad, y posición de los individuos en el espacio muestral, como puede observarse en el gráfico 11, el cual ya se representó en el capítulo anterior. La diferencia más importante entre los dos gráficos, y por lo tanto en las distintas agrupaciones obtenidas para los LMOOC PA e IP, se debe a que los participantes constantes aparecen agrupados en un solo clúster en el LMOOC IP, mientras que en PA aparecen representados en dos subgrupos distintos. Otras diferencias menores fueron explicadas debido a la incorrecta agrupación por parte del algoritmo de algunas de las muestras.

GRÁFICO 11. YUXTAPOSICIÓN DE LOS DISTINTOS PERFILES PARA LOS DATOS DE PA (IZQUIERDA) E IP (DERECHA).



En segundo lugar, tal y como se ha descrito en la pregunta de investigación previa, los perfiles identificados exhiben características comunes en ambos LMOOC. Aunque, como se explicará más adelante, existen ligeras diferencias entre los dos LMOOC con respecto a los rangos de valores de cada una de las variables, dichos valores tienden a mostrar una proporcionalidad similar en cada uno de los perfiles. Tomemos como ejemplo el valor del número de sesiones. En todos los perfiles, la media de sesiones dedicadas al curso es mayor en PA que IP. Sin embargo, la diferencia en el número de sesiones de cada uno de los perfiles de PA es proporcional a la diferencia en el número de sesiones de cada uno de los perfiles de IP. Esto demuestra que los perfiles identificados comparten rasgos comunes y que esta relación de perfiles es por lo tanto extensible a otros LMOOC de la misma tipología.

En tercer lugar, cabe destacar que, aunque ambos MOOC comparten plataforma y tipología (MOOC de lenguas con formato de xMOOC lanzados en la plataforma edX), las características de los dos LMOOC analizados son notablemente distintas. En primer lugar, los cursos no tuvieron la misma estructura, ya que IP estaba dividido en tres módulos y PA en cuatro. En consecuencia, la duración de los cursos fue también distinta. IP tuvo una duración de tres semanas, mientras que PA estuvo disponible durante seis semanas, el doble de tiempo. En segundo lugar, el perfil de los participantes de

cada curso exhibía diferencias significativas. IP estaba dirigido a trabajadores con un nivel B1 de inglés, mientras que PA estaba diseñado para migrantes. Como se ha explicado anteriormente, las diferencias entre los dos perfiles produjeron variaciones en cuanto a la dedicación al curso, ya que el perfil de estudiantes de IP tiene *a priori* un tiempo más limitado para dedicar al LMOOC, mientras que se deduce que los de PA tienen mayor tiempo disponible y una mayor necesidad de asimilar los contenidos del curso. El que la misma taxonomía de participantes haya sido identificada en LMOOC de rasgos tan dispares hace prever que los perfiles descritos son comunes, como mínimo, a todos los cursos de la misma tipología, esto es, a todos los LMOOC con un diseño de tipo xMOOC.

En conclusión, es esperable que esta misma taxonomía sea identificada en diferentes tipos de LMOOC, con las posibles variaciones descritas en función de la presencia de los dos subgrupos de las categorías de participantes ocasionales y constantes.

8.3.2. 2ª PREGUNTA DE LA INVESTIGACIÓN

La segunda pregunta planteada en el capítulo de introducción fue la siguiente: «¿existen diferencias en los distintos perfiles en función del LMOOC del que se obtienen los datos?». El análisis llevado a cabo en el capítulo anterior ofrece una respuesta afirmativa a esta pregunta, ya que tanto las variables utilizadas para la definición de los perfiles como ciertas variables correlacionadas con ellas presentan valores diferentes para cada uno de los dos LMOOC. Los valores de estas variables ya han sido representados en distintas tablas, tanto en el apartado anterior como en el capítulo 7. Sin embargo, estas variables volverán a ser analizadas en detalle en los siguientes apartados al responder a las preguntas específicas asociadas a esta segunda pregunta general de la investigación.

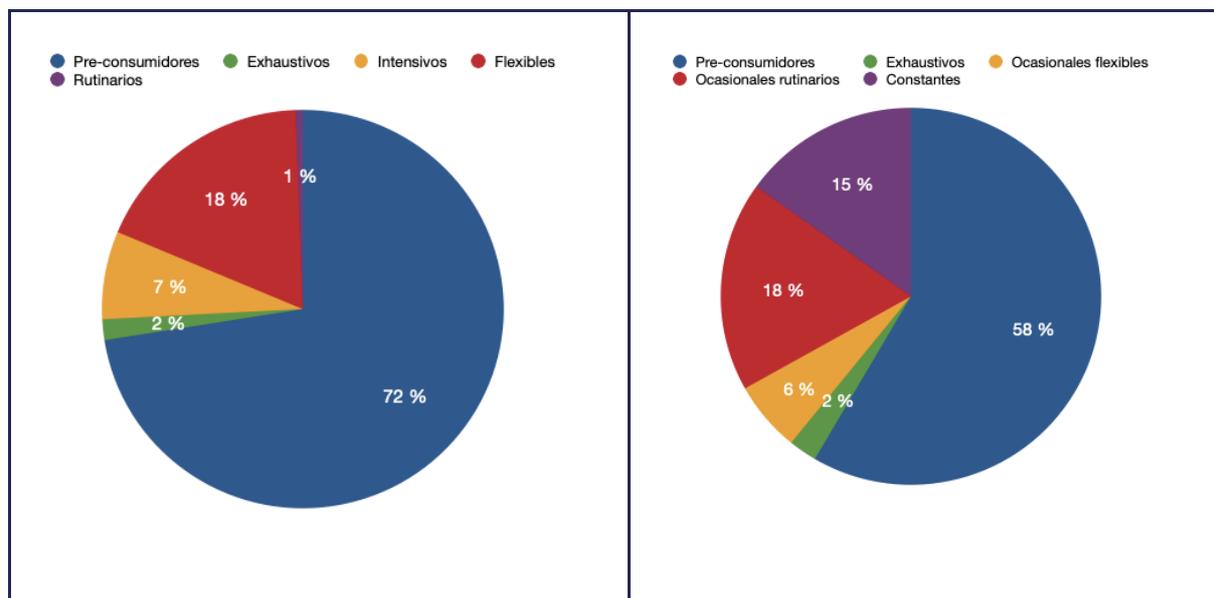
Para analizar y comparar las características de cada uno de los perfiles, se tomarán como referencia los valores obtenidos de la ejecución del algoritmo k-medias en lugar de los valores originales del estudio de del Peral (2019). En el estudio de del Peral, la forma de delimitar los distintos perfiles fue mediante el establecimiento de valores barrera tanto en el número de sesiones como en la

dispersión de los instantes de inicio de las sesiones. La clasificación resultante estaba basada, por lo tanto, en límites apriorísticos y se utilizó LA para encontrar todas las muestras del conjunto de datos que se adaptaban a los rasgos predefinidos para cada categoría. Estos límites, tanto en el número de sesiones como en la dispersión de los comienzos de sesiones, eran límites rígidos y no fluidos. Una diferencia de una sesión o de un minuto entre dos muestras podía hacer que fueran categorizadas en grupos distintos. Por el contrario, el algoritmo k-medias es un algoritmo de descubrimiento de estructuras cuyo objetivo es intentar encontrar una estructura en los datos sin ninguna idea *a priori* de lo que se debería encontrar (Baker e Inventado, 2014). De este modo, el algoritmo k-medias clasifica las muestras a partir del valor medio de los objetos pertenecientes al grupo (MacKay, 2015; Aldowah, 2019).

Tomemos como ejemplo la prevalencia de los perfiles. Como se observa en el gráfico 12, los porcentajes de prevalencia en las dos formas de delimitar los perfiles, de forma rígida en el estudio inicial de del Peral (gráfico de la izquierda) y mediante el algoritmo k-medias (gráfico de la derecha), no coinciden. Únicamente existe coincidencia en el grupo de participantes exhaustivos, los cuales representan un 2% del total en ambas formas de calcular los perfiles. El porcentaje de los participantes ocasionales, los cuales se corresponden con la suma de los perfiles flexible y rutinario de la clasificación de del Peral, tienen valores similares, pero no exactos, con 15% para los ocasionales, y un 19% para la suma de flexibles y rutinarios. Esto se debe claramente a que en la clasificación de del Peral se han contabilizado en estos perfiles participantes con un número de sesiones entre 2 y 15, pero que realmente pertenecen a la categoría de participantes ocasionales por su bajo nivel de desempeño.

Por último, puede observarse cómo la categoría de pre-consumidores de la clasificación de del Peral (72%) es más numerosa que la obtenida mediante el algoritmo k-medias. Esto se debe a que un porcentaje considerable de los participantes pre-consumidores de del Peral ha sido clasificado como ocasionales (rutinarios o flexibles) por el algoritmo, probablemente debido a la realización de actividades que no aportan puntuación a la nota final, como el cuestionario inicial o los vídeo tutoriales.

GRÁFICO 12. PORCENTAJES DE PREVALENCIA DE CADA PERFIL DE PARTICIPANTES EN EL ESTUDIO ORIGINAL DE DEL PERAL (IZQUIERDA) Y MEDIANTE EL ALGORITMO K-MEDIAS (DERECHA).



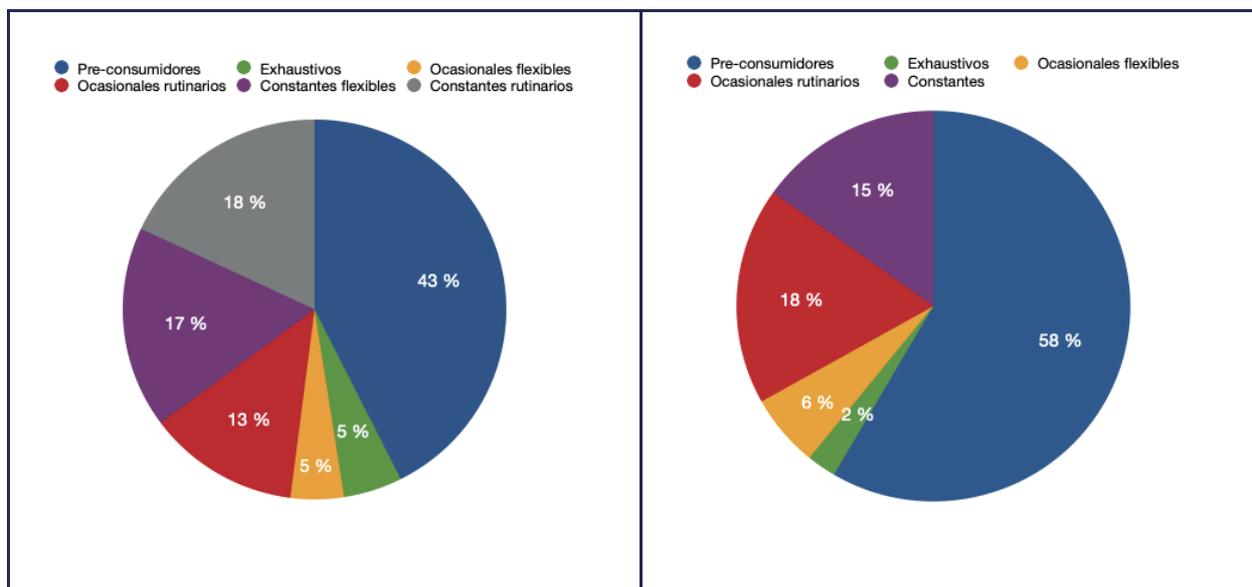
Por lo tanto, para comparar tanto la prevalencia como el resto de las características de cada uno de los perfiles en los LMOOC PA e IP se recurrirá a los datos obtenidos como resultado de aplicar la misma metodología al conjunto de datos generados en cada uno de los LMOOC, esto es, a los datos obtenidos de la ejecución del algoritmo k-medias.

¿QUÉ DIFERENCIAS HAY RESPECTO A LA PREVALENCIA DE PERFILES?

Como puede comprobarse en el gráfico 13, existen grandes diferencias en el porcentaje de participantes que pertenecen a cada uno de los perfiles para los LMOOC PA (izquierda) y IP (derecha).

En el LMOOC IP hay una mayor proporción de participantes que pertenecen a perfiles con menores niveles de actividad y de desempeño. De este modo, el porcentaje de participantes que pertenecen a la categoría de pre-consumidores es mucho mayor en IP (58%) que en PA (43%), al igual que es mayor también el porcentaje de participantes ocasionales en IP (24%) que en PA (18%). Por el contrario, en PA hay un mayor número de participantes exhaustivos, un 5% frente a un 2% de IP, y de participantes constantes, un 35% frente a un 15% de IP.

GRÁFICO 13. PORCENTAJES DE PREVALENCIA DE CADA PERFIL DE PARTICIPANTES OBTENIDOS MEDIANTE EL ALGORITMO K-MEDIAS PARA PA (IZQUIERDA) E IP (DERECHA).



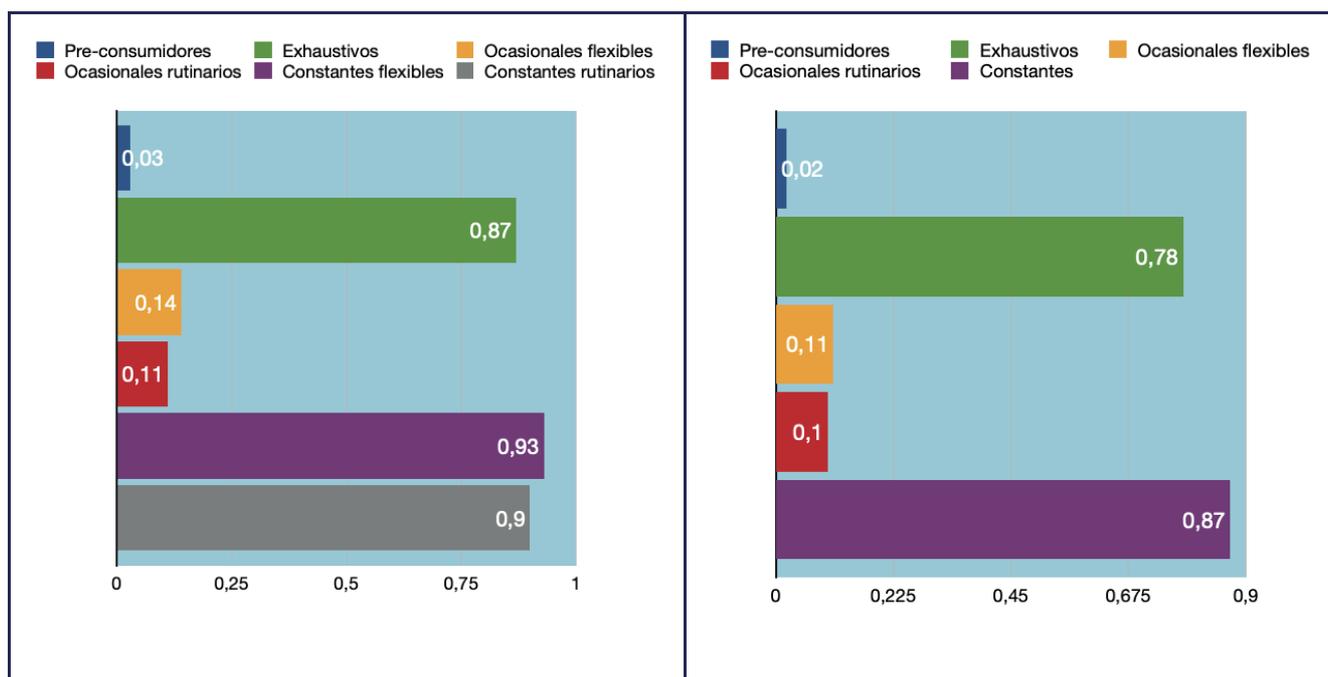
Por lo tanto, puede concluirse que en el LMOOC PA la proporción de participantes pertenecientes a los dos perfiles más activos, constantes y exhaustivos, es superior que en el LMOOC IP.

¿QUÉ DIFERENCIAS HAY RESPECTO AL GRADO DE DESEMPEÑO?

Como puede observarse en el gráfico 14, las calificaciones finales son bastante similares en ambos LMOOC, además de ser coherentes con las características de cada uno de los distintos perfiles de participantes. De este modo, los perfiles con una mayor implicación en el curso, los exhaustivos y los constantes, presentan unas notas medias altas, entre el 8 y el 9, mientras que los perfiles con menor vinculación con el curso, los pre-consumidores y los participantes ocasionales, presentan notas en torno al 0 y al 1, respectivamente.

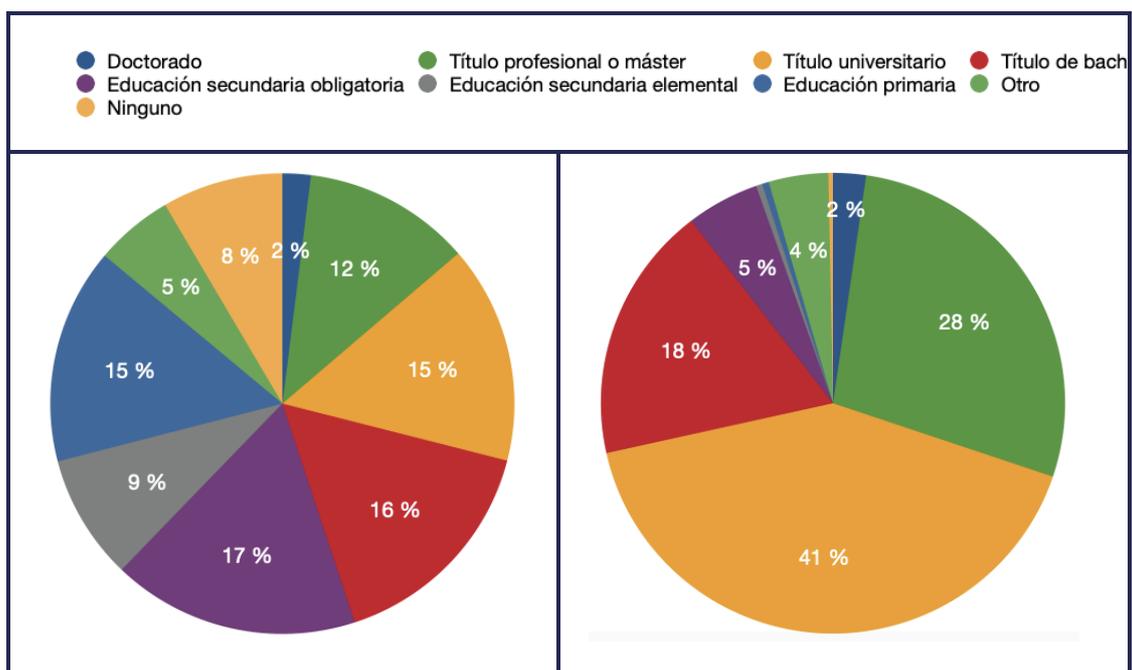
Aunque bastante similares, la media de las calificaciones finales es siempre ligeramente inferior para todos los perfiles del LMOOC IP, excepto en el perfil de participantes exhaustivos, donde la diferencia es más acusada, de casi una décima (lo cual supone un punto en una escala de 10).

GRÁFICO 14. CALIFICACIONES MEDIAS DE CADA PERFIL DE PARTICIPANTES DE PA (IZQUIERDA) E IP (DERECHA).



Este hecho en sí no sería reseñable si no se dispusiera de información demográfica del conjunto de participantes de ambos cursos. Afortunadamente, la plataforma donde se alojaron los LMOOC permite acceder a ciertos datos personales de los estudiantes, como la fecha de nacimiento; el género, o el nivel de estudios. En este caso particular es realmente interesante notar que, como se muestra en el gráfico 15, el nivel de estudios de los participantes de IP es considerablemente superior al de los participantes de PA. De este modo, el 71% de los participantes de IP cuenta con educación superior (estudios universitarios, de postgrado o de doctorado), porcentaje significativamente superior a la media en estos cursos, la cual se halla en torno al 60% (Díez-Arcón, 2021), mientras que el porcentaje de estudiantes de IP que no tienen completada la educación secundaria obligatoria es de tan solo el 2%. Por el contrario, el porcentaje de participantes de PA que no han finalizado la educación secundaria obligatoria es del 29% y el porcentaje de aquellos que poseen estudios universitarios es del 32%.

GRÁFICO 15. PORCENTAJES DEL NIVEL EDUCATIVO DE LOS PARTICIPANTES DE PA (IZQUIERDA) E IP (DERECHA).



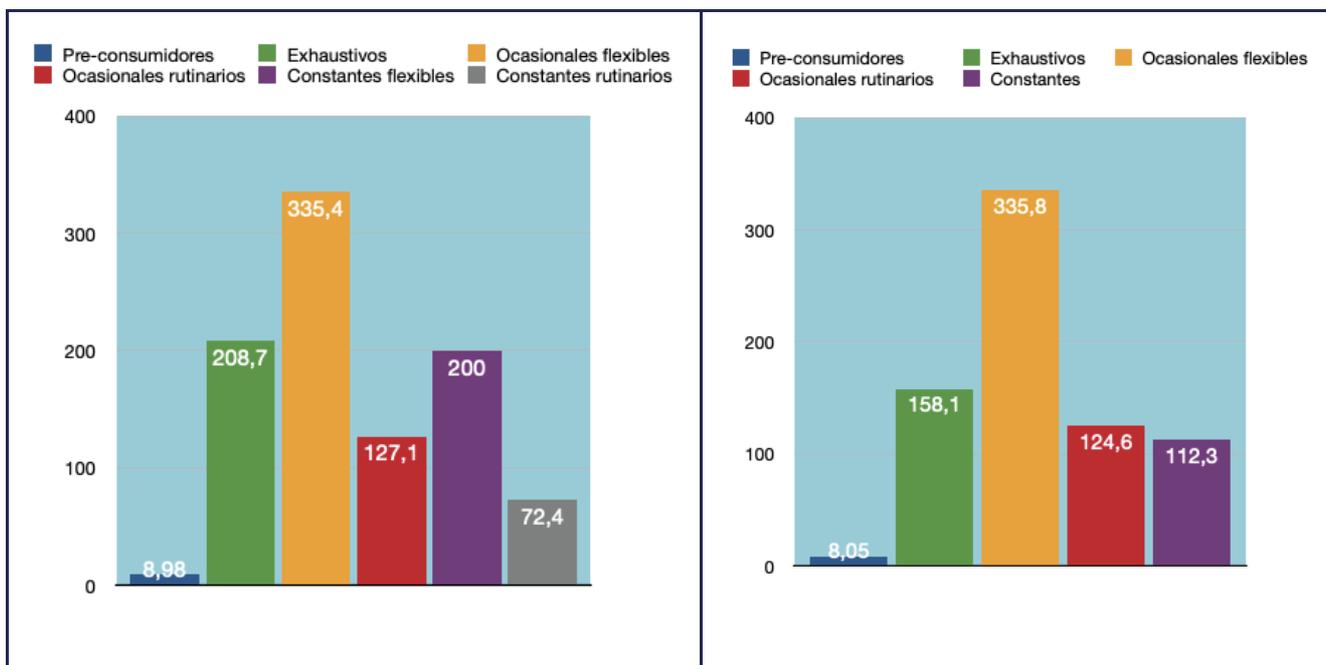
A la vista de los datos, es sorprendente que las calificaciones de todos los perfiles de PA sean superiores a las de IP. Este hecho apunta claramente a que el mayor desempeño de los participantes de PA se debe a una mayor implicación con el LMOOC, ya que su motivación para avanzar en el curso y asimilar los contenidos es más poderosa que la de los participantes de IP.

¿QUÉ DIFERENCIAS HAY RESPECTO A LA VARIABILIDAD EN LOS TIEMPOS DE ACCESO AL CURSO?

La variabilidad de las horas de inicio de las sesiones presenta valores muy similares en varios de los perfiles identificados. De hecho, como puede observarse en el gráfico 16, este valor de dispersión es muy similar para ambos grupos de pre-consumidores (8 minutos para IP y 9 para PA), lo cual es esperable en un grupo en el que gran parte de los integrantes solo cuenta con una sesión.

Más sorprendente es el caso de los participantes ocasionales, tanto los flexibles como los rutinarios, con valores prácticamente idénticos. Las diferencias más notables se hayan en los perfiles de participantes exhaustivos, con 50 minutos más de variación para PA, y en el grupo de los participantes constantes, cuya media ponderada sería de 134 minutos para PA, mientras que la media para IP es de 112.

GRÁFICO 16. VARIACIÓN EN LOS TIEMPOS DE INICIO DE SESIÓN PARA CADA PERFIL DE PARTICIPANTES DE PA (IZQUIERDA) E IP (DERECHA).

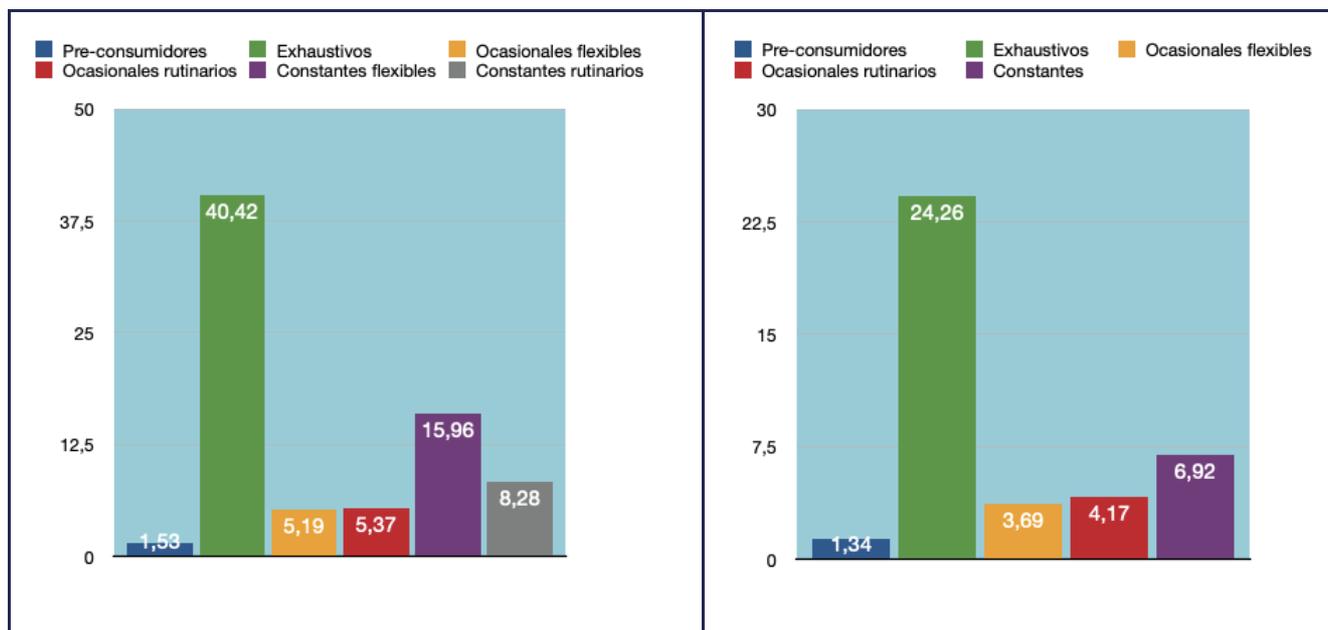


Se puede concluir, por lo tanto, que existe diferencia en cuanto a la dispersión de las horas de inicio de las sesiones únicamente en algunos de los perfiles, especialmente en el perfil de participantes exhaustivos.

¿QUÉ DIFERENCIAS HAY RESPECTO AL NÚMERO DE SESIONES TOTALES EN EL CURSO?

La media en el número de sesiones entre los participantes del LMOOC PA es claramente superior a la del LMOOC IP para todos los perfiles. Como puede observarse en el gráfico 17, la diferencia es mínima en el perfil de pre-consumidores, ya que en ambos casos el número de sesiones está cerca del valor mínimo de 1. Sin embargo, para el resto de los grupos la diferencia se hace cada vez más evidente según aumenta el grado de implicación en cada uno de los perfiles. Así pues, para los participantes ocasionales la diferencia es de algo más de una sesión en los subtipos rutinario y flexible, valor que aumenta para los usuarios constantes, cuyos valores en los dos subgrupos de PA superan a los de IP. Esta diferencia alcanza el valor máximo en el perfil de participantes exhaustivos, con una media de 16 sesiones más en PA que en IP.

GRÁFICO 17. NÚMERO MEDIO DE SESIONES DE CADA PERFIL DE PARTICIPANTES DE PA (IZQUIERDA) E IP (DERECHA).



Por lo tanto, puede concluirse que los usuarios de PA tienden a dedicar más sesiones de estudio al MOOC que los de IP.

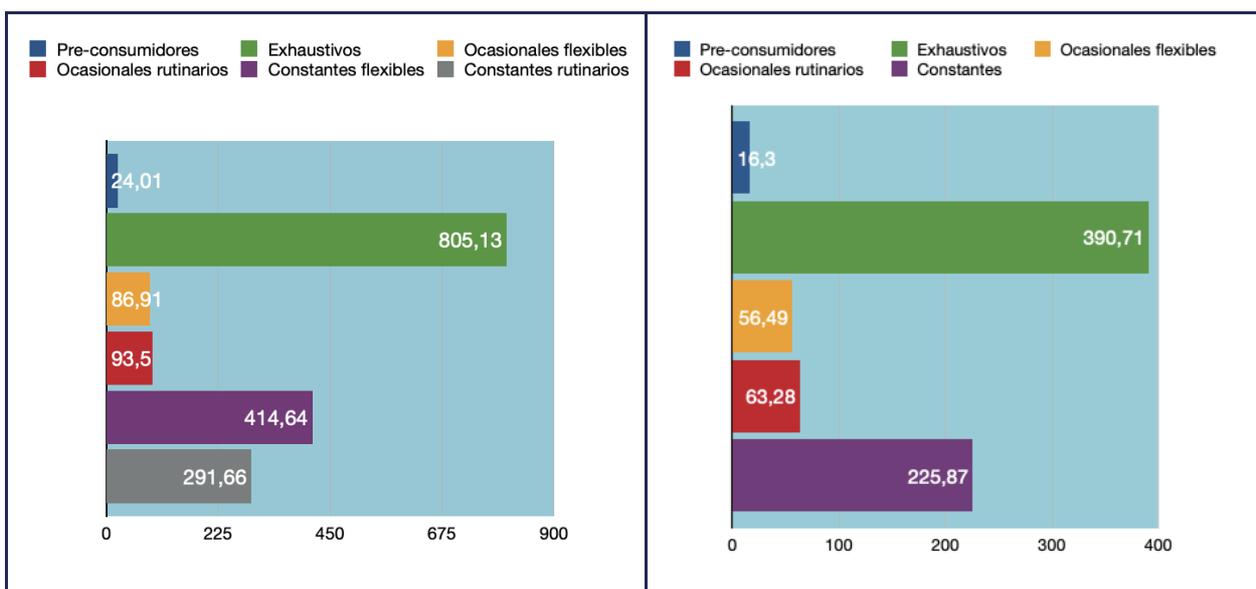
¿QUÉ DIFERENCIAS HAY RESPECTO AL TIEMPO TOTAL INVERTIDO EN EL CURSO?

Al igual que sucedía con la media del número de sesiones, todos los perfiles de participantes del LMOOC PA dedican más tiempo al curso que los del LMOOC IP.

De nuevo, la diferencia en el número total de minutos dedicados de promedio por los pre-consumidores de ambos cursos es baja, de apenas 8 minutos, aunque esta diferencia va incrementándose en el resto de los perfiles de forma proporcional al nivel de implicación de cada grupo, como puede observarse en el gráfico 18. De este modo, para los participantes ocasionales la diferencia es de unos 30 minutos adicionales de media en PA, de 65 para los constantes rutinarios, de unos 190 minutos para los constantes flexibles y de 415 para los exhaustivos.

Por lo tanto, puede afirmarse claramente que los participantes de PA dedican de promedio más tiempo al curso que los participantes de IP.

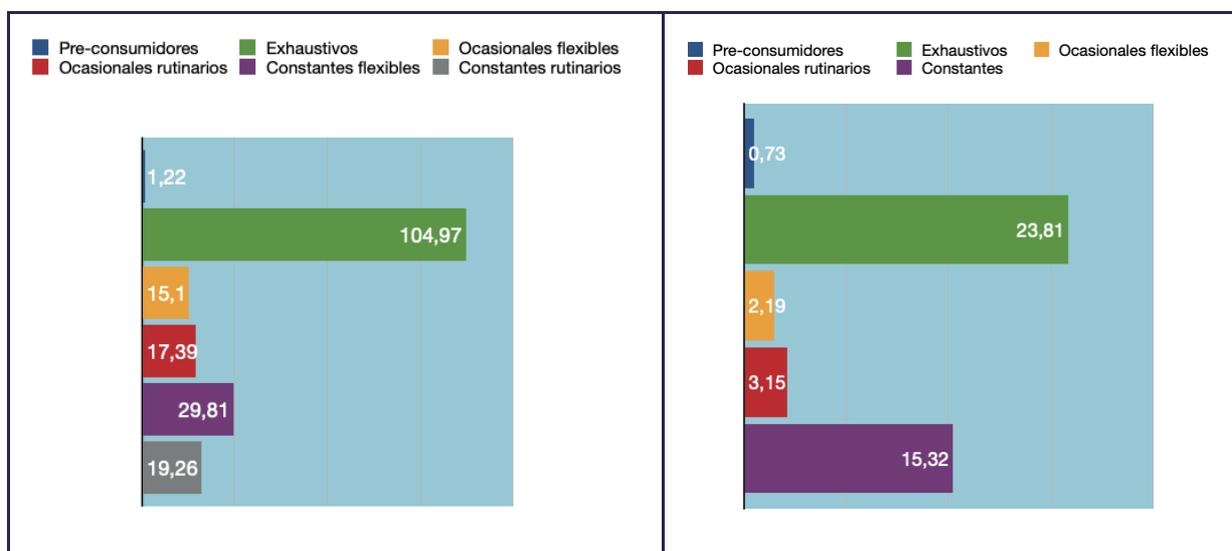
GRÁFICO 18. TIEMPO MEDIO DEDICADO AL CURSO POR LOS PARTICIPANTES DE PA (IZQUIERDA) E IP (DERECHA).



¿QUÉ DIFERENCIAS HAY RESPECTO AL ACCESO A LOS FOROS DEL CURSO?

De nuevo, según puede comprobarse en el gráfico 19, los estudiantes del LMOOC PA exhiben una mayor actividad en los foros que los de IP. El acceso a los foros es mayor en todos los perfiles, con cifras entre cinco y siete veces superiores en todos los grupos excepto en el perfil de participantes constantes, donde la diferencia sigue siendo significativa, aunque claramente inferior que en el resto de los perfiles.

GRÁFICO 19. NÚMERO MEDIO DE ACCESO A LOS FOROS PARA CADA PERFIL DE PARTICIPANTES DE PA (IZQUIERDA) E IP (DERECHA).



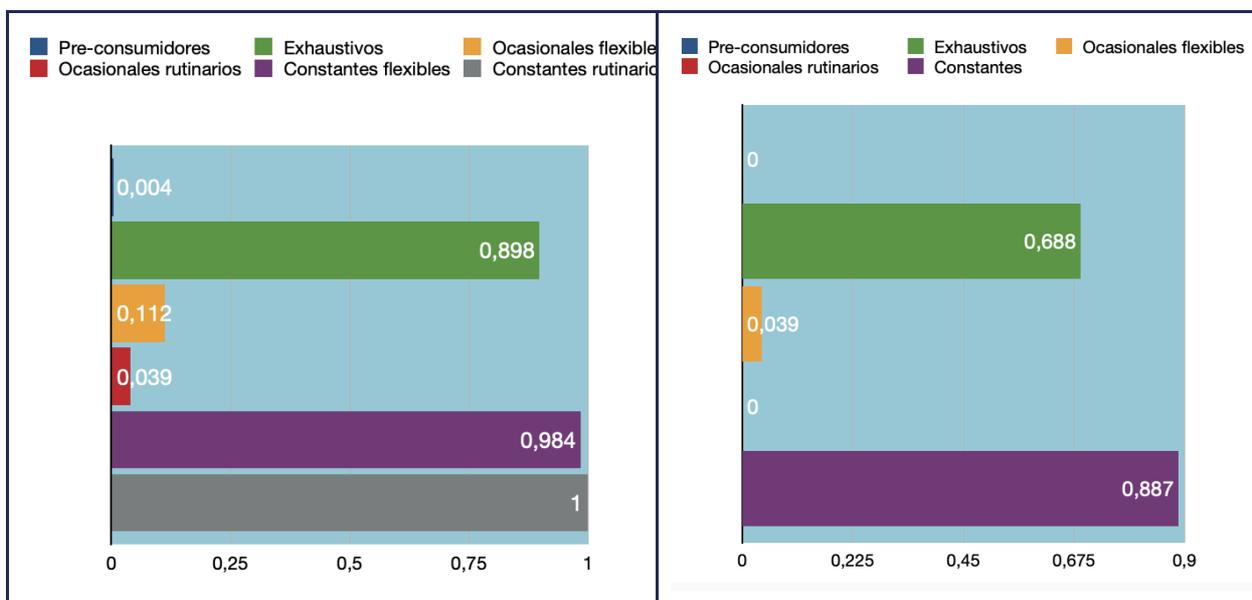
Por consiguiente, también en el caso del acceso a los foros puede afirmarse categóricamente que los participantes de PA son más activos que los participantes de IP.

¿QUÉ DIFERENCIAS HAY RESPECTO A LAS TASAS DE ELEGIBILIDAD PARA LA ACREDITACIÓN DEL CURSO?

Las tasas de elegibilidad para el certificado final también son superiores en el LMOOC PA para todos los perfiles, como bien puede observarse en el gráfico 20. Esta variable está correlacionada con la del desempeño en el curso, ya que, para que un participante sea elegible para el certificado final del curso, este debe tener una nota final igual o superior a 5.

Pues bien, todos los perfiles de PA, incluso el de pre-consumidores, tienen algún participante elegible para la obtención del certificado, mientras que en IP existen dos perfiles en los que ningún participante es elegible, el de pre-consumidores y el de participantes ocasionales rutinarios. En el perfil de participantes ocasionales flexibles, aunque la elegibilidad para el certificado final de PA es bajo, el porcentaje es de casi el triple que en IP, mientras que en los participantes exhaustivos la diferencia entre PA e IP es de más de 20 puntos porcentuales. Finalmente, en el perfil de participantes constantes, en PA el porcentaje de estudiantes elegibles roza el 100%, mientras que en IP es del 89%.

GRÁFICO 20. PROPORCIÓN DE PARTICIPANTES ELEGIBLES PARA EL CERTIFICADO FINAL PARA CADA PERFIL DE PARTICIPANTES DE PA (IZQUIERDA) E IP (DERECHA).



En resumen, puede afirmarse claramente que las tasas de elegibilidad para el certificado final son superiores en PA para todos los perfiles de participantes. De nuevo es relevante destacar que, como puede observarse en el gráfico 15 previamente representado, el nivel educativo de los participantes de IP es notablemente superior a los de PA y, aun así, los resultados obtenidos por los participantes en el curso son inferiores. Esto refuerza el indicio de que los mejores resultados por parte de los participantes de PA se deben a causas motivacionales y a un mayor tiempo disponible para dedicarlo al curso.

¿QUÉ RASGOS INHERENTES AL DISEÑO DE LOS CURSOS PRODUCEN TALES DIFERENCIAS?

Como se analizó en el capítulo sexto, los dos cursos comparten algunas características de diseño, como la división en módulos temáticos; la variedad en la tipología de actividades; el empleo de vídeos para ofrecer los contenidos, o el uso de foros. Sin embargo, ciertas diferencias en ambos LMOOC pueden ayudar a explicar los mejores resultados de los participantes de PA. En primer lugar, la duración de PA es significativamente mayor que la de IP, con cuatro módulos de contenidos divididos en seis semanas en PA y tres módulos divididos en tres semanas para IP. El que la duración de PA fuera del doble puede haber favorecido un incremento en los accesos al foro, en el número de sesiones y en el tiempo dedicado al curso por parte de sus estudiantes. En segundo lugar, en PA se realizó un esfuerzo consciente por facilitar la descarga de los materiales del curso, en especial los vídeos, los cuales estuvieron disponibles en una menor resolución. De este modo, es posible que el tiempo invertido por los participantes de PA esté subestimado y haya sido incluso mayor que el registrado.

La dispersión de las horas de inicio de las sesiones es la única variable que no presenta grandes diferencias en los perfiles de ambos cursos. En el único grupo en el que se observa una diferencia reseñable es en el de los participantes exhaustivos, que para PA muestra mayores valores de variabilidad que en IP. Este comportamiento es esperable, ya que el número de sesiones promedio de los participantes exhaustivos de PA (40) es significativamente mayor que los de IP (24). En consecuencia, es lógico que, ante un mayor número de sesiones dedicadas al curso por parte de este perfil de participantes, favorecida por una

mayor duración del curso, la probabilidad de que estas sesiones se sitúen en franjas horarias distintas sea mayor.

En conclusión, aunque las características del LMOOC PA parecen favorecer la dedicación de un tiempo mayor al curso, este hecho por sí solo no explicaría por qué los participantes de PA presentan valores notablemente superiores a los de IP, a pesar de que su nivel educativo es, como promedio, considerablemente inferior. La respuesta hay que buscarla, por lo tanto, en las características del alumnado.

¿QUÉ RASGOS DEL ALUMNADO DE LOS CURSOS PRODUCEN TALES DIFERENCIAS?

Las preguntas anteriormente analizadas confirman que existen diferencias significativas entre ambos LMOOC respecto a los valores de la mayoría de las variables, más específicamente en el número de sesiones; el tiempo total en el curso; la nota final; el acceso a los foros, y las tasas de elegibilidad en el curso, con valores superiores en el LMOOC PA a pesar de un nivel educativo medio más bajo de sus participantes.

El LMOOC IP estaba dirigido y diseñado, en general, para trabajadores con un nivel B1 de inglés. A modo de prueba piloto, también se recompensó con 1 crédito ECTS a aquellos estudiantes del grado de turismo de la UNED que obtuvieron un certificado en el curso. Por consiguiente, el público objetivo de este LMOOC fueron estudiantes universitarios y trabajadores con un nivel medio del idioma; trabajadores que en un gran porcentaje poseían, como se ha indicado anteriormente, estudios superiores. En consecuencia, tomando en cuenta el perfil personal de los participantes del LMOOC IP, puede concluirse que los individuos pertenecientes a estos perfiles se caracterizan por tener un tiempo limitado para dedicar al MOOC, ya que deben compatibilizarlo con su empleo y/o con sus estudios. De las circunstancias personales del perfil medio de participantes de IP se deduce, por lo tanto, que la posibilidad y la necesidad de profundizar en los contenidos del curso tienden a ser limitadas, a pesar de que se supone que estos estudiantes han adquirido un alto nivel de autorregulación del proceso de aprendizaje a lo largo de su vida educativa.

Por el contrario, el público objetivo para el que se diseñó el curso PA es el de participantes migrantes que, debido a sus circunstancias personales, se ven obligados a aprender español. Por lo tanto, este perfil de estudiantes presenta una motivación más acusada y una necesidad inmediata de asimilar los contenidos del curso y profundizar en ellos. De manera adicional, las tasas de ocupación entre la población migrante suelen ser bajas, en gran medida por las dificultades con el idioma, y por las trabas legales y burocráticas para obtener permisos de trabajo. Menores tasas de ocupación se traducen en mayor tiempo disponible, lo cual favorece un mayor compromiso con el curso y más tiempo dedicado a la interacción con los contenidos del LMOOC. A la luz de las estadísticas analizadas anteriormente, el tiempo dedicado al LMOOC revierte positivamente en las calificaciones y en el porcentaje de estudiantes que superan el curso.

Como se ha explicado previamente, todos los perfiles de PA dedican al curso una media de tiempo superior a los de IP, además de un mayor número de sesiones. También son notablemente superiores sus cifras de acceso a los foros, ya que estos se convirtieron en un lugar fructífero para el intercambio de vivencias y de ideas. Y también es superior el porcentaje de estudiantes pertenecientes a perfiles de participantes que se caracterizan por una mayor vinculación al curso: los participantes exhaustivos y los constantes. Por lo tanto, la evidencia apunta a que las diferencias en la prevalencia de los perfiles de participantes, con mayor porcentaje de estudiantes con niveles altos actividad y desempeño en PA, y las diferencias en las notas medias; el acceso a los foros, y las tasas de elegibilidad para el certificado final, con resultados superiores también para PA a pesar de un nivel educativo medio inferior de sus estudiantes, se deben a las características intrínsecas del alumnado objetivo de cada curso, lo cual favorece una mayor implicación de los participantes del LMOOC PA.

8.3.3. CONFIRMACIÓN DE LA HIPÓTESIS

La hipótesis planteada inicialmente es la siguiente: «es posible identificar mediante LA la taxonomía de perfiles de usuario de LMOOC basada en patrones temporales de uso descrita por del Peral (2019) para el LMOOC IP en un nuevo LMOOC, PA, pero los niveles de prevalencia y desempeño variarán en

función de las características particulares de cada LMOOC y de sus participantes». Una vez respondidas las preguntas de la investigación, tanto las generales como las específicas, puede confirmarse que la primera parte de la hipótesis se ha verificado parcialmente, ya que ha sido posible identificar la taxonomía de perfiles de usuario de LMOOC basada en patrones temporales de uso descrita por del Peral en el LMOOC PA, pero no de forma unívoca. La discrepancia en el número de perfiles se debe a que dos de los perfiles identificados en el nuevo LMOOC, PA, son en efecto dos sub-perfiles de uno de los perfiles descritos en la taxonomía original de del Peral. Respecto a la segunda parte de la hipótesis, se ha verificado que, en efecto, las tasas de prevalencia y desempeño de los LMOOC PA e IP son distintas y que tales diferencias son atribuibles a una mayor duración del curso PA y, especialmente, a las características particulares del respectivo alumnado.

Como parte de la metodología del proyecto de LA, se ha ejecutado el algoritmo k-medias a partir de los datos originales del LMOOC IP, con el objetivo de identificar mediante una técnica alternativa de LA la taxonomía de perfiles descrita por del Peral. En consecuencia, resulta procedente extender la hipótesis planteada inicialmente para el LMOOC PA también al LMOOC original IP, por lo que, puede también confirmarse de forma parcial que, al emplear una técnica alternativa de LA, ha sido posible identificar en el LMOOC IP la misma taxonomía de perfiles de usuario de LMOOC basada en patrones temporales de uso descrita por del Peral y hallada en PA. Tras la ejecución del algoritmo k-medias para IP, se ha identificado el mismo número de perfiles descritos originalmente por del Peral, aunque de nuevo no se ha hallado una correspondencia unívoca. El motivo es que, en primer lugar y al igual que sucedía en PA, dos de los grupos identificados en IP se consideran como dos subgrupos de una de las categorías de del Peral y, en segundo lugar, dos de los grupos identificados por del Peral, el de los participantes rutinarios y el de los participantes flexibles, se consideran subgrupos de la categoría de participantes constantes, la cual aparece como una categoría única en el LMOOC IP.

En conclusión, aunque no se ha encontrado en los clústeres resultantes del algoritmo k-medias una coincidencia unívoca con los perfiles definidos por del Peral, las diferencias halladas se han explicado por la existencia de distintos

subgrupos, de tal modo que los rasgos de los perfiles descritos en la taxonomía original de del Peral encuentran su correspondencia en alguno de los grupos resultantes de la ejecución del algoritmo k-medias en ambos LMOOS. Por lo tanto, puede considerarse probada la hipótesis inicial.

8.4. LIMITACIONES, APORTACIONES DEL ESTUDIO Y FUTURAS LÍNEAS DE INVESTIGACIÓN

8.4.1. LIMITACIONES

Como es común en los trabajos de investigación, este estudio presenta una serie de limitaciones que es necesario explicitar, ya que los resultados del trabajo deben ser interpretados teniendo en cuenta las circunstancias exactas en las que la metodología es aplicada.

Una de las limitaciones, la cual ya se explicó en el capítulo seis, más concretamente en la sección dedicada a la fase de pre-procesado de los datos, es la pérdida de datos debido a la corrupción de algunas entradas en los ficheros del curso PA. Durante la fase de pre-procesado se identificaron entradas incompletas en los ficheros de *log* que se intentaron reparar manualmente si el fragmento presente en el fichero de *log* contenía los campos que eran leídos dentro del programa. De lo contrario, si alguno de los campos que se utilizaban en el programa no estaba presente o estaba incompleto, el objeto JSON se eliminaba del fichero. La consecuencia fue la pérdida de una decena de las acciones que los participantes realizaron en el LMOOC. La incidencia de estos datos que no pudieron repararse es realmente insignificante por dos motivos principales. En primer lugar, el valor de estos datos es mínimo entre los casi dos millones de objetos JSON almacenados en los ficheros de *log*. Por otro lado, la pérdida de estos datos podría producir una diferencia en la agrupación de una muestra únicamente si el evento registrado perteneciera a una sesión compuesta de un único evento, algo muy poco probable. Si el evento omitido está localizado entre otros dos eventos de una sesión, no se pierde información para el estudio, y si es el último o el primer evento de una sesión, esta sesión será registrada como más corta de lo que realmente es. Esto último tampoco impacta en la identificación de perfiles, ya que la variable

empleada a tal efecto es el número de sesiones. Por lo tanto, puede afirmarse con total seguridad que esta pérdida de datos no ha tenido impacto en la identificación de los distintos perfiles.

Otra de las limitaciones de este estudio, que en este caso sí ha podido tener incidencia en la identificación de los perfiles de participantes, es la sobreestimación y la subestimación del número y la duración de las sesiones.

Los casos que han podido producir una subestimación del número y la duración de las sesiones son varios. En primer lugar, dentro del LMOOC IP se incluía una actividad especial voluntaria basada en la descarga y utilización de una aplicación de podcasts, la cual, aunque no se tomaba en cuenta para el cálculo de la calificación final del curso, debería haberse contabilizado para el cálculo de las sesiones y de los minutos invertidos en el LMOOC. Sin embargo, dicha aplicación era ajena a la plataforma donde se alojó el curso, por lo que no se tuvo acceso a los *logs* generados por dicha aplicación. Por lo tanto, la interacción con dicha aplicación no se ha tenido en cuenta para el análisis de los datos de uso de los usuarios y el consiguiente trazado de los perfiles, de modo que es posible que se hayan infravalorado algunos de los perfiles del LMOOC IP. En segundo lugar, es imposible contabilizar el tiempo que los participantes dedican al estudio del material descargado. Esto es especialmente relevante en el caso del curso PA, ya que en el diseño se priorizó que los participantes pudiesen avanzar en el curso en los momentos en los que no dispusieran de conexión a internet. En consecuencia, puede que haya perfiles dentro del LMOOC PA que también estén infrarrepresentados, en especial el de los participantes exhaustivos. También es probable que las estadísticas de los participantes constantes estén subestimadas y el número medio de sesiones y el tiempo total dedicado al curso por parte de estos perfiles sea en realidad superior. Por último, en los *logs* de la plataforma solo se registra la última acción realizada por los participantes, por ejemplo, pulsar el botón de reproducir un vídeo, acceder a una página del curso, etc. El tiempo que un participante dedica a esa acción no se registra, por lo que la duración real de todas las sesiones de ambos LMOOC siempre será ligeramente superior a la calculada. Como este fenómeno es transversal a todos los perfiles y a los dos LMOOC, y

como la variable afectada es el tiempo de la sesión y no el número de sesiones, no supone un impacto en el cálculo de los perfiles.

Los casos que han podido producir una sobrestimación en el número y la duración de las sesiones son principalmente los de participantes que están realizando multitarea, esto es, están interactuando con más de una página o aplicación, o realizando más de una acción mientras tienen una sesión iniciada en el LMOOC. Cuando dichos usuarios intercalan acciones ajenas al curso con las sesiones de estudio, la duración de las sesiones se incrementa de forma artificial. De manera análoga, si estas actividades paralelas se extienden durante un tiempo superior a los 30 minutos fijados como umbral para diferenciar las distintas sesiones, se registrará el inicio de una nueva sesión, aunque pueda ser razonable considerar que la nueva acción es en realidad una extensión de la sesión original. Este fenómeno es más probable que se produzca en aquellos perfiles que presentan mayores tasas de actividad en el curso: los participantes exhaustivos y los constantes. El efecto que puede producirse es que se haya incrementado el número de sesiones y el tiempo total en esos grupos, o que haya habido un trasvase mínimo del grupo de participantes constantes al de los exhaustivos. Sin embargo, debido a que este fenómeno es común a los dos LMOOC, no supone un impacto determinante en el estudio.

Por último, es necesario recordar que, aunque en el diseño de los LMOOC PA e IP se intentó fomentar la interacción y la variedad en el tipo de actividades de evaluación, el modelo de ambos cursos es el de xMOOC, claramente el tipo más numeroso entre los MOOC ofertados actualmente. Como se explicó en el capítulo tercero, esta tipología de MOOC es de naturaleza tecno-céntrica, tanto en su diseño y estructura como en su filosofía, ya que tiende a adaptar los formatos de aprendizaje desarrollados tradicionalmente por las universidades en el marco de sus acciones de *e-learning* (Vázquez et al., 2013; Cabero et al., 2014; Jacoby, 2014). En consecuencia, en los xMOOC las vídeo clases y las herramientas de evaluación estandarizadas y automatizadas desempeñan un papel muy significativo (Cabero et al., 2014). Las diferencias entre los xMOOC y la otra gran categoría de estos cursos, los cMOOC, son notables. En los cMOOC el desarrollo del conocimiento; la creatividad, y la comunicación de los

participantes se efectúa fuera de la plataforma que aloja el MOOC, a través de elementos de internet tales como blogs; redes sociales; wikis; mundos virtuales, o webs de alojamiento de vídeos. Los estudiantes acceden a la información, reflexionan sobre ella, entregan sus actividades y con ello se retroalimenta el proceso (Masters, 2011; Clark, 2013; Seidametova, 2018). La abrumadora escasez de LMOOC de esta tipología y la enorme dificultad para obtener datos educativos de esta modalidad de MOOC ha imposibilitado la inclusión de uno de estos cursos en el estudio. Por lo tanto, aunque los datos obtenidos en este estudio indican que es previsible que la taxonomía descrita sea inherente a todos los LMOOC, incluso a todos los MOOC independientemente de su temática, es necesario confirmar dicha hipótesis con un estudio específico con datos obtenidos de un cMOOC de lenguas.

8.4.2. APORTACIONES DEL ESTUDIO

Vivimos en una sociedad marcada por la aceleración, en la que los ciudadanos sienten la presión continua de incrementar su productividad y aprovechar al máximo el tiempo disponible. Esta necesidad constante de priorizar la productividad implica que, incluso fuera de los entornos laborales y educativos, los individuos se sientan condicionados a maximizar sus logros y su rendimiento (Concheiro, 2016; Kohli, 2020; Skliar, 2021). La ocupación y la falta de tiempo incluso se han convertido en un signo de prestigio (Bellezza et al, 2017).

Por otra parte, esta celeridad de nuestra sociedad moderna subraya la importancia del aprendizaje a lo largo de toda la vida como mecanismo de adaptación. El aprendizaje a lo largo de toda la vida se refiere a los aprendizajes en todos los espacios y mediante todos los recursos disponibles, combinando aprendizajes en entornos formales; no-formales, e informales. Según la UNESCO-UIL el aprendizaje a lo largo de la vida está «basado en valores de inclusión, emancipación, humanísticos y democráticos; es global y parte integrante de la perspectiva de una sociedad basada en el conocimiento» (2010, p.28). Por lo tanto, la UNESCO plantea el aprendizaje a lo largo de la vida como un enfoque para la política educativa y como un nuevo paradigma para la educación en el siglo XXI, ya que supone un medio de mejora de la calidad de vida de las

personas y de contribución a la igualdad social (UNESCO, 2020). En este contexto, los MOOC, un movimiento respaldado por organismos internacionales como la ONU, se erigen como un fenómeno de globalización; democratización, y apertura del conocimiento que cuenta con la potencialidad de satisfacer un derecho humano fundamental como es el derecho a la educación (Ruiz Bolívar, 2015). El objetivo del presente trabajo es, por lo tanto, el de contribuir a salvar la brecha entre la necesidad de formación del ciudadano moderno, la cual tiene en los MOOC una herramienta de gran potencia y versatilidad, y su creciente dificultad para gestionar un menor tiempo libre disponible para dedicar a los cursos.

En los últimos años, el fenómeno de los MOOC ha despertado el interés de la comunidad científica y han proliferado las investigaciones que tienen a estos cursos como objeto de estudio. De este modo, tal y como se describió en el capítulo cinco, se han descrito numerosas taxonomías de participantes de MOOC con el objetivo de arrojar luz sobre las preferencias y los hábitos de los estudiantes, y de esta forma maximizar su desempeño en los cursos. La mayoría de estas taxonomías describen un grado cuantitativo o cualitativo de interacción con ciertos elementos del MOOC, aportando información valiosa sobre el grado de vinculación y compromiso de los participantes de los MOOC, y del tipo de contenido preferido por los mismos. Sin embargo, son pocas las clasificaciones que estudian algún aspecto que revele cómo los participantes de MOOC gestionan el tiempo que emplean en el curso. De este modo, un usuario que completa un MOOC en un par de sesiones intensivas, realizando todas las actividades y participando en los foros, sería categorizado de la misma forma que un usuario que programe la realización del curso en las mismas franjas horarias, o que un usuario que dedica decenas de sesiones a revisar los contenidos del curso, aunque claramente se trate de perfiles de estudiantes con características y preferencias de aprendizaje distintas. Por este motivo, la investigación llevada a cabo en esta tesis supone una aproximación a un área de estudio con escasa literatura científica: el perfilado de participantes de LMOOC en función a patrones temporales de interacción con los cursos.

La identificación de perfiles de usuario de LMOOC basados en patrones temporales de uso tiene como principal objetivo posibilitar el diseño

metodológico de LMOOC teniendo en cuenta los rasgos de cada tipo de estudiante, de tal forma que los cursos se adapten a las preferencias de aprendizaje específicas y a las necesidades de cada perfil de participantes. La identificación de perfiles podría combinarse con encuestas iniciales que permitieran conocer los objetivos y motivaciones de los alumnos y, lo que es más relevante para este estudio, sus limitaciones de tiempo, y sus horarios y hábitos de estudio. De esta forma, el perfilado de los participantes se podría realizar en una etapa incipiente del MOOC, de tal modo que los estudiantes pudieran recibir ayuda individualizada desde el mismo comienzo del curso. Mediante las técnicas de LA, se puede monitorizar la actividad de los estudiantes a la luz de los datos recogidos en la encuesta y así implementar mecanismos de intervención dirigidos a los alumnos que no están cumpliendo con los objetivos que se trazaron inicialmente.

En su trabajo de revisión bibliográfica sobre las tasas de deserción en los MOOC, Bezerra y da Silva (2017) enumeran los 24 motivos principales identificados en la literatura científica por los que los participantes abandonan los cursos, entre los que se encuentran los siguientes:

- Falta de tiempo para dedicar al curso.
- Dificultad para gestionar el tiempo.
- Alta carga de trabajo.
- Falta de urgencia para finalizar el curso.

Estos cuatro motivos enumerados están íntimamente relacionados con los patrones temporales analizados en este trabajo de investigación. El perfilado de estudiantes realizado en esta tesis permitiría asociar estos motivos a grupos específicos de participantes, de tal modo que puedan emplearse las LA para establecer estrategias de identificación temprana de los alumnos pertenecientes a los perfiles con menor rendimiento y mayores tasas de deserción. De este modo, se facilitaría la implementación de mecanismos de intervención que prevengan el abandono entre estos grupos de estudiantes. Algunas de estas estrategias podrían ser sistemas de recompensa; recordatorios por email y notificaciones; ayuda personalizada mediante sistemas de inteligencia artificial; guías de gestión del tiempo basadas en el historial del

alumno, o modificaciones de los caminos de aprendizaje que permitieran a los participantes completar el curso con éxito.

Una forma en que los distintos patrones de gestión del tiempo por parte de los estudiantes pueden ser tomados en cuenta a la hora de diseñar los MOOC es la inclusión de distintas tipologías de actividades. Actividades de duración más corta pueden ser particularmente apropiadas para los perfiles de usuarios ocasionales, para de esta forma incrementar su motivación, e incluso para los participantes exhaustivos, que tienden a dedicar al curso sesiones de duración más corta. Por otro lado, actividades que conlleven más dedicación temporal pueden ser adecuadas para los perfiles constantes, que exhiben mayores tasas de tiempo en el curso y mejores resultados. Esta misma flexibilidad es aplicable a los contenidos, en particular los vídeos, los cuales podrían dividirse o acortarse en función de los distintos perfiles. De este modo, podrían crearse rutas de aprendizaje individualizadas en función de la tipología de participantes, cada una con las actividades y la forma de presentación de contenidos óptimas para cada perfil.

Uno de los perfiles a los que podría aplicarse dichas estrategias es el de los participantes exhaustivos. Este trabajo ha confirmado lo que era un indicio en el trabajo inicial de del Peral (2019), que una mayor dedicación en tiempo, sesiones y actividad en los foros por parte de este grupo de participantes no se traduce en mejores calificaciones, ya que, aunque el rendimiento de este grupo de estudiantes es alto, es superado tanto por el de los participantes constantes del LMOOC IP como por los participantes constantes rutinarios y flexibles en el LMOOC PA. Este menor rendimiento de los participantes exhaustivos es un fenómeno contraintuitivo, ya que un mayor tiempo dedicado al curso debería traducirse en un mayor rendimiento y unas mejores calificaciones. De nuevo, la detección temprana de los individuos pertenecientes a este grupo permitiría la introducción de mecanismos de recompensa que fomenten que los participantes maximicen el rendimiento del tiempo dedicado al curso. Podrían establecerse para este grupo de participantes micro-pruebas o micro-tareas regulares que contribuyeran a una mejor asimilación de los conceptos impartidos en el curso y a la identificación de las posibles dificultades con ciertos contenidos del MOOC. Adicionalmente, este grupo de estudiantes con grandes

tasas de participación en los foros serían los candidatos perfectos para el establecimiento de programas de tutorización por pares que permitieran proporcionar ayuda más personalizada a participantes de grupos con tasas menores de rendimiento, como los ocasionales. Este tipo de programas ha sido implementado en ciertos MOOC y han contribuido a la mejora de las tasas de finalización y del grado de satisfacción de los alumnos (Garreta et al., 2015; Dhorne et al., 2017).

Junto con los participantes exhaustivos, el grupo de participantes constantes es el que exhibe mayores niveles de actividad y capacidad de autorregulación durante el transcurso del MOOC. La identificación temprana de los integrantes de ambos grupos permitiría que también pudieran ser objeto de intervenciones destinadas a incrementar su participación en las discusiones de los foros o a la realización de actividades desafiantes, con un mayor nivel de dificultad, de forma que se fomente la motivación entre los participantes de estos perfiles. Uno de los mecanismos que podrían emplearse para incrementar la vinculación de los participantes del curso son la gamificación y la concesión de insignias, por ejemplo, al completar un cierto número de cuestionarios; al contribuir en los foros con un cierto número de comentarios, o al alcanzar una nota determinada en una actividad. Estas estrategias favorecerían una mejor asimilación de los contenidos y, por tanto, un mayor rendimiento de los participantes, además de contribuir a la reducción de las tasas de abandono.

Por último y aunque no era el objetivo principal de esta tesis, este trabajo también arroja luz sobre el papel de la motivación como factor de éxito por parte de los participantes de los MOOC. Aunque hay autores que sugieren que este tipo de cursos es más adecuado para individuos con educación superior, ya que sus estrategias de autorregulación en el aprendizaje están más desarrolladas (Cassidy et al. 2014), las conclusiones de este estudio muestran que participantes con, *a priori*, mayores niveles de motivación, como los del LMOOC PA, obtienen mejores resultados a pesar de poseer un nivel educativo claramente inferior a los del otro LMOOC analizado, IP.

FIGURA 26. POSIBLES ADAPTACIONES QUE UNA IDENTIFICACIÓN TEMPRANA DE PERFILES PERMITE HACER DE LA RUTA DE APRENDIZAJE DE LOS PARTICIPANTES DE MOOC



8.4.3. FUTURAS LÍNEAS DE INVESTIGACIÓN

Durante los últimos años, los MOOC han probado ser un fructífero campo de estudio en el sector de la educación, con numerosas investigaciones que ofrecen perspectivas originales y enriquecedoras sobre esta iniciativa que busca impartir educación de calidad de forma masiva y de manera gratuita. Aunque el presente trabajo profundiza en un aspecto de los MOOC sobre el que aún no existía conocimiento suficiente, esto es, el perfilado de participantes basado en patrones temporales, esta tesis supone una primera aproximación al estudio de los hábitos temporales de los estudiantes de MOOC. Este análisis, por

lo tanto, abre la puerta a distintas líneas de investigación que puedan continuar profundizando y ampliando el conocimiento sobre esta perspectiva particular de los MOOC.

La línea más inmediata y evidente es la replicación del presente estudio con los datos de tipologías de MOOC distintas a las analizadas en este trabajo. Aunque hay sobrados indicios de que la taxonomía identificada puede ser generalizable a todos los LMOOC e incluso a todos los MOOC, se necesitan estudios adicionales que lo confirmen. Por lo tanto, un siguiente paso sería la realización del mismo proceso de análisis, siguiendo la misma metodología de LA, a partir de los datos de un MOOC de temática distinta al aprendizaje de lenguas. También sería sumamente interesante ampliar el enfoque del estudio con MOOC de tipo conectivista, ya que estos presentan unas características muy particulares. Sin embargo, la enorme distribución de la actividad de este tipo de MOOC dificulta en gran medida, si no imposibilita, la recogida de datos educativos en los que basar las investigaciones.

Por otro lado, el algoritmo empleado para la identificación de los distintos perfiles ha sido el algoritmo k-medias, un algoritmo robusto utilizado frecuentemente en la literatura científica para el perfilado de participantes de MOOC (Kizilcec et al., 2013; Ferguson y Clow, 2015; Kovanovic et al., 2016; Rodrigues et al., 2016; Arora et al., 2017; Khalil y Ebner, 2017; Maya-Jariego et al., 2020). Por consiguiente, una futura línea de investigación sería comprobar si el empleo de un algoritmo alternativo sobre los mismos datos produce clústeres o agrupamientos similares. Sería interesante el análisis mediante un método de partición distinto, por ejemplo, el algoritmo k-medoids, o mediante otros tipos de algoritmos de descubrimiento de estructuras, como los métodos basados en densidad o, muy especialmente, los métodos jerárquicos, ya que varios de los grupos identificados en este trabajo presentan una estructura jerárquica dividida en subgrupos. Adicionalmente, podría realizarse una triangulación de la metodología empleada en este trabajo con otros enfoques de investigación, tales como encuestas y entrevistas, de manera que se pueda profundizar en el conocimiento de las rutinas de estudio de los participantes y de cómo estas impactan en su rendimiento.

La identificación de grupos de usuarios con menor implicación temporal en los cursos, debido principalmente a un abandono temprano, también abre las puertas a posteriores investigaciones. La taxonomía descrita puede ser la base para la identificación de los perfiles de estudiantes que necesitan apoyo externo o incentivos para avanzar en el curso. Podría estudiarse el comportamiento específico de interacción con los cursos durante los primeros días de impartición de los MOOC, con el objetivo de identificar rasgos comunes entre los perfiles con menor vinculación que permitan detectar de forma temprana a los individuos en riesgo de abandono.

De manera adicional, los motivos de deserción de los cursos relacionados con factores temporales, como la diferencia en los husos horarios o un inicio tardío en el curso por parte de los estudiantes (Bezerra y da Silva, 2017), también pueden ser objeto de nuevas investigaciones. Estos estudios podrían tener como meta el análisis de la incidencia de estos factores en las tasas de abandono o en la pertenencia a determinados grupos de la taxonomía.

Otra posibilidad de estudio dentro del tema definido por esta tesis sería analizar cómo distintas estrategias y técnicas educativas, tales como la gamificación; la metodología *flipped* aplicada a los MOOC; las insignias, o la tutorización inciden en la prevalencia y en el rendimiento de los distintos perfiles (Anderson et al., 2014; Klemke et al, 2018; Saputro et al., 2019; Sunar et al., 2020).

Por último, ante los mejores resultados exhibidos por los participantes del LMOOC PA, aunque sus participantes contasen con un menor nivel medio de estudios que los de IP, este trabajo abre la puerta al empleo de los LMOOC como complemento metodológico de cursos de niveles pre-universitarios, más concretamente de secundaria y de ciclos formativos. Al igual que el LMOOC IP se propuso como complemento a una asignatura del Grado de Turismo de la UNED, podría investigarse sobre el uso de MOOC como medida de atención a la diversidad, bien como estrategia de refuerzo para alumnos con problemas de base en ciertas materias, bien como material de ampliación para alumnos aventajados o de altas capacidades.

BIBLIOGRAFÍA

- Abe, K. (2004). Japanese attitudes towards foreign languages. *Collegium antropologicum*, 28(1), 115-122.
- Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Parada, H. A. y Muñoz-Organero, M. (2014). Delving into Participants' Profiles and Use of Social Tools in MOOC. *IEEE Transactions on Learning Technologies* 7 (3), 260-266. doi: 10.1109/TLT.2014.2311807
- Alboukadel (2018). Cluster Analysis in R Simplified and Enhanced [Entrada de blog]. *Datanovia*. Recuperado el 1 de marzo de 2022 de <https://www.datanovia.com/en/blog/cluster-analysis-in-r-simplified-and-enhanced/>
- Aldowah, H., Al-Samarraie, H. y Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13-49.
- Alexander, B., Ashford-Rowe, K., Barajas-Murphy, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R. y Weber, N. (2019). *EDUCAUSE Horizon Report: 2019 higher education edition*. Louisville: EDUCAUSE. Recuperado el 7 de junio de 2022 de <https://library.educause.edu/resources/2019/4/2019-horizon-report>
- Álvarez, D. (2 de abril de 2012). Crypt4you, el MOOC de la Universidad Politécnica de Madrid [Entrada de blog]. *e-aprendizaje*. Recuperado el 7 de junio de 2022 de <https://e-aprendizaje.es/2012/04/02/crypt4you-el-mooc-de-la-universidad-politecnica-de-madrid/>
- Alvarez, I. (2014). Ethical and aesthetic considerations in language MOOCs. En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOCs: Providing learning, transcending boundaries* (pp. 127–142). Berlin: De Gruyter Open.
- Amorós-Negre, C., Recio Diego, Á. y Tomé Cornejo, C. (2018). La calidad de los MOOC como reto para la enseñanza de lenguas en entornos digitales. En: Fernández Juncal, C. y N. Hernández Muñoz (eds.) *Vías de transformación en la enseñanza de lenguas con mediación tecnológica. Circulo de Lingüística Aplicada a la Comunicación* 76, 49-66 <http://dx.doi.org/10.5209/CLAC.62497>
- Anderson, A., Huttenlocher, D., Kleinberg, J. y Leskovec, J. (2014). Engaging with massive online courses. *Proceedings of the 23rd International Conference on World wide web* (pp. 687–698). New York: ACM. doi: 10.1145/2566486.2568042
- Anderson, M. y Gavan, C. (2016). *Developing Effective Educational Experiences through Learning Analytics*. Hershey, PA: IGI Global.
- Andrade, A., Ehlers, U. D., Caine, A., Carneiro, R., Conole, G., Kairamo et al. (2011). *Beyond OER: Shifting focus to open educational practices*. Open Education Quality Initiative. Recuperado el 7 de junio de 2022 de

<https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/OPAL2011.pdf>

- Area, M. y Adell, J. (2009). eLearning: Enseñar y aprender en espacios virtuales. En J. De Pablos (Coord). *Tecnología Educativa. La formación del profesorado en la era de Internet*. Málaga: Aljibe.
- Arneil, S. y Holmes, M. (2001). Hot Potatoes, History and Future. [Presentación]. Recuperado el 7 de junio de 2022 de <http://web.uvic.ca/hrd/eurocall2001/HotPotPastFuture/PastFutureHome.htm>
- Arnold, D. N. (1997). Computer-aided instruction. *Encarta encyclopédia*. New Mexico: Microsoft.
- Arora, S., Goel, M., Sabitha, A. S. y Mehrotra, D. (2017). Learner groups in massive open online courses. *American Journal of Distance Education*, 31(2), 80–97. doi: <https://doi.org/10.1080/08923647.2017.1300461>.
- Aviv, R., Erlich, Z., Ravid, G. y Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks*, 7(3), 1–23.
- Baker, R. S., Costa, E., Amorim, L., Magalhães, J., y Marinho, T. (2012). Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. *Jornada de Atualização em Informática na Educação*, 1, 1-29.
- Baker, R. S. e Inventado, P.S. (2014). Educational Data Mining and Learning Analytics. En J.A. Larusson y B. White (Eds.), *Learning Analytics: From Research to Practice* (pp. 61-75). Nueva York: Springer.
- Baker, R. S. y Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Bakharia, A. y Dawson, S. (2011). SNAPP: A bird's-eye view of temporal participant interaction. En D. Gasevic, G., Conole, G., Siemens y P., Long, P. (Eds). *LAK11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Banff, Canada, 27 February - 1 March 2011* (pp. 168–173). Nueva York: ACM.
- Barberá, E. (2008). *Aprender e-learning*. Paidós: Barcelona
- Bárcena, E., Martín-Monje, E., y Read, T. (2015). Potentiating the human dimension in Language MOOCs. En *EMOOCs 2015. Proceedings of the European MOOCs Stakeholders Summit 2015* (pp. 46-54). P.A.U. Education.
- Bárcena, E., Martín-Monje, E. y de la Torre, M. J. (2016). Innovación metodológica y tecnológica en la enseñanza del inglés para turismo a distancia. *Ibérica, Revista de la Asociación Europea de Lenguas para Fines Específicos*, (31), 39-61.

- Bárcena, E., Read, T. y Sedano, B. (2020). An Approximation to Inclusive Language in LMOOCs Based on Appraisal Theory. *Open Linguistics*, 6(1), 38-67. doi: <https://doi.org/10.1515/opli-2020-0003>
- Bárkányi, Z. (2018). Can You teach me to speak? Oral practice and anxiety in a language MOOC. En F. Rosell-Aguilar, T. Beaven & M. Fuertes Gutierrez (Eds.), *Innovative language teaching and learning at university: Integrating informal learning into formal language education* (pp. 9–16). Voillans, France: Research-publishing.net.
- Bárkányi, Z. y Melchor-Couto, S. (2017). Foreign Language Anxiety on a Massive Open Online Language Course. En K. Borthwick, L. Bradley & S. Thouësny (Eds.), *CALL in a climate of change: Adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 24–29). Voillans, France: Research-publishing.net.
- Bartalesi-Graf, D. (2017). From a Small Liberal Arts College to the World: Our Blended Courses, SPOC, and MOOCs in Italian Studies. En *Digital Education: Out to the World and Back to the Campus* (Vol. 10254, Lecture Notes in Computer Science, pp. 3-9). Cham: Springer International Publishing.
- BBC (s.f.). About BBC Learning English. Recuperado el 7 de junio de 2022 de <http://www.bbc.co.uk/learningenglish/english/hygiene>
- Beaven, T., Codreanu, T. y Creuzé, A. (2014). Motivation in a language MOOC: Issues for course designers. En E. Martín-Monje y E. Barcena (Eds.), *Language MOOCs: Providing learning, transcending boundaries* (pp. 48–66). Berlin: De Gruyter Open.
- Beetham, H., Falconer, I., McGill, L., y Littlejohn, A. (2012). *Open practices: Briefing paper*. JISC. Recuperado el 7 de junio de 2022 de <https://oersynth.pbworks.com/w/file/attach/58444186/Open%20Practices%20briefing%20paper.pdf>
- Beirne, E. (2017). LMOOCs, Classifying Design: Survey Findings from LMOOC Providers. *Research-publishing.net*, 2017-2023.
- Bellezza, S., Paharia, N., & Keinan, A. (2017). Conspicuous consumption of time: When busyness and lack of leisure time become a status symbol. *Journal of Consumer Research*, 44(1), 118-138.
- Benevolo, C. y Negri, S. (2007). Evaluation of Content Management Systems (CMS): a Supply Analysis. *The Electronic Journal Information Systems Evaluation* 10(1), 9-22.
- Bezerra, L. N., y da Silva, M. T. (2017). A review of literature on the reasons that cause the high dropout rates in the MOOCs. *Revista Espacios*, 38(05).
- Bienkowski, M., Feng, M. y Means, B. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. Recuperado el 7 de junio de 2022 de <https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf>

- Bliss, T. J. y Smith, M. (2017). A Brief History of Open Educational Resources. En R. S.Jhangiani y R. Biswas-Diener (Eds.) *Open: The Philosophy and Practices that are Revolutionizing Education and Science*. pp. 9–27. London: Ubiquity Press. doi: <https://doi.org/10.5334/bbc.b>.
- Blyth, C. (29 de mayo de 2013). The Mixxer Launches Spanish and English Language MOOCs [Entrada de blog]. COERLL. Recuperado el 1 de junio de 2022 de <https://blog.coerll.utexas.edu/update-on-the-mixxer/>
- Bolinger, D. (1934). Spanish on the air in Wisconsin. *Modern Language Journal*, 18(4), 217-221.
- Bourrel, J. R., Vidal, M. y Mahieux, F. (2008). *L'histoire du Cned Depuis 1939*. Ministère de l'éducation nationale et Ministère de l'enseignement supérieur et de la recherche. Recuperado el 7 de junio de 2022 de <https://docplayer.fr/11710781-L-histoire-du-cned-d-e-p-u-i-s-1-9-3-9.html>
- Britain, S. y Liber, O. (2004). A framework for pedagogical evaluation of virtual learning environments. JTAP. Recuperado el 7 de junio de 2022 de <http://www.jisc.ac.uk/media/documents/programmes/jtap/jtap-041.pdf>
- Brown, M., McCormack, M., Reeves, J., Brooks, D. C., Grajek, S., Alexander, B., Bali, M., Bulger, S., Dark, S., Engelbert, N., Gannon, K., Gauthier, A., Gibson, D., Gibson, R., Lundin, B., Veletsianos, G. y Weber, N. (2020). *EDUCAUSE Horizon Report, Teaching and Learning Edition*. Louisville: EDUCAUSE. Recuperado el 7 de junio de 2022 de https://library.educause.edu/media/files/library/2020/3/2020_horizon_report_pdf.pdf
- Bryman, A. (2012). *Social research methods (4th ed.)*. Oxford: Oxford University Press.
- Butcher, N. (2015). *Guía Básica de Recursos Educativos Abiertos (REA)*. París: UNESCO. Recuperado el 7 de junio de 2022 de <http://unesdoc.unesco.org/images/0021/002158/215804e.pdf>
- Cabero, J. (2015). Visiones educativas sobre los MOOC. *RIED. Revista Iberoamericana de Educación a Distancia*, 18 (2), 39-60. doi: <http://dx.doi.org/10.5944/ried.18.2.13718>
- Cabero, J., Llorente, M.C. y Vázquez, A.I. (2014). Las tipologías de MOOC: su diseño e implicaciones educativas. Profesorado. *Revista de Curriculum y formación del profesorado*, 18(1), 14-26.
- Cabrerizo Aparicio, C. (2018). ¿Por qué es necesario un cambio metodológico? *Revista Ventana Abierta*. Septiembre (30). Recuperado el 7 de junio de 2022 de <http://revistaventanaabierta.es/1536-2/>
- Calvet Liñán, L., y Juan Pérez, Á. A. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *RUSC. Universities and Knowledge Society Journal*, 12(3). 98-112. doi: <http://dx.doi.org/10.7238/rusc.v12i3.2515>

- Cassidy, D., Breakwell, N. y Bailey, J. (2014). Keeping them clicking: promoting student engagement in MOOC design. *The All Ireland Journal of Teaching and Learning in Higher Education*, 6(2), 1–15.
- Castrillo, M. D. (2014). Language Teaching in MOOCs: the Integral Role of the Instructor. En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOCs. Providing Learning, Transcending Boundaries* (pp. 67-90). Berlín: De Gruyter Open.
- Castrillo, M. D. y Mañana, J. (2017). Evolución del perfil sociodemográfico del participante en MOOC (2013-2016): estudio de un caso en lenguas extranjeras. En C. Delgado, C. Alario-Hoyos, y R. Hernández Rizzardini (Eds.). *Jornada de MOOCs en Español (EMOOCs-ES) EMOOCs 2017. Conference Book*. Springer
- Castrillo M. D., Martín-Monje, E., y Vázquez-Cano, E. (2018). *Guía práctica para el diseño y tutorización de MOOC*. MiríadaX. Recuperado el 7 de junio de 2022 de <https://www.uned.es/universidad/inicio/en/dam/jcr:ff99c67c-6f14-4baa-a84c-0cced629a214/Gu%C3%ADaDise%C3%B1oMOOC.pdf>
- Castrillo, M. D., y Sedano, B. (2021). Joining Forces Toward Social Inclusion: Language MOOC Design for Refugees and Migrants through the Lens of Maker Culture. *CALICO Journal*, 38(1), 79–102. <https://doi.org/10.1558/cj.40900>
- Chacón-Beltrán, R. (2018). Vocabulary learning strategies outside the classroom context what adults is learn in a technology-based learner-centred environment. *The Language Learning Journal*, 46(5), 583–593. doi:10.1080/09571736.2018.1503135
- Chang, S. C. (2011). A contrastive study of grammar translation method and communicative approach in teaching English grammar. *English language teaching*, 4(2), 13.
- Chatti, M. A., Dyckhoff, A.L., Schroeder, U. y Thüs, H. (2012). A Reference Model for Learning Analytics. *International Journal of Technology Enhanced Learning*, 4(5/6), 318–331.
- Chen, J. (2017). An e-Portfolio-based Model for the Application and Sharing of College English ESP MOOCs. *Higher Education Studies*, 7, 35-42.
- Chen, B., Haklev, S., Harrison, L., Najafi, H. y Rolheiser, C. (2015). How do MOOC learners' intentions relate to their behaviors and overall outcomes. *Proceedings of the AERA annual meeting*.
- Chen, L. y Wang, G. (2008). An efficient piecewise hashing method for computer forensics. *First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008)*, 635-638.
- Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D. y Emanuel, E. (6 de noviembre de 2013). The MOOC phenomenon: Who takes massive open online courses and why? Recuperado el 7 de junio de 2022 de http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2350964

- Cisel, M., Mano, M., Bachelet, R. y Silberzahn, P. (2015). A Tale of Two MOOCs: Analyzing Long-Term Course Dynamics. *European MOOCs Stakeholders Summit (eMOOCs), May 2015, Mons, Belgium.*
- Clark, D. (16 de abril de 2013). MOOCs: taxonomy of 8 types of MOOC [Entrada de blog]. *Donald Clark Plan B*. Recuperado el 7 de junio de 2022 de <http://donaldclarkplanb.blogspot.com/2013/04/moocs-taxonomy-of-8-types-of-mooc.html>
- Clements, K., West, R. E., y Hunsaker, E. (2020). Getting started with open badges and open microcredentials. *International Review of Research in Open and Distributed Learning*, 21(1) doi: <http://dx.doi.org/10.19173/irrodl.v21i1.4529>
- Coffrin, C., Corrin, L., de Barba, P. y Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. *Proceedings of the 4th international conference on learning analytics and knowledge (LAK '14)*, 83–92. Indianapolis, USA: ACM. doi: <https://doi.org/10.1145/2567574.2567586>
- Coleman, C. A., Seaton, D. T. y Chuang, I. (2015). Probabilistic use cases: Discovering behavioral patterns for predicting certification. *2nd ACM conference on learning@scale*. Vancouver, Canada: ACM.
- Colpaert, J. (2014). Conclusion. Reflections on Present and Future: towards an Ontological Approach to LMOOCs. En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOCs. Providing Learning, Transcending Boundaries* (pp. 161-172). Berlín: De Gruyter Open.
- Comisión Europea (2001). *The eLearning Action Plan. Designing tomorrow's education*. Communication from the Commission to the Council and the European Parliament, COM(2001) 172.
- Comisión Europea (2017). *Comunicación de la comisión al parlamento europeo, al consejo, al comité económico y social europeo y al comité de las regiones. Reforzar la identidad europea mediante la Educación y la Cultura. Contribución de la Comisión Europea a la reunión de dirigentes en Gotemburgo el 17 de noviembre de 2017*. Recuperado el 7 de junio de 2022 de [http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2017/0673/COM_COM\(2017\)0673_ES.pdf](http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2017/0673/COM_COM(2017)0673_ES.pdf)
- Concheiro, L. (2016). *Contra el tiempo*. Barcelona: Editorial Anagrama.
- Conde Gafaro, B. (2020). MOOCs in the Language Classroom: Using MOOCs as Complementary Materials to Support Self-Regulated Language Learning. En A. Andújar (Ed.), *Recent Tools for Computer- and Mobile-Assisted Foreign Language Learning* (pp. 194-211). Hershey: IGI Global.
- Conole, G. (2014). A new classification schema for MOOCs. *The International Journal for Innovation and Quality in Learning*, 2(3), 65-77.

- Conole, G. (2016). MOOCs as disruptive technologies: Strategies for enhancing the learner experience and quality of MOOCs. *Revista de Educación a Distancia (RED)*, 39, 1-17. doi: 10.6018/red/50/2.
- Conole, G. y Ehlers, U. D. (2010): Open Educational Practices: Unleashing the power of OER. *En UNESCO Workshop on OER in Namibia*. Windhoek. Recuperado el 7 de junio de 2022 de https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/OEP_Unleashing-the-power-of-OER.pdf
- Consejo de Europa (1982). *Recommendation No R (82)18 of the Committee of Ministers to member States concerning modern languages*. Recuperado el 7 de junio de 2022 de <https://rm.coe.int/16804fa45e>
- Consejo de Europa (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Secretaría General Técnica del MEC, Anaya e Instituto Cervantes.
- Cook, T. D., y Reichardt, C. S. (Eds.). (1979). *Qualitative and quantitative methods in evaluation research (Vol. 1)*. Beverly Hills, CA: Sage publications.
- Cotton, K. (2008). Computer-assisted instruction. *Encyclopedia of Special Education*, 514-520.
- Coursera (s.f.). About. Recuperado el 7 de junio de 2022, de <https://about.coursera.org/>
- Crystal, D. (2008). Two thousand million? *English Today*, 24(1), 3-6. doi:10.1017/S0266078408000023
- D'Antoni, S. (2009) Open Educational Resources: reviewing initiatives and issues. *Open Learning: The Journal of Open, Distance and e-Learning*, 24(1), 3-10, doi: 10.1080/02680510802625443
- Daniel, B.K. (Ed.) (2016). *Big Data and Learning Analytics in Higher Education: Current Theory and Practice*. Nueva York: Springer.
- Daniel, J. I. (1999). Computer-aided instruction on the World Wide Web: The third generation. *The Journal of Economic Education*, 30(2), 163-174.
- Daradoumis, T., Rodríguez-Ardura, I., Faulin, J., y Martínez-López, F. J. (2010). CRM Applied to Higher Education: Developing an e-Monitoring System to Improve Relationships in e-Learning Environments. *International Journal of Services Technology and Management*, 14(1), 103-125. doi: <http://dx.doi.org/10.1504/IJSTM.2010.032887>
- Davidian, M., y Louis, T. A. (2012). Why statistics? *Science*, 336(6077), 12.
- Davies, G., Otto, S. E. K., y Rüschoff, B. (2013). Historical perspectives on CALL. En M. Thomas, H. Reinders y M. Warschauer (Eds.), *Contemporary Computer-Assisted Language Learning*. Londres: Bloomsbury.

- Dawson, S. (2010). 'Seeing' the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, 41(5), 736-752.
- De Laat, M., Lally, V., Lipponen, L. y Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: a role for social network analysis. *International Journal of Computer Supported Collaborative Learning*, vol. 2, 87-103.
- DeLaRosa, Alexander (27 de agosto de 2021). Log Monitoring: not the ugly sister. [Entrada de blog]. *Pandora FMS*. Recuperado el 18 de febrero de 2022 de <https://pandorafms.com/blog/log-monitoring/>
- Demiray, U. e İşman, A. (2003). History of distance education. En A. İşman, M. Barkan y U. Demiray (Eds.), *Online education book*. Recuperado el 7 de junio de 2022 de <http://www.tojet.net/e-book/ebook.htm#2>
- Dhorne, L., Deflandre, J., Bernaert, O., Bianchi, S. y Thirouard, M. (2017). Mentoring Learners in MOOCs: A New Way to Improve Completion Rates? En *Digital Education: Out to the World and Back to the Campus* (Vol. 10254, Lecture Notes in Computer Science, pp. 29-37). Cham: Springer International Publishing.
- Díaz Alcaraz, F. (2002). *Didáctica y currículo: un enfoque constructivista*. Cuenca: Ediciones de la Universidad de Castilla La Mancha.
- Dichev, C., Dicheva, D., y Irwin, K. (2018). Gamification driven learning analytics. *Proceedings of the International Conference on E-Learning, ICEL, 2018*, 70-76.
- Díez-Arcón, P. (2021). Perfiles del participante en LMOOC: análisis bibliográfico y estudio de caso. *Bellaterra Journal of Teaching & Learning Language & Literature*, 14(3), e979. <https://doi.org/10.5565/rev/jtl3.979>
- Dillenbourg, P., Schneider, D.K. y Synteta, P. (2002). Virtual Learning Environments. In A. Dimitracopoulou (Ed). *Proceedings of the 3rd Hellenic Conference Information & Communication Technologies in Education*, 3-18. Rodas: Kastaniotis Editions.
- Dixon, E., y Thomas, M. (Eds.). (2015). *Researching language learner interactions online: From social media to MOOCs*. San Marcos, Texas: Computer Assisted Language Instruction Consortium.
- Dormezil, S., Khoshgoftaar, T. y Robinson-Bryant, F. (2019). *Differentiating between educational data mining and learning analytics: A bibliometric approach*. LABBEC Workshop (Learning analytics: Building bridges between the Education and the Computing communities; pp. 1-6), Montreal, Canada.
- Downes, S. (23 de abril de 2012). The Rise of MOOC. [Entrada de blog]. *Stephen Downes*. Recuperado el 7 de junio de 2022 de <http://www.downes.ca/post/57911>

- Downes, S. (17 de enero de 2013). What Makes a MOOC Massive? [Entrada de blog]. *Half an Hour*. Recuperado el 7 de junio de 2022 de <https://halfanhour.blogspot.com/2013/01/what-makes-mooc-massive.html>
- ds106. (s.f.). A History of ds106. Recuperado el 7 de junio de 2022 de <https://ds106.us/history/>
- Durall, E., Gros, B., Maina, M., Johnson, L. y Adams, S. (2012). *Perspectivas tecnológicas: educación superior en Iberoamérica 2012-2017*. Austin, Texas: The New Media Consortium.
- Durksen, T., Chu, M., Ahmad, Z., Radil, A. y Daniels, L. (2016). Motivation in a MOOC : A probabilistic analysis of online learners' basic psychological needs. *Social Psychology of Education*, 19(2), 241-260.
- Dyckhoff, A. L. (2011). Implications for Learning Analytics Tools: A Meta-Analysis of Applied Research Questions. *International Journal of Computer Information Systems and Industrial Management Applications*, 3, 594-601.
- ElAtia, S., Ipperciel, D. y Zaïane, O.R. (Eds.) (2016). *Data Mining and Learning Analytics: Applications in Educational Research*. John Wiley & Sons.
- Elias, T. (2011). Learning analytics: Definitions, processes and potential. Recuperado el 7 de junio de 2022 de <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
- Federal Writers' Project. (1939). *Radio in education*. Harrisburg: Department of Public Instruction Commonwealth of Pennsylvania.
- Fenstermacher, G. D. (1989): Tres aspectos de la filosofía de la investigación sobre la enseñanza. En M. C. Wittrock (Ed.). *La investigación de la enseñanza, I. Enfoques, teorías y métodos*. Madrid: Paidós /M.E.C.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304-317.
- Ferguson, R., y Clow, D. (2015). Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs). *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 51-58). ACM.
- Fernández, T. (10 de enero de 2019). Estos son los profesionales más buscados en 2019. *Expansión*. Recuperado el 7 de junio de 2022 de <https://www.expansion.com/>
- Fernández-Cruz, F. y Fernández-Díaz, M. (2016). Los docentes de la generación Z y sus competencias digitales. *Comunicar*, 24(46), 97-105. doi: <http://dx.doi.org.ezproxy.uned.es/10.3916/C46-2016-10>
- Feustle Jr, J. A. (1987). Software Reviews. *Hispania* 70(3), 708-710.

- Fiaidhi, J. (2014). The next step for learning analytics. *IT Professional*, 16(5), 4-8.
- Finnstats (20 de abril de 2021). Cluster Analysis in R [Entrada de blog]. *R-bloggers*. Recuperado el 1 de marzo de 2022 de <https://www.r-bloggers.com/2021/04/cluster-analysis-in-r/>
- Fischetti, E., y Gisolfi, A. (1990). From computer-aided instruction to intelligent tutoring systems. *Educational Technology*, 30(8), 7-17.
- Fleming, T. y Toutant, T. (2008). «A Modern Box of Magic»: School Radio in British Columbia, 1927-1984. *International Journal of E-Learning & Distance Education/Revue internationale du e-learning et la formation à distance*, 10(1), 53-73.
- Fonseca, L. (16 de agosto de 2019). Clustering Analysis in R using K-means [Entrada de blog]. *Towards Data Science*. Recuperado el 1 de marzo de 2022 de <https://towardsdatascience.com/clustering-analysis-in-r-using-k-means-73eca4fb7967>
- Fuchs, C. (2016). The structural and dialogic aspects of language massive open online courses (LMOOCs): A case study. En C. Wang y L. Winstead (Eds.), *Handbook of research on foreign language education in the digital age* (pp. 10-32). Pennsylvania: IGI Global.
- Fung, B., Wang, K., Chen, R. y Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4), article 14, 1-53. doi: 10.1145/1749603.1749605
- FunMOOC. (s.f.). About Fun. Recuperado el 31 de mayo de 2022 de <https://www.fun-mooc.fr/en/about-fun/>
- FutureLearn. (s.f.). Our Story. Recuperado el 7 de junio de 2022 de <https://www.futurelearn.com/about-futurelearn>
- García Aretio, L. (1999). Historia de la educación a distancia. *RIED. Revista Iberoamericana de Educación a Distancia*, 2(1). doi: <https://doi.org/10.5944/ried.2.1.2084>
- García Aretio, L. (2006). La Universidad Nacional de Educación a Distancia (UNED) de España. *Revista Iberoamericana De Educación a Distancia*, 9(1), 17-51.
- García Aretio, L. (2011). Perspectivas teóricas de la educación a distancia y virtual. *Revista española de pedagogía*, 249, 255-271.
- García Aretio, L. (2021). COVID-19 y educación a distancia digital: preconfi namiento, confi namiento y posconfi namiento. *RIED. Revista Iberoamericana de Educación a Distancia*, 24(1), pp. 09-32. doi: <http://dx.doi.org/10.5944/ried.24.1.28080>
- García, M. F. y Appel, C. (2016). *Synchronous tandem language learning in a MOOC context: A study on task design and learner performance*. EuroCALL 2016: CALL Communities and Culture, Limassol, Cyprus.

- García, M. S. y Cuevas, B. S. (2018). Nuevos roles en la docencia y tutorización de un MOOC integrado en la enseñanza formal. En M. d. C. Ortega Navas, M. Á. López González y P. A. Hernández (Coords.). *Innovación educativa en la era digital: libro de actas* (pp. 225-234). UNED, Universidad Nacional de Educación a Distancia.
- Garreta, M., Aguado, G., Mor, Y., Fernandez, C., & Riviou, K. (2015). A Peer-Mentoring Approach for the Continuous Professional Development of Educators in a MOOC Setting. Proceedings of the European Distance and E-Learning Network 2015 Annual Conference Barcelona, 9-12 June, 2015.
- Garrison, D. R. (1985). Three generations of technological innovation in distance education. *Distance Education* 6(2), 235-241, doi: 10.1080/0158791850060208
- Gasevic, D., Conole, G., Siemens, G., Long, P. (Eds). (2011) *LAK11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Banff, Canada, 27 February - 1 March 2011*. Nueva York: ACM. Recuperado el 7 de junio de 2022 de <https://www.solaresearch.org/publications/conference-proceedings/>
- GATE (Gabinete de Tele-Educación). (s.f.). MOOCs. *Guía para el docente*. Madrid: Universidad Politécnica de Madrid. Disponible en <http://serviciosgate.upm.es/docs/asesoramiento/Guia%20MOOC%20para%20eI%20profesor.pdf>
- Geser, G. (2007). *Open Educational Practices and Resources. OLCOS Roadmap, 2012*. Salzburg Research / EduMedia Group.
- Ghilay, Y. (2019). Effectiveness of Learning Management Systems in Higher Education. *Journal of Online Higher Education*, 3(2), 29-50.
- Gibbons, A. y O'Neal, A. (2014). TICCIT: Building Theory for Practical Purposes. *International Journal of Designs for Learning*, 5(2), 1-19.
- Gillani, N., Eynon, R., Osborne, M., Hjorth, I. y Roberts, S. (2014). Communication Communities in MOOCs. *Cornell University*. Recuperado el 7 de junio de 2022 de <https://arxiv.org/abs/1403.4640v2>
- Gilliland, B., Oyama, A. y Stacey, P. (2018). Second language writing in a MOOC: Affordances and missed opportunities. *The Electronic Journal for English as a Second Language*, 22(1), 1-25.
- Gilliot, J.-M., Garlatti, S., Rebai, I., y Belen-Sapia, M. (2013). Le concept de iMOOC pour une ouverture maîtrisée. *Séminaire MOOC et EIAH 2013*, 27-36. Toulouse, France. Recuperado el 7 de junio de 2022 de <https://ateliermooceiah2013.files.wordpress.com/2013/05/gilliot.pdf>
- Gimeno-Sanz, A. (2016). Moving a step further from «integrative CALL». What's to come? *Computer Assisted Language Learning* 29(6), 1102-1115, doi: <http://dx.doi.org/10.1080/09588221.2015.1103271>

- Jimeno-Sanz, A., Navarro-Laboulais, C., y Despujol-Zabala, I. (2017). Additional functionalities to convert an xMOOC into an xLMOOC. En C. Delgado Kloos, P. Jermann, M. Pérez-Sanagustín, D. Seaton y S. White (Eds.), *Digital education: Out to the world and back to the campus. Proceedings of the European MOOC Stakeholders Summit 2017* (pp. 48-57). Cham, Switzerland: Springer.
- Given, L. M. (2008). *The SAGE Encyclopedia of Qualitative Research Methods*. Los Angeles: SAGE Publications.
- Godwin-Jones, R. (2014). Global reach and local practice: The promise of MOOCs. *Language Learning & Technology*, 18(3), 5–15. Recuperado el 7 de junio de 2022 de <http://llt.msu.edu/issues/october2014/emerging.pdf>
- Goldstein, P. J. y Katz, R. N. (2005). *Academic Analytics: The Uses of Management Information and Technology in Higher Education*. EDUCAUSE Center for Applied Research.
- Gómez, I. (1996). Enseñanza y aprendizaje. *Cuadernos de pedagogía*, 250(1), 54-59.
- Gómez, J. (ed.) (2017). *UNIVERSITIC 2017. Análisis de las TIC en las Universidades Españolas*. Madrid: Crue Universidades Españolas.
- Gomez, M. (2014). Use and Mastery of Virtual Learning Environment in Brazilian Open University. *International Journal of Instruction*, 7(2), 119-134.
- González Ornelas, V. (2001). *Estrategias de enseñanza y aprendizaje*. México D.F.: PAX.
- Google (2022). Google Analytics Cookie Usage on Websites. Recuperado el 22 de febrero de 2022 de <https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage>.
- Grimes, S. (2013). An Introduction to Text Analytics: Social, Online and Enterprise [Presentación de PowerPoint]. Text and Social Analytics Summit, workshop presentation, 2013. Recuperado el 7 de junio de 2022 de www.slideshare.net/SethGrimes/an-introduction-to-text-analytics-2013-workshop-presentation.
- Grünewald, F., Mazandarani, E., Meinel, C., Teusner, R., Totschnig, M. y Willems, C. (2013). OpenHPI - a Case-Study on the emergence of two learning communities. *Proc. IEEE Global Eng. Edu. Conf.* (pp. 13-15).
- GSM Association (2019). *The Mobile Economy 2019*. Recuperado el 7 de junio de 2022 de <https://www.gsmaintelligence.com/research/?file=b9a6e6202ee1d5f787cfebb95d3639c5&download>
- Gu, J., Reeves, S., Stephen, A. y Wang, J. (2016). Introduction, Definition, and History of Distance Education. En A. D. Ritzhaupt (ed.). *Distance Education. A guide for theory and practice*.

- Gutiérrez Gutiérrez, D. P. (2017). La revolución tecnológica de la información y los retos para la sociedad actual. *Ar@cne: Revista electrónica de recursos en internet sobre geografía y ciencias sociales*. [En línea]. Barcelona: Universidad de Barcelona, nº 220, 1 de junio de 2017. Recuperado el 7 de junio de 2022 de <http://www.ub.edu/geocrit/aracne/aracne-220.pdf>
- Guzmán Parra, V. F. y Vila Oblitas, J. R. (2011). Recursos educativos abiertos y uso de internet en enseñanza superior: el proyecto OpenCourseWare. *EduTec. Revista Electrónica de Tecnología Educativa*, 38. Recuperado el 7 de junio de 2022 de <https://www.edutec.es/revista/index.php/edutec-e/article/view/383/120>
- Haggard, S. (2013). The Maturing of the MOOC. Literature review of Massive Open Online Courses and other forms of online distance learning [BIS Research Paper No. 130]. London: Department for Business Innovation & Skills - UK Government. Recuperado el 7 de junio de 2022 de https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/240193/13-1173-maturing-of-the-mooc.pdf
- Halawa, S., Greene, D. y Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. En U. Cress y C. Delgado Kloos (Eds.), *Proceedings of the European MOOC Stakeholder Summit* (pp. 58-65). P.A.U. Education.
- Hashim, H., Salam, S., Mohamad, S. N. M. y Sazali, N. S. S. (2018). The designing of adaptive self-assessment activities in second language learning using massive open online courses (MOOCs). *International Journal of Advanced Computer Science and Applications*, 9(9), 276-282.
- Hawkrige, D., & McCormick, B. (1983). China's television universities. *British Journal of Educational Technology*, 14(3), 160-173.
- Haythornthwaite, C. y de Laat, M. (2010). Social networks and learning networks: using social network perspectives to understand social learning. *Proceedings of the 7th International Conference on Networked Learning, 2010*, 183-190.
- He J. y Yuan, C. (2019). The 40-year History of the OUC: A University Open for All, Building a Defence Line for Educational Equality. Recuperado el 7 de junio de 2022 de <http://en.ouchn.edu.cn/index.php/news-v2/2578-the-40-year-history-of-the-ouc-a-university-open-for-all-building-a-defence-line-for-educational-equality>
- Henry, M., y Marrs, D. (2015). Cada Día Spanish: An analysis of confidence and motivation in a social learning language MOOC. International Conference on e-Learning, Las Palmas, Spain.
- Hill, P. (2013, 10 de marzo). Emerging student patterns in MOOC: A (Revised) graphical view [Entrada de blog]. Recuperado el 7 de junio de 2022 de <https://mfeldstein.com/emerging-student-patterns-in-moocs-a-revised-graphical-view/>
- Hinchey, P. (2008). *Action Research Primer*. New York: Peter Lang Publishing.

- Ho, A. D., Chuang, I., Reich, J., Coleman, C. A., Whitehill, J., Northcutt, C. G., et al. (2015). HarvardX and MITx: Two years of open online courses fall 2012-summer 2014. (HarvardX Working Paper No. 10). <https://doi.org/10.2139/ssrn.2586847>.
- Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., y Chuang, I. (2014). *HarvardX and MITx: The first year of open online courses* (HarvardX and MITx Working Paper No. 1).
- Hodgkinson-Williams, C., y Gray, E. (2009). Degrees of openness: The emergence of OER at the University of Cape Town. *International Journal of Education and Development Using Information and Communication Technology*, 5(5), 101–116. Recuperado el 7 de junio de 2022 de <http://ijedict.dec.uwi.edu/viewarticle.php?id=864>
- Holmes, W., Nguyen, Q., Zhang, J., Mavrikis M. y Rienties, B. (2019) Learning analytics for learning design in online distance learning. *Distance Education*, 40(3), 309-329. doi: [10.1080/01587919.2019.1637716](https://doi.org/10.1080/01587919.2019.1637716)
- Huber, P. J. (1981). *Robust statistics*. New York: John Wiley.
- Hylén, J. (2005). Open educational resources: Opportunities and challenges. OECD-CERI. Recuperado el 7 de junio de 2022 de <http://www.oecd.org/education/ceri/37351085.pdf>
- Ibañez Moreno, A. y Traxler, J. (2016). MALL-based MOOCs for language teachers: Challenges and opportunities. *Porta Linguarum*, 1, 73–85.
- Ifenthaler, D., Mah, D. y Yau, Y.J. (Eds.) (2019). *Utilizing Learning Analytics to Support Study Success*. Springer.
- Instituto Cervantes (2021). *El español en el mundo 2021*. Recuperado el 7 de junio de 2022 de https://cvc.cervantes.es/lengua/anuario/anuario_21/el_espanol_en_el_mundo_anuario_instituto_cervantes_2021.pdf
- Interactive Advertising Bureau (2013). IAB Audience Reach Measurement Guidelines. Recuperado el 22 de febrero de 2022 de <https://www.iab.com/wp-content/uploads/2015/06/AudienceReachMeasurementGuidelines.pdf>
- International Educational Data Mining Society. (s.f.). Home. Recuperado el 7 de junio de 2022 de www.educationaldatamining.org/
- ITU (3 de febrero de 2022). ICT Statistics. Recuperado el 29 de mayo de 2022, de <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- Jacoby, J. (2014). The disruptive potential of the massive open online course: A literature review. *Journal of Open, Flexible and Distance Learning*, 18(1), 73-85.
- Jansen, D. y Schuwer, R. (2015). Institutional MOOC strategies in Europe: Status report based on a mapping survey conducted in October - December 2014. EADTU report.

- Johns Hopkins University, World Bank y UNICEF (2021). COVID-19 Global Education Recovery Tracker. Last updated as of 2022-02-11. Recuperado el 29 de mayo de 2022 de <https://www.covideducationrecovery.global/maps/education-status/> Baltimore, Washington DC, New York: JHU, World Bank, UNICEF.
- Johnson, L., Adams, S. y Cummins, M. (2012) *The NMC Horizon Report: 2012 Higher Education Edition*. Austin, Texas: The New Media Consortium,.
- Johnson, L., Smith, R., Willis, H., Levine, A. y Haywood, K. (2011). *The 2011 Horizon Report*. Austin, Texas: The New Media Consortium.
- Johnstone, S. M. (2005). Open educational resources serve the world. *Educause Review*. Recuperado el 7 de junio de 2022 de <https://er.educause.edu/-/media/files/article-downloads/eqm0533.pdf>
- Jones, K., Thomson, J. y Arnold, K. (2014). Questions of Data Ownership on Campus. *EDUCASE Review*. Recuperado el 7 de junio de 2022 de <http://www.educause.edu/ero/article/questions-data-ownership-campus>
- Jordan, K. (2014). Initial Trends in Enrolment and Completion of Massive Open Online Courses. *International Review of Research in Open and Distance Learning*, 15(1), 133-160.
- Kabacoff, R. (2011). *R in action: Data analysis and graphics with R (1st ed.)*. Shelter Island: Manning Publications.
- Kahan, T., Soffer, T. y Nachmias, R. (2017). Types of Participant Behavior in a Massive Open Online Course. *The International Review of Research in Open and Distributed Learning*, 18(6). doi: <https://doi.org/10.19173/irrodl.v18i6.3087>
- Kandeel, R. H. (2019). Practices among French-as-a-foreign-language learners when writing a summary in the Paroles de FLE LMOOC. *Íkala*, 24(3), 487-502. <https://doi.org/10.17533/udea.ikala.v24n03a10>
- Kaplan, A. M. y Haenlein, M (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. *Business Horizons*. 59 (4), 441–450. doi:[10.1016/j.bushor.2016.03.008](https://doi.org/10.1016/j.bushor.2016.03.008).
- Kapler, T., y Wright, W. (2005). GeoTime information visualization. *Information Visualization*, 4(2), 136- 146.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. San Francisco: John Wiley & Sons.
- Karoly, L. y Panis, C. (2004). *The 21st Century at Work: Forces Shaping the Future Workforce and Workplace in the United States*. Santa Monica, CA; Arlington, VA; Pittsburgh, PA: RAND Corporation. Recuperado el 7 de junio de 2022 de <http://www.jstor.org/stable/10.7249/mg164dol.12>
- Keegan, D. (1996). *The Foundations of Distance Education (3° ed.)*. Routledge: London.

- Khalil, M. y Ebner, M. (2015). Learning Analytics: Principles and Constraints. En S. Carliner, C. Fulford y N. Ostashewski (Eds.), *Proceedings of EdMedia 2015--World Conference on Educational Media and Technology* (pp. 1789-1799). Montreal, Quebec, Canada: Association for the Advancement of Computing in Education (AACE). Recuperado el 7 de junio de 2022 de <https://www.learntechlib.org/primary/p/151455/>
- Khalil, M. y Ebner, M. (2016). What is learning analytics about? A survey of different methods used in 2013–2015. *Proceedings of Smart Learning Conference, Dubai, UAE, March 7–9, 2016* (pp. 294–304). Dubai: HBMSU Publishing House.
- Khalil, M. y Ebner, M. (2017). Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories. *Journal of Computing in Higher Education* 29, 114–132. doi: <http://dx.doi.org/10.1007/s12528-016-9126-9>
- Khan, S. (2012). *The one world schoolhouse: Education reimaged*. London: Hodder & Stoughton.
- Kizilcec, R. F., Piech, C. y Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *3rd International Conference on Learning Analytics and Knowledge* (pp. 170–179). New York: ACM. doi: 10.1145/2460296.2460330
- Klemke, R., Eradze, M. y Antonaci, A. (2018). The Flipped MOOC: Using Gamification and Learning Analytics in MOOC Design: A Conceptual Approach. *Education Sciences*, 8(1), 25.
- Kline, P. (1993). *An easy guide to factor analysis*. London: Routledge.
- Kohli, P. (26 de abril de 2020). Society has a Toxic Obsession with Productivity. *The UCSD Guardian*. Recuperado el 23 de mayo de 2022 de <https://ucsdguardian.org/2020/04/26/society-has-a-toxic-obsession-with-productivity/>
- Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *Educause Review*, 48(3), 62–63.
- Kong, N. (2011). Establishing a Comprehensive English Teaching Pattern Combining the Communicative Teaching Method and the Grammar-Translation Method. *English Language Teaching*, 4(1), 76-78.
- Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *International Review of Research in Open and Distance Learning* 12(3). Recuperado el 7 de junio de 2022 de <http://www.irrodl.org/index.php/irrodl/article/view/882>
- Kovanović, V., Joksimović, S., Gašević, D., Owers, J., Scott, A. M., y Woodgate, A. (2016). Profiling MOOC Course Returners: How Does Student Behavior Change Between Two Course Enrollments? *Proceedings of the Third (2016) ACM Conference on*

Learning@ Scale (pp. 269–272). ACM. doi:
<https://doi.org/10.1145/2876034.2893431>

- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Pergamon Press Inc.
- Krumm, A., Means, B. y Bienkowski, M. (2018). *Learning Analytics Goes to School. A Collaborative Approach to Improving Education*. Routledge.
- Lane, L. M. (15 de agosto de 2012). Three Kinds of MOOCs [Entrada de blog]. Lisa M Lane. Recuperado el 7 de junio de 2022 de <http://lisahistory.net/wordpress/musings/three-kinds-of-moocs/>
- Lane, J., y Kinser, K. (29 de septiembre de 2012). MOOC's and the McDonaldization of global higher education [Entrada de blog]. World.edu. Recuperado el 7 de junio de 2022 de <https://world.edu/moocs-and-the-mcdonaldization-of-global-higher-education/>
- Lang, C., Siemens, G., Wise, A. y Gašević, D. (Eds.) (2017). *Handbook of Learning Analytics*. SOLAR.
- Larsson, J.A. y White, B. (Eds.) (2014). *Learning Analytics: From research to practice*. Nueva York: Springer.
- Lee, S., Kadel, R. S., Madden, A., y Gazi, Y. (2018). How Can Learner Analytics Data Inform Language MOOC Design? *Boğaziçi Üniversitesi Eğitim Dergisi*, 35(1), 19-29.
- Leedham, M., Lillis, T. y Twiner, A. (2020). Exploring the core 'preoccupation' of social work writing: A corpus-assisted discourse study. *Journal of Corpora and Discourse Studies*, 3, 1-26.
- Leitner P., Ebner M. y Ebner M. (2019) Learning Analytics Challenges to Overcome in Higher Education Institutions. En: Ifenthaler, D., Mah, D. y Yau, Y.J. (Eds.) *Utilizing Learning Analytics to Support Study Success*. Springer.
- León Urrutia, M., Vázquez Cano, E., y López-Meneses, E. (2017). MOOC learning analytics using real-time dynamic metrics. *@tic. revista d'innovació educativa* (18), 38-47. doi:10.7203/attic.18.10022
- Lester, J., Klein, C., Johri, A. y Rangwala, H. (Eds.) (2018). *Learning Analytics in Higher Education: Current Innovations, Future Potential, and Practical Applications*. Routledge.
- Lewin, T. (19 de septiembre de 2012). Education Site Expands Slate of Universities and Courses. *New York Times*. Recuperado el 7 de junio de 2022 de <http://www.nytimes.com/2012/09/19/education/coursera-adds-more-ivy-league-partner-universities.html>
- Leys, C., Christophe Ley, C., Klein, O., Bernard, P. y Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around

the median. *Journal of Experimental Social Psychology*, 49(4), 764-766.
<http://dx.doi.org/10.1016/j.jesp.2013.03.013>

Li, Q y Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in Massive Open Online Courses. *Computers & Education*, vol. 127, 41-65, doi: <https://doi.org/10.1016/j.compedu.2018.08.005>.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *Internet and Higher Education*, 29, 40-48. doi: <https://doi.org/10.1016/j.iheduc.2015.12.003>.

Liu, B. (2006). *Web Data Mining*. Berlin Heidelberg: Springer.

Liyanagunawardena, T. R., Adams, A. A., y Williams, S.A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *International Review of Research in Open and Distance Learning*, 14 (3) 202-227.

Liyanagunawardena, T., Williams, S., y Adams, A. (2013). The Impact and Reach of MOOCs: A Developing Countries' Perspective. *eLearning Papers*, 33. Recuperado el 7 de junio de 2022 de <https://centaur.reading.ac.uk/32452/>

Lodge, J., Horvath J. y Corrin, L. (Eds.) (2018). *Learning Analytics in the Classroom. Translating Learning Analytics Research for Teachers*. Routledge.

Long, P. y Siemens, G. (2011). Penetrating the fog: analytics in learning and education. *Educause Review Online*, 6(5), 31-40.

López Meneses, E., Vázquez Cano, E., y Román, P. (2015). Análisis e implicaciones del impacto del movimiento MOOC en la comunidad científica: JCR y Scopus (2010-13). *Comunicar: Revista Científica de Comunicación y Educación*, 22(44), 73-80. doi: <http://dx.doi.org/10.3916/C44-2015-08>

Luik, P., Suviste, R., Lepp, M., Palts, T., Tõnisson, E., Säde, M., y Papli, K. (2019). What motivates enrolment in programming MOOCs? *British Journal of Educational Technology*, 50(1), 153-165.

MacKay, D. J. C. (2005). *Information Theory, Inference and Learning Algorithms* (7.2 ed.). Cambridge: Cambridge University Press.

Mallart i Navarra, J. (2001). Didáctica: concepto, objeto y finalidades. En F. Sepúlveda y N. Rajadell (coords.) *Didáctica general para psicopedagogos*. Madrid: UNED, 23-57.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. y Hung Byers, A. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute.

Markoff J. (15 de agosto de 2011). Virtual and Artificial, but 58,000 Want Course. *The New York Times*. Recuperado el 7 de junio de 2022 de http://www.nytimes.com/2011/08/16/science/16stanford.html?_r=1&

- Martí, J. (2012, 24 de agosto). Tipos de MOOC [Entrada de blog]. Xarxatic. Recuperado el 7 de junio de 2022 de <http://www.xarxatic.com/tipos-de-MOOC/>
- Martín Peris, E. (Ed.). (2008). *Diccionario de términos clave de ELE*. Madrid: SGEL. Recuperado el 7 de junio de 2022 de http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/indice.htm
- Martín-Monje, E., y Bárcena, E. (Eds.) (2014). *Language MOOC. Providing Learning, Transcending Boundaries*. Berlín: De Gruyter Open.
- Martín-Monje, E., Bárcena, E., y Read, T. (2014). La interacción entre compañeros y el feedback lingüístico en los COMA de lenguas extranjeras. *Profesorado, Revista de Currículum y Formación del Profesorado*, 18(1), 167-183.
- Martín-Monje, E., Castrillo, M. D. y Mañana-Rodríguez, J. (2018). Understanding online interaction in language MOOCs through learning analytics. *Computer Assisted Language Learning*, 31(3), 251-272. doi: 10.1080/09588221.2017.1378237
- Martín-Monje, E., Read, T., y Bárcena, E. (2017). The importance of forum interaction and success rates in language MOOCs. EMOOCS 2017: European MOOCs Stakeholders Summit.
- Maselena, A., Sabani, N., Huda, M., Ahmad, R., Jasmi, K. A. y Basiron, B. (2018). Demystifying learning analytics in personalised learning. *International Journal of Engineering & Technology*, 7(3), 1124-1129.
- Masters, K. (2011). A Brief Guide to Understanding MOOCs. *The Internet Journal of Medical Education*, 1(2). doi: 10.5580/1f21.
- Maya-Jariego, I., Holgado, D., González-Tinoco, E., Castaño-Muñoz, J. y Punie, Y. (2020). Typology of motivation and learning intentions of users in MOOCs: the MOOCKNOWLEDGE study. *Educational technology research and development*, 68(1), 203-224. doi: <https://doi.org/10.1007/s11423-019-09682-3>
- Mazza, R. (2009). *Introduction to Information Visualization*. London: Springer-Verlag.
- McAuley, A., Stewart, B., Siemens, G. y Cormier, D. (2010). *The MOOC model for digital practice*. University of Prince Edward Island. Recuperado el 7 de junio de 2022 de https://www.oerknowledgecloud.org/archive/MOOC_Final.pdf
- McCorkle, B., Halasek, K., Clinnin, K. y Selfe, C. L. (2016). Negotiating world Englishes in a writing-based MOOC. *Composition Studies*, 44(1), 53-71.
- McKay, T. (2015) An Introduction to Modern Learning Analytics. *Practical Learning Analytics (MOOC)*. Edx. Recuperado el 7 de junio de 2022 de <https://www.edx.org/course/practical-learning-analytics>
- McLoughlin, L. y Magnoni, F. (2017). The move-me project: Reflecting on xMOOC and cMOOC structure and pedagogical implementation. En K. Qian y B. Stephen (Eds.), *Beyond the language classroom: Researching MOOCs and other innovation* (pp. 59-69). Voillans, France: Research-publishing.net.

- McMinn, S. (2017). Participants' engagement in and perceptions of English language MOOCs. *Faces of English Education: Students, Teachers, and Pedagogy*, (pp. 176–196). Abingdon, UK: Taylor and Francis.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Nueva York: Cambridge University Press.
- Medina Salguero, R y Aguaded Gómez, J. I. (2014). Los MOOC en la plataforma educativa MiriadaX. *Profesorado: Revista De Curriculum Y Formación Del Profesorado*, 18(1), 137-153.
- Meinel, C. y Willems, C. (2013). *openHPI. The MOOC Offer at Hasso Plattner Institute*. Potsdam: Hasso Plattner Institute y Universidad de Potsdam.
- Merrill, M. D., Schneider, E. W. y Fletcher, K. A. (1979). *TICCIT*. Englewood Cliffs NJ: Educational Technology Publications.
- Methode Toussaint-Langenscheidt (3 de marzo de 2020). *Wikipedia, La enciclopedia libre*. Recuperado el 7 de junio de 2022 de https://de.wikipedia.org/w/index.php?title=Methode_Toussaint-Langenscheidt&oldid=197363310
- Metwally, A. H. S. y Yining, W. (2017). *Gamification in Massive Open Online Courses (MOOCs) to support Chinese language learning*. 2017 International Conference of Educational Innovation through Technology (EITT).
- Milligan, C., Littlejohn, A. y Margaryan, A. (2013). Patterns of engagement in connectivist MOOC. *Journal of Online Learning & Teaching* 9(2), 149-159.
- Mirth, D. (1996). Corbett and radio at the University of Alberta Extension. *Canadian Journal for the Study of Adult Education*, 10(1), 47-66.
- MOOC para la enseñanza de idiomas (16 de febrero de 2022). *Wikipedia, La enciclopedia libre*. Recuperado el 1 de junio de 2022 de https://es.wikipedia.org/w/index.php?title=MOOC_para_la_ense%C3%B1anza_de_idiomas&oldid=141710300
- Moreira Teixeira, A. y Mota, J. (2014). A Proposal for the Methodological Design of Collaborative Language MOOCs En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOC. Providing Learning, Transcending Boundaries* (pp. 33-47). Berlín: De Gruyter Open.
- Morgan, G. (2003). *Faculty Use of Course Management Systems (Vol. 2)*. Colorado: ECAR, EDUCAUSE Center for Applied Research.
- Mullaney, T. (2014). Making Sense of Moocs: A Reconceptualization of HarvardX Courses and Their Students. Recuperado el 7 de junio de 2022 de <https://ssrn.com/abstract=2463736> or <http://dx.doi.org/10.2139/ssrn.2463736>
- Narang, V. K. (2020). Design principles of MOOC: an effective approach from the learner's point of view. *Studies in Indian Place Names*, 40(33), 72-75.

- Nelson, R. J. y Wood, R. E. (1975). Radio in Foreign Language Education. *CAL-ERIC/CLL Series on Languages and Linguistics*, 11.
- Nic Giolla Mhichíl, M., Mac Lochlainn, C. y Beirne, E. (2018). Integrating #FLIrish101 in an LMOOC – learner engagement and pedagogical approach. En P. Taalas, J. Jalkanen, L. Bradley y S. Thouësny (Eds.), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 225-229). Research-publishing.net. <https://doi.org/10.14705/rpnet.2018.26.841>
- Niemeyer, P. y Leuck, D. (2013). *Learning Java (4th ed.)*. O'Reilly.
- Niemi, D., Pea, R. D., Saxberg, B. y Clark, R. E. (Eds.) (2018). *Learning Analytics in Education*. Information Age Publishing.
- Noffsinger, J. S. (1926). *Correspondence Scholls, Lyceums, hautauquas*. New York: Macmillan.
- Noffsinger, J. S. (1931). *Home study blue book*. Washington: The National Home Study Council.
- de Nooy, W., Mrvar, A. y Batagelj, V. (2018). *Exploratory Social Network Analysis with Pajek: Revised and Expanded Edition for Updated Software (3rd Edition)*. Nueva York: Cambridge University Press.
- Nunn, S., Avella, J. T., Kanai, T. y Kebritchi, M. (2016). Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learning (Newburyport, Mass.)*, 20(2), 13-29.
- Oblinger, D. G. (2012). Let's talk analytics. *EDUCAUSE Review*, 47(4), 10-13.
- OECD (2020). *PISA 2018 Results (Volume V): Effective Policies, Successful Schools* Paris: PISA, OECD Publishing. <https://doi.org/10.1787/ca768d40-en>.
- OpenEdOz (2016). *Students, universities and Open Education. Final Report 2016*. Recuperado el 7 de junio de 2022 de https://openedoz.org/wp-content/uploads/2017/02/ID14-3972_CSU_Wills_Final-Report_2016.pdf
- Orsini-Jones, M., Gafaro, B. C. y Altamimi, S. (2017). Integrating a MOOC into the postgraduate ELT curriculum: Reflecting on students' beliefs with a MOOC blend. En K. Qian & S. Bax (Eds.), *Beyond the language classroom: Researching MOOCs and other innovations* (pp. 71–83). Voillans, France: Research-publishing.net.
- Otto, S. E. K. (2017). From past to present: a hundred years of technology for L2 learning. En C. A. Chapelle y S. Sauro (eds.). *The Handbook of Technology and Second Language Teaching and Learning*. Hoboken: Wiley.
- Ozturk, H. T. (2015). Examining value change in MOOCs in the scope of connectivism and Open Educational Resources movement. *The International Review of Research in Open and Distributed Learning*, 16(5). Recuperado el 7 de junio de 2022 de <http://www.irrodl.org/index.php/irrodl/article/view/2027>

- Panagiotidis, P. (2019). MOOCs for language learning. Reality and prospects. En K. Graziano (Ed.) *Society for Information Technology & Teacher Education International Conference* (pp. 540-546). Association for the Advancement of Computing in Education (AACE).
- Pappano, L. (2 de noviembre de 2012). The year of the MOOC. *The New York Times*. Recuperado el 7 de junio de 2022 de <https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html>
- Parker, L. W. (1937). *School broadcasting in Great Britain*. Nueva York: National Advisory Council on Radio in Education.
- Pastor Cesteros, S. (2004). Aprendizaje de segundas lenguas. Lingüística aplicada a la enseñanza de idiomas. Alicante: Publicaciones de la Universidad de Alicante
- Patru, M. y Balaji, V. (Eds.). (2016). *Making Sense of MOOCs: A Guide for Policy-Makers in Developing Countries*. París/ Burnaby: UNESCO and Commonwealth of Learning. Recuperado el 7 de junio de 2022 de <http://unesdoc.unesco.org/images/0024/002451/245122E.pdf>
- Pedreño, A., Moreno, L., Ramón, A. y Pernías, P. (2013). UniMOOC: Trabajo colaborativo e innovación educativa. *Campus Virtuales*, 2(1), 10-18.
- del Peral Pérez, J. J. (2019). Patrones temporales de participación en MOOC. Estudio de un MOOC de lenguas. *RIED. Revista Iberoamericana de Educación a Distancia*, 22(2). doi: <http://dx.doi.org/10.5944/ried.22.2.23109>
- Petersen, R. J. (2012). Policy dimensions of analytics in higher education. *EDUCAUSE Review*, 47(4), 44-49.
- Pilli, O. y Admiraal, W. (2016). A Taxonomy of Massive Open Online Courses. *Contemporary Educational Technology*, 7(3), 223-240.
- Poellhuber, B., Roy, N., y Bouchoucha, I. (2019). Understanding Participant's Behaviour in Massively Open Online Courses. *International Review of Research in Open and Distributed Learning*, 20(1).
- Polo, J.D. (24 de enero de 2013). UNED Coma – Cursos Online Masivos Abiertos de la UNED [Entrada de blog]. *wwwwhatsnew*. Recuperado el 7 de junio de 2022 de <https://wwwwhatsnew.com/2013/01/24/uned-coma-cursos-online-masivos-abiertos-de-la-uned/>
- Poy, Y., y Gonzales-Aguilar, A. (2014). Factores de éxito de los MOOC: algunas consideraciones críticas. *Revista Ibérica de Sistemas y Tecnologías de Información*, SPE1 (2014), 105-118.
- Qayyum, A. y Zawacki-Richter, O. (2019) The State of Open and Distance Education. En O. Zawacki-Richter y A. Qayyum (eds) *Open and Distance Education in Asia, Africa and the Middle East*. Singapore: Springer Open.

- R (s.f.). What is R? Recuperado el 7 de junio de 2022 de <https://www.r-project.org/about.html>
- Radio ECCA (s.f.). Conozca Radio ECCA. Recuperado el 7 de junio de 2022 de <http://www2.radioecca.org/conozcanos#60>
- Ramanarayanan, V., Pautler, D., Lange, P., Tsuprun, E., Ubale, R., Evanini, K. y Suendermann-Oeft, D. (2018). Toward scalable dialog technology for conversational language learning: A case study of the ToEFL® MOOC. INTERSPEECH. Recuperado el 7 de junio de 2022 de <https://rutujaubale.github.io/papers/interspeech2018-dialogdemo.pdf>
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519-549.
- Read, T. (2014). The Architectonics of Language MOOCs. En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOCs. Providing Learning, Transcending Boundaries* (pp. 91-105). Berlín: De Gruyter Open.
- Read, T. y Bárcena, E. (2020). Toward a Framework for Language MOOCs and Mobile Assisted Language Learning. *Propósitos y Representaciones*, 8(1), e470. doi: <http://dx.doi.org/10.20511/pyr2020.v8n1.470>
- Reich, J. (8 de diciembre de 2014). MOOC Completion and Retention in the Context of Student Intent. *EDUCAUSE Review*. Recuperado el 7 de junio de 2022 de <http://www.educause.edu/ero/article/mooc-completion-and-retention-context-student-intent>
- Research and Markets (2019). *Online Education Market & Global Forecast, by End User, Learning Mode (Self-Paced, Instructor Led), Technology, Country, Company*.
- Richards, J. C. y Rodgers, T. S. (1986). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.
- Roberts, E. M., Stewart, B. M. y Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software, Articles*, 91(2), 1-40, doi: 10.18637/jss.v091.i02. Recuperado el 7 de junio de 2022 de <https://www.jstatsoft.org/article/view/v091i02>
- Rodrigo, C. (2014). Accessibility in language MOOCs. En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOCs: Providing learning, transcending boundaries* (pp. 106–126). Berlín: De Gruyter Open.
- Rodrigues, R. L., Ramos, J. L., Silva, J. C. S., Gomes, A. S., de Souza, F. D. F. y Maciel, A. M. A. (2016). Discovering level of participation in MOOCs through clusters analysis. *16th international conference on Advanced Learning Technologies* (pp. 232–233)
- Romero, C. y Ventura, S. (2006). *Data mining in E-learning*. Southampton, England: Wit-Press.

- Romero, C. y Ventura, S. (2007) Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.
- Romero, C. y Ventura, S. (2020) Educational data mining and learning analytics: An updated survey. *WIREs Data Mining Knowl Discov*, 10:e1355. Recuperado el 7 de junio de 2022 de <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1355>
- Romero, C., Cerezo, R., Bogarín, A. y Sánchez-Santillán, M. (2016). Educational Process Mining: a tutorial and case study using Moodle data sets. En S. ElAtia, D. Ipperciel y O. R. Zaiane (Eds.), *Data Mining and Learning Analytics* (pp. 1-28). Wiley.
- Rosenberg, M. J. (2001). *e-Learning. Strategies for delivering knowledge in the Digital Age*. New Cork: McGraw-Hill.
- Rubio, F. (2014). Teaching pronunciation and comprehensibility in a Language MOOC. En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOCs. Providing Learning, Transcending Boundaries* (pp. 143-160). Berlín: De Gruyter Open.
- Ruiz Bolívar, C. (2015). El MOOC: ¿un modelo alternativo para la educación universitaria?. *Apertura (Guadalajara, Jal.)*, 7(2), 86-100. Recuperado el 7 de junio de 2022 de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-61802015000200086&lng=es&tlng=es.
- Salinas, J. y Sureda, J. (1992): Aprendizaje abierto y educación a distancia. En J. M. Sancho, (Coord.). *European Conference about Information Technology in Education: A Critical Insight. Proceedings* (pp. 677-686). Congr s Europeu T.I.E., Barcelona.
- Sallam, M. H., Mart n-Monje, E. y Li, Y. (2020). Research trends in language MOOC studies: a systematic review of the published literature (2012-2018). *Computer Assisted Language Learning*. doi: 10.1080/09588221.2020.1744668
- Sanders, R. H. (1984). Pilot-Spion: A computer game for German students. *Die Unterrichtspraxis/Teaching German*, 17(1), 123-129.
- Santos, J. L., Klerkx, J., Duval, E., Gago, D. y Rodr guez, L. (2014). *Success, activity and drop-outs in MOOCs an exploratory study on the UNED COMA courses*. 4th International Conference on Learning Analytics and Knowledge, LAK 2014, Indianapolis, IN; United States.
- Sanz Gil, M. (2021). LMOOC para la integraci n de personas desplazadas. *Anales de Filolog a Francesa*, (29), 463–477. <https://doi.org/10.6018/analesff.483181>
- Saputro, R., Salam, S., Zakaria, M. y Anwar, T. (2019). A gamification framework to enhance students' intrinsic motivation on MOOC. *Telkomnika*, 17(1), 170-178.
- Schildt, H. (2021). *Java: the complete reference*. (12th ed.). McGraw-Hill
- Sclater, N. (2017). *Learning Analytics Explained*. New York: Routledge.

- Seaman, J. E., Allen, I. E., y Seaman, J. (2018). *Grade increase: Tracking distance education in the United States*. Oakland, CA: Babson Survey Research Group.
- Seidametova, Z. (2018). Design and Development of MOOCs. En V. Ermolayev, M. C. Suárez-Figueroa, V. Yakovyna, V. Kharchenko, V. Kobets, H. Kravtsov, V. Peschanenko, Y. Prytula, M. Nikitchenko y A. Spivakovsky (Eds.), *Proceedings of the 14th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops* (Vol. 2104, 462-471). CEUR Workshop Proceedings.
- Şeker, S. (2013). Computer - Aided Learning in Engineering Education. *Procedia - Social and Behavioral Sciences*, 83(C), 739-742.
- Semenova, T. (2020). The role of learners' motivation in MOOC completion. *Open Learning: The Journal of Open, Distance and e-Learning*. doi: 10.1080/02680513.2020.1766434
- Shafranovich, Y. (2005). Common Format and MIME Type for CSV Files. RFC 4180. IETF. doi:10.17487/RFC4180.
- Shah, D. (18 de enero de 2018). By the Numbers: MOOCs in 2017 [Entrada de blog]. *Class Central*. Recuperado el 29 de mayo de 2022 de <https://www.classcentral.com/report/mooc-stats-2017/>
- Shah, D. (11 de diciembre de 2018). By the Numbers: MOOCs in 2018 [Entrada de blog]. *Class Central*. Recuperado el 29 de mayo de 2022 de <https://www.classcentral.com/report/mooc-stats-2018/>
- Shah, D. (2 de diciembre de 2019). By the Numbers: MOOCs in 2019 [Entrada de blog]. *Class Central*. Recuperado el 29 de mayo de 2022 de <https://www.classcentral.com/report/mooc-stats-2019/>
- Shah, D. (30 de noviembre de 2020). By the Numbers: MOOCs in 2020 [Entrada de blog]. *Class Central*. Recuperado el 29 de mayo de 2022 de <https://www.classcentral.com/report/mooc-stats-2020/>
- Shah, D. (1 de diciembre de 2021). By the Numbers: MOOCs in 2021 [Entrada de blog]. *Class Central*. Recuperado el 29 de mayo de 2022 de <https://www.classcentral.com/report/mooc-stats-2021/>
- Shah, D., Pickard, L y Ma, R. (11 de enero de 2022). Massive List of MOOC Platforms Around The World in 2022 [Entrada de blog]. *Class Central*. Recuperado el 31 de mayo de 2022 de <https://www.classcentral.com/report/mooc-platforms/>
- Shalatska, H. M. (2018). The efficiency of MOOCs implementation in teaching English for professional purposes. *Information Technologies and Learning Tools*, 66(4), 186–196.
- Sharma, K., Jermann, P. y Dillenbourg, P. (2015). Identifying styles and paths toward success in MOOCs. *International Educational Data Mining Society*.

- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist* 57(10), 1380–1400.
- Siemens, G., y Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. En S. B. Shum, D. Gasevic, y R. Ferguson (Eds.). *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252-254). New York, NY: ACM.
- Siemens, G. y Gasevic, D. (2012). Guest Editorial - Learning and Knowledge Analytics. *Educational Technology & Society* 15(3), 1-2.
- Singh, X. P. (2015). The History and the Current Status of Computer Assisted Language Learning. *The Journal of English Language Teaching (India)* LVII(5), 25-35.
- Singh, D., y Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524.
- Skliar, C. (2021). Postludio Salirse del tiempo: la educación como instante en comunidad. *RIESED - Revista Internacional De Estudios Sobre Sistemas Educativos*, 3(12), 242-251. Recuperado el 7 de junio de 2022 de <http://www.riesed.org/index.php/RIESED/article/view/151>
- Slade, S. y Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1509–1528.
- Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., y Gasevic, D. (2017). Tools for Educational Data Mining. *Journal of Educational and Behavioral Statistics*, 42(1), 85-106.
- Smith, B. (2015). *Beginning JSON*. Apress.
- Smith, M. y Casserly, C. (2006). *The Promise of Open Educational Resources*. *Change: The Magazine of Higher Learning* 38, 8-17. doi: 10.3200/CHNG.38.5.8-17.
- Sokolik, M. (2014). What Constitutes an Effective Language MOOC? En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOCs. Providing Learning, Transcending Boundaries* (pp. 16-32). Berlín: De Gruyter Open.
- Solans García, M. y Estebas Vilaplana, E. (2020). Los MOOC en el proceso de enseñanza y aprendizaje universitaria: El caso de un LMOOC de pronunciación inglesa a través de textos literarios. En M. Solans García (Coord.), *Estrategias de aprendizaje en el entorno digital: propuestas desde el ámbito universitario* (pp. 13-28). Ediciones Egregius.
- Spirtes, P., Glymour, C. y Scheines, R. (2000). *Causation, prediction, and search*. New York: MIT Press.
- Stacey, P. (2014). Pedagogy of MOOCs. *The International Journal for Innovation and Quality in Learning*, 2(3), 111-115.

- Stephens, M. (1972). The BBC's Contribution to Education in the Commonwealth. *Journal of the Royal Society of Arts*, 120(5188), 240-249.
- Stevens, V. (2013). What's with the MOOCs? *TESL-EJ: Teaching English as a Second or Foreign Language*, 16(4).
- Stracke, C. M., Downes, S., Conole, G., Burgos, D y Nascimbeni, F. (2019). Are MOOCs Open Educational Resources? A literature review on history, definitions and typologies of OER and MOOCs. *Open Praxis*, 11(4), 331-341.
- Sunar, A., Abbasi, R., Davis, H., White, S. y Aljohani, N. (2020). Modelling MOOC learners' social behaviours. *Computers in Human Behavior*, 107. doi: 10.1016/j.chb.2018.12.013.
- Sursock, A. (2015). *Trends 2015: Learning and Teaching in European Universities*. European University Association. Recuperado el 7 de junio de 2022 de <https://eua.eu/downloads/publications/trends%202015%20learning%20and%20teaching%20in%20european%20universities.pdf>
- Suthers, D. y Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. (pp. 64-74). New York: ACM.
- Tafazoli, D. y Golshan, N. (2014). Review of Computer-Assisted Language Learning: History, Merits & Barriers. *International Journal of Language and Linguistics. Special Issue: Teaching English as a Foreign/Second Language*. Vol. 2, No. 5-1, 32-38, doi: 10.11648/j.ijll.s.2014020501.15
- Tatman, R. (2018). Data Cleaning Challenge: Scale and Normalize Data [Entrada de blog]. *Kaggle*. Recuperado el 7 de junio de 2022 de <https://www.kaggle.com/rtatman/data-cleaning-challenge-scale-and-normalize-data>
- Taylor, C., Veeramachaneni, K. y O'Reilly, U-M. (2014). Likely to stop? Predicting Stopout in Massive Open Online Courses. *Cornell University*, arXiv:1408.3382v1. doi: <https://arxiv.org/abs/1408.3382v1>
- Teixeira, A. M. y Mota, J. (2014). A Proposal for the methodological design of collaborative language MOOCs. En E. Martín-Monje y E. Bárcena (Eds.), *Language MOOC. Providing Learning, Transcending Boundaries* (pp. 33-47). Berlín: De Gruyter Open.
- Tejada Fernández, J. (2011). ¿Qué son la didáctica y la teoría del currículum? En I. Cantón Mayo y M. Pino-Juste (coords). *Diseño y desarrollo del currículum* (19-40). Madrid: Alianza Editorial
- The Open University (s.f.). Exhibition: The OU Story. Recuperado el 7 de junio de 2022 de <https://www.open.ac.uk/library/digital-archive/exhibition/53>

- Tibshirani, R., Walther, G. y Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.
- Triguero, S., González, Moyano, J. M., García, S., Alcalá-Fdez, J., Luengo, J., Fernández, A., del Jesus, M. J., Sánchez, L. y Herrera, F. (2017). KEEL 3.0: An Open Source Software for Multi-Stage Analysis in Data Mining. *International Journal of Computational Intelligence Systems* 10, 1238-1249.
- Troncarelli, D., y Villarini, A. (2017). Internationalization of higher education and the use of MOOCs to improve second language proficiency: The MOVE-ME project. En K. Qian & S. Bax (Eds.), *Beyond the language classroom: Researching MOOCs and other innovations* (pp. 5–14). Voillans, France: Research-publishing.net.
- Tseng, S. F., Tsao, Y. W., Yu, L. C., Chan, C. L. y Lai, K. R. (2016). Who will pass? Analyzing learner behaviors in MOOCs. *Research and Practice in Technology Enhanced Learning*, 11(8). <https://doi.org/10.1186/s41039-016-0033-5>
- Tuomi, I. (2013). Open Educational Resources and the Transformation of Education, *European Journal of Education*, 48(1), 58–78. doi: <https://doi.org/10.1111/ejed.12019>
- Tyler, T. F. (1933). *An appraisal of radio broadcasting in the land-grant colleges and state universities*. Washington D.C.: National Committee on Education by Radio.
- Udacity (s.f.). About Us. Recuperado el 7 de junio de 2022 de <https://www.udacity.com/us#>
- UN (2020). Policy Brief: Education during COVID-19 and beyond (August 2020). United Nations. Recuperado el 7 de junio de 2022 de https://www.un.org/sites/un2.un.org/files/sg_policy_brief_covid-19_and_education_august_2020.pdf
- UNESCO (1949). *Broadcasting to schools: Report on the organization of school broadcasting services in various countries*. Paris: UNESCO.
- UNESCO (2016). *Statistical yearbook (2016th ed.)*. New York: United Nations.
- UNESCO (16 de diciembre de 2020). From COVID-19 learning disruption to recovery: A snapshot of UNESCO's work in education in 2020. Recuperado el 7 de junio de 2022 de <https://en.unesco.org/news/covid-19-learning-disruption-recovery-snapshot-unescos-work-education-2020>
- UNESCO (2020). *El enfoque de Aprendizaje a lo Largo de Toda la Vida: Implicaciones para la política educativa en América Latina y el Caribe*. París: UNESCO.
- UNESCO (s.f.). Open Educational Resources (OER). Recuperado el 7 de junio de 2022 de <https://en.unesco.org/themes/building-knowledge-societies/oer>
- UNESCO-UIL (2010). *Marco de Acción de Belém*. CONFINTEA VI. Hamburgo, UNESCO-UIL

- Universia España (16 de mayo de 2013). La UNED se alza ganadora de la I Edición de los premios Miríada X. *Universia*. Recuperado el 7 de junio de 2022 de <https://www.universia.net/es/actualidad/orientacion-academica/uned-se-alza-ganadora-i-edicion-premios-miriada-x-1023529.html>
- UTEID (Unidad de Tecnología Educativa e Innovación Docente) (2014). *Guía metodológica para la planificación, diseño e impartición de MOOCs y SPOCs*. Madrid: Universidad Carlos III de Madrid.
- Vázquez Cano, E. y López Meneses, E. (2014). Los MOOC y la Educación Superior: La expansión del conocimiento. *Profesorado: Revista De Curriculum Y Formación Del Profesorado*, 18(1), 3-12.
- Vázquez Cano, E., López Meneses, E. y Sarasola Sánchez-Serrano, J. (2013). *La expansión del conocimiento en abierto: Los MOOC*. Octaedro - Universitat de Barcelona. Institut de Ciències de l'Educació.
- Ventura, P., Bárcena, E. y Martín-Monje, E. (2014). Analysis of the impact of social feedback on written production and student engagement in Language MOOCs. *Procedia-Social and Behavioral Sciences*, 141, 512-517.
- Ventura, P. y Martín-Monje, E. (2016). Learning specialised vocabulary through Facebook in a massive open online course. En A. Pareja-Lora, C. Calle-Martínez & P. Rodríguez-Arancón (Eds.), *New perspectives on teaching and working with languages in the digital era* (pp. 117-128). Dublin, Ireland: Research-publishing.net.
- Virvou, M, Alepis, E., Tsihrintzis, G.A. y Jain, L.C. (Eds.) (2019). *Machine Learning Paradigms: Advances in Learning Analytics*. Springer.
- Voice of America (s.f.). VOA History. Recuperado el 7 de junio de 2022 de <https://www.insidevoa.com/p/5829.html>
- Voice of America Learning English (s.f.). About us. Recuperado el 7 de junio de 2022 de <https://learningenglish.voanews.com/p/5373.html>
- Vorobyeva, A. A. (2018). Language acquisition through massive open online courses (MOOCs): Opportunities and restrictions in educational university environment. *XLinguae*, 11(2), 136-146. doi:10.18355/XL.2018.11.02.11
- Walton, A., Robles, J., Cataylo, S., Horn, L., Thornton, J., y Whitfield, K. (2014). The Employer Potential of MOOCs: A Mixed-Methods Study of Human Resource Professionals' Thinking on MOOCs. *The International Review of research in open and distance learning*, 15 (5), 1-25.
- Wang, J., An, N. y Wright, C. (2018). Enhancing beginner learners' oral proficiency in a flipped Chinese foreign language classroom. *Computer Assisted Language Learning*, 31 (5-6), 490-521. doi:10.1080/09588221.2017.1417872

- Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. *6th international conference on learning analytics & knowledge, Edinburgh, UK.*
- Warschauer, M. (1996). Computer Assisted Language Learning: an Introduction. En S. Fotos (Ed.) *Multimedia language teaching* (pp. 3-20). Tokyo: Logos International.
- Warschauer, M. (2004). Technological Change and the Future of CALL. En S. Fotos & C. Brown (Eds.) *New Perspectives on CALL for Second and Foreign Language Classrooms* (5-25). Mahwah: Lawrence Erlbaum Associates.
- Wasserman, S. y Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Waterman, K. K. y Bruening, P. J. (2014). Big Data analytics: Risks and responsibilities. *International Data Privacy Law, 4*(2), 89-95.
- Waters, B. R., Balfanz, D., Durfee, G., y Smetters, D. K. (2004). Building an encrypted and searchable audit log. *Proceedings of the Network and Distributed System Security Symposium, NDSS 2004*. The Internet Society.
- Watkinson, J. D. (1996). «Education for Success»: The International Correspondence Schools of Scranton, Pennsylvania. *The Pennsylvania Magazine of History and Biography*. Vol. CXX (4).
- Weller, M. (2007). *Virtual learning environments: Using, choosing and developing your VLE*. Abingdon: Routledge.
- White, C. (2003). *Language Learning in Distance Education*. Cambridge: Cambridge University Press.
- Whitmer, J., Schiorring, E. y James, P. (2014). *Patterns of persistence: What engages students in a remedial English writing MOOC?* 4th International Conference on Learning Analytics and Knowledge, LAK 2014, Indianapolis, IN; United States.
- Wiley, D. (5 de marzo de 2014). The Access Compromise and the 5th R. [Mensaje en blog]. *Iterating toward openness*. Recuperado el 7 de junio de 2022 de <https://opencontent.org/blog/archives/3221>
- Wiley, D., Reeves, T. y Reynolds, T. (2015). The MOOC misstep and the Open Education infrastructure. In C. Bonk y M. Lee (Eds.), *MOOCs and Open Education around the world* (pp. 3-11). London: Routledge.
- Wilkins, D. A. (1973). *The linguistic and situational content of the common core in a unit/credit system*. Estrasburgo: Council of Europe.
- Worldometer (2022). Población mundial. Recuperado el 7 de junio de 2022 de <https://www.worldometers.info>
- Xiao, J. (2008). Teaching English at a distance in China's Radio and Television Universities. *Open Learning: The Journal of Open, Distance and E-Learning, 23*(3), 171-183.

- Xinying, Z. (2017). Researching into a MOOC embedded flipped classroom model for college English Reading and Writing course. En Q. Kan y S. Bax (Eds.), *Beyond the language classroom: researching MOOCs and other innovations* (pp. 15-27). Research-publishing.net. <https://doi.org/10.14705/rpnet.2017.mooc2016.668>
- Xu, B. y Yang, D. (2016). Motivation classification and grade prediction for MOOCs learners. *Computational Intelligence and Neuroscience*, 2016(4), 1–7. doi: <https://doi.org/10.1155/2016/2174613>.
- Yousef, A., Chatti, M., Schroeder, U., y Wosnitza, M. (2014). What Drives a Successful MOOC? An Empirical Examination of Criteria to Assure Design Quality of MOOCs. *2014 IEEE 14th International Conference on Advanced Learning Technologies*, 44-48.
- Zabalza Beraza, M. (1984). El análisis de la enseñanza desde el modelo comunicacional. *Enseñanza & Teaching* vol. 2. Recuperado el 7 de junio de 2022 de <http://revistas.usal.es/index.php/0212-5374/article/view/3090/3118>

ANEXO I. CÓDIGO FUENTE DEL PROGRAMA JAVA DE PRE-PROCESADO

```
package com.lmooc;

import java.io.BufferedReader;
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.IOException;
import java.io.PrintStream;
import java.text.SimpleDateFormat;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Calendar;
import java.util.Date;
import java.util.GregorianCalendar;
import java.util.StringTokenizer;

import org.json.simple.JSONObject;
import org.json.simple.parser.JSONParser;
import org.json.simple.parser.ParseException;

public class ReadJSON {
    static ArrayList<User> users;

    public static void main(String[] args) {
        String csvFile = args[0];
        String logsDir = args[1];
        String excluded = args[2];
        PrintStream outputStream = null;
        PrintStream console = System.out;

        users = new ArrayList<User>();
        try{
            parseLogs(logsDir);
            processEvents();
            parsecsv(csvFile);
        } catch (IOException | ParseException e) {
            e.printStackTrace();
        }
        calculateDuration();
        Utils.generateStatistics(users, excluded);
    }
}
```

```

    private static void parsecsv(String csvFile) throws
FileNotFoundException, IOException {
        String line;
        FileReader csvReader = new FileReader(csvFile);
        BufferedReader bufferedReadercsv = new
BufferedReader(csvReader);

        while ((line = bufferedReadercsv.readLine()) !=
null) {

            StringTokenizer csv = new
StringTokenizer(line, ";");
            String userId = csv.nextToken();

            for (int i = 0; i < users.size(); i++) {

                if
(users.get(i).getId().equals(userId)) {
                    users.get(i).grade =
Float.parseFloat(csv.nextToken());
                    users.get(i).actAver =
Float.parseFloat(csv.nextToken());
                    users.get(i).actFinalAver =
Float.parseFloat(csv.nextToken());
                    if (csv.nextToken().equals("Y"))
                        users.get(i).certEligible =
true;

                    if (csv.nextToken().equals("Y"))
                        users.get(i).certSent = true;

                }

            }
            bufferedReadercsv.close();
        }

        private static void parseLogs(String logsDir) throws
FileNotFoundException, IOException, ParseException {
            File dir = new File(logsDir);
            File[] directoryListing = dir.listFiles();
            Arrays.sort(directoryListing);
            if (directoryListing != null) {
                int numLines=0;
                for (File inputFile : directoryListing)
                    String lineLog = null;
                    FileReader fileReader = new
FileReader(inputFileName);
                    BufferedReader bufferedReader = new
BufferedReader(fileReader);

```

```

        while ((lineLog =
bufferedReader.readLine()) != null) {
            if (lineLog.endsWith("}")) {
                numLines++;
                JSONObject obj = (JSONObject)
new JSONParser().parse(lineLog);

                String userName =
obj.get("username").toString();
                String event_type =
null;
                if
(obj.get("event_type") != null)
                    event_type =
obj.get("event_type").toString();
                String referer = null;
                if (obj.get("referer")
!= null)
                    referer =
obj.get("referer").toString();
                String time = null;
                if (obj.get("time") !=
null)
                    time =
obj.get("time").toString();
                time = time.substring(0,
19).replace('T', ' ');
                if (obj.get("session")
!= null) {
                    }
                    Date date = null;
                    Date begin = null;
                    Date end = null;
                    try {
                        date = new
SimpleDateFormat("yyyy-MM-dd HH:mm:ss").parse(time);
                        begin = new
SimpleDateFormat("yyyy-MM-dd HH:mm:ss").parse("2019-01-15
00:00:00");
                        end = new
SimpleDateFormat("yyyy-MM-dd HH:mm:ss").parse("2019-03-10
23:59:59");
                    } catch (Exception e) {

                        System.out.println("Error formatting Date");
                    }
                    User user =
containsUser(userName);
                    if (date.after(begin) &&
date.before(end)) {

```

```

        if (user != null) {
            if (event_type
!= null && event_type.startsWith("/courses/course-
v1:UNED+MOONLITE_I_001+2018") ) {

                user.addEvent(date, event_type);

            } else
if(referer.contains("MOONLITE_I_001+2018")){

                user.addEvent(date, referer);

            }
            } else {
                user = new
User(userName);
                if (event_type
!= null && event_type.startsWith("/courses/course-
v1:UNED+MOONLITE_I_001+2018")) {

                    user.addEvent(date, event_type);

                    users.add(user);

                } else
if(referer.contains("MOONLITE_I_001+2018")){

                    user.addEvent(date, referer);

                    users.add(user);

                }

            }

        }
        bufferedReader.close();
    }
    System.out.println("LINES: " + numLines);
}
}

public static User containsUser(String userName) {
    for (User user : users) {
        if (user.getId().equals(userName))
            return user;
    }
    return null;
}
}

```

```

public static void calculateDuration () {
    for (User user : users)
        user.calculateDuration ();
}

public static void processEvents() {
    for (User user : users) {
        for (Date key : user.events.keySet()) {
            user.processEvent(key,
user.events.get(key));
        }
        user.MAD = calculateMAD(user);
    }
}

public static double calculateMAD(User user) {
    double[] times = new
double[user.sessions.size()];
    double[] abs = new double[user.sessions.size()];
    for (int i = 0; i < user.sessions.size(); i++) {
        Session session = user.sessions.get(i);
        Calendar cal1 =
GregorianCalendar.getInstance();
        cal1.setTime(session.getBegin());
        int hour = cal1.get(Calendar.HOUR_OF_DAY);
        int min = cal1.get(Calendar.MINUTE);
        int sec = cal1.get(Calendar.SECOND);
        double timeInSeconds = hour * 3600 + min *
60 + sec;
        session.setBeginInSeconds(timeInSeconds);
        times[i] = timeInSeconds;
    }

    double mean = median(times);
    int midnight = 86400;
    for (int i = 0; i < user.sessions.size(); i++) {
        double diff1;
        diff1 =
user.sessions.get(i).getBeginInSeconds() - mean >= 0
?
user.sessions.get(i).getBeginInSeconds() - mean
: mean -
user.sessions.get(i).getBeginInSeconds();
        double diff2 = midnight -
user.sessions.get(i).getBeginInSeconds() + mean;
        if (diff1 < diff2)
            abs[i] = diff1 >= 0 ? diff1 : -diff1;
        else abs[i] = diff2 >= 0 ? diff2 : -
diff2;
    }
}

```

```

    }

    return median(abs);
}

static double median(double[] values) {
    Arrays.sort(values);
    double median;
    int totalElements = values.length;
    if (totalElements % 2 == 0) {
        double sumOfMiddleElements =
(values[totalElements / 2] + values[totalElements / 2 -
1]);
        median = ((double) sumOfMiddleElements) / 2;
    } else {
        median = (double) values[values.length / 2];
    }
    return median;
}
}

```

```
package com.lmooc;
```

```
import java.util.Date;
```

```
public class Event {
```

```
    String url;
```

```
    Date date;
```

```
    public Event(String url, Date date) {
```

```
        this.url = url;
```

```
        this.date = date;
```

```
    }
```

```
    public String getUrl() {
```

```
        return url;
```

```
    }
```

```
    public void setUrl(String url) {
```

```
        this.url = url;
```

```
    }
```

```
    public Date getDate() {
```

```
        return date;
```

```
    }
```

```
    public void setDate(Date date) {
```

```

        this.date = date;
    }
}

package com.lmooc;

import java.util.Date;

public class Session {
    String id;
    Date begin;
    Date end;
    double secondsBegin;

    public void setId(String id){
        this.id = id;
    }
    public String getId(){
        return id;
    }
    public void setBegin(Date begin){
        this.begin = begin;
    }
    public Date getBegin(){
        return begin;
    }
    public void setBeginInSeconds(double seconds){
        this.secondsBegin = seconds;
    }
    public double getBeginInSeconds(){
        return this.secondsBegin;
    }
    public void setEnd(Date end){
        this.end = end;
    }
    public Date getEnd(){
        return end;
    }

    public int getDuration(){
        return 0;
    }
}

```

```

package com.lmooc;
import java.util.ArrayList;
import java.util.Calendar;
import java.util.Date;

```

```

import java.util.GregorianCalendar;
import java.util.TreeMap;
import java.util.concurrent.TimeUnit;

public class User {
    String id;
    ArrayList<Session> sessions;
    TreeMap<Date, String> events;
    Date lastDate;
    int counter=0;
    String lastSession="session"+counter;
    int userType=10;
    long duration;
    double MAD = 0;
    boolean homepage = false;
    boolean welcome = false;
    boolean ini = false;
    boolean initial_survey = false;
    boolean one0 = false;
    boolean one1 = false;
    boolean one2 = false;
    boolean one3 = false;
    boolean one4 = false;
    boolean one5 = false;
    boolean one6 = false;
    boolean two0 = false;
    boolean two1 = false;
    boolean two2 = false;
    boolean two3 = false;
    boolean two4 = false;
    boolean two5 = false;
    boolean two6 = false;
    boolean three0 = false;
    boolean three1 = false;
    boolean three2 = false;
    boolean three3 = false;
    boolean three4 = false;
    boolean three5 = false;
    boolean three6 = false;
    boolean four0 = false;
    boolean four1 = false;
    boolean four2 = false;
    boolean four3 = false;
    boolean four4 = false;
    boolean four5 = false;
    boolean four6 = false;
    boolean final_survey = false;
    boolean progress = false;
    boolean transcripts = false;
    boolean resources = false;
}

```

```

boolean request_cert = false;
boolean vocab = false;
int forum = 0;

float grade, act1, act2, act3, act4, act5, act6, act7,
act8, act9, act10, act11, act12, actAver, actFinal1,
actFinal2, actFinal3, actFinal4, actFinalAver;
boolean certEligible = false;
boolean certSent = false;
public void setId(String id){
    this.id = id;
}
public String getId(){
    return id;
}
public User (String id){
    this.id = id;
    sessions = new ArrayList<Session>();
    events = new TreeMap<Date, String>() ;
}

public void addEvent(Date time, String url){
    events.put(time, url);
}
public void processEvent(Date time, String url){
    if(lastDate==null){
        Session s = new Session();
        s.setId(lastSession);
        s.setBegin(time);
        s.setEnd(time);
        sessions.add(s);
        lastDate = time;
    }else{
if(Utills.getDateDiff(lastDate,time,TimeUnit.MINUTES)>30){
        counter++;
        lastSession="session"+counter;
        Session s = new Session();
        s.setId(lastSession);
        s.setBegin(time);
        s.setEnd(time);
        sessions.add(s);
        lastDate = time;
    } else{
        for(int i=0; i<sessions.size(); i++) {
if(sessions.get(i).getId().equals(lastSession)) {
                sessions.get(i).setEnd(time);
                lastDate = time;
            }
}
}
}
}

```

```

        }
    }
}

    increaseCounters(url);
}

    public void calculateDuration(){
        for(Session session : sessions){
            long diff = Uutils.getDateDiff(session.getBegin(),
session.getEnd(), TimeUnit.MINUTES);
            duration = duration +
Uutils.getDateDiff(session.getBegin(), session.getEnd(),
TimeUnit.MINUTES);
        }
    }

    public Session containsSession(String sessionNumber){
        for(Session session : sessions){
            if(sessionNumber==session.getId()) return
session;
        }
        return null;
    }

    public void increaseCounters(String url){

        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/course/")) homepage = true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/902a4e0dd1114258a2c
9221b5ad9ab98/d6bf24eb4c67420ca91b7756a525e0af/")) welcome
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/902a4e0dd1114258a2c
9221b5ad9ab98/5ef8f146d31743c58bd11523cbf8c7d4/")) ini=
true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/902a4e0dd1114258a2c
9221b5ad9ab98/fe8158d628d345b7b432f7c2ba286464/"))
initial_survey = true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57
16a1a23934d84/2c7aba49af244f1984f455569ef0e60e/")) one0 =
true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57

```

```

16a1a23934d84/4d10bad23b7347a3875c2272754286d3/")) one1 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57
16a1a23934d84/b7e2e13218974816a1500d6c8bb6410b/")) one2 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57
16a1a23934d84/e10b642d51074491b0ea56fa54a568ef/")) one3 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57
16a1a23934d84/455c3b20c69e472abd72e0be7221f95e/")) one4 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57
16a1a23934d84/621976f987304dec8c3488cb207c42b4/")) one5 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57
16a1a23934d84/fa3507abd9bd46569edbe02b81582e17/")) one6 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/41126fc5aa1d47aea57
16a1a23934d84/426b71d88f4a4754bac3fb5e2a1b6a51/"))
forum++;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/59813c692d5e44ce8b96f01b0cdaeab9/")) two0 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/e41ffa70ec594afdae93edeb92d343ff/")) two1 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/ca98241418db45fc925a2562ac2966d8/")) two2 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/96c0d0ffb7014579b331fc04441f914e/")) two3 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/671f90fa38c345ed993becae38017140/")) two4 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/fceb8f055ae346a09ded134be001fba5/")) two5 =
true;

```

```

        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/a3fa76e04eff47f2a6d5d699e4f7555c/")) two6 =
true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/3b2d2bf1f32a4bdda61
4917aaa67df78/51679f1fe4b046d6bd5d44c2daa32b0e/"))
forum++;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/39055c31d427453d92e133c2dec4e8f/")) three0
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/e1744dff0f3549e28bf2d8b6cdf8e8fc/")) three1
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/6aa96e1bd0914ee496eafb4fdd33eaa/")) three2
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/a3fd35cb5b0043d9a649b73c37d118c6/")) three3
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/1f305ff3f8324e6d9907c052da1ccef7/")) three4
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/3b900575775d4248915f2ef63d217da3/")) three5
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/d18b0e6689094d01befc698ebbeb3710/")) three6
= true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/00ccd40329d34752914
23755eaed58a1/53b5984233c84dea9f3392c093e7a6aa/"))
forum++;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924
d80a02aa83143/7ab3641393de44c7b8215e80d01f313a/")) four0 =
true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924
d80a02aa83143/03d6335005c44162a25a4f1e6d6b8730/")) four1 =
true;
        if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924

```

```

d80a02aa83143/344bcd7ecb0f4a2ba2c585a515b3a81d/")) four2 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924
d80a02aa83143/34134acfab24452ab6286337e93a3084/")) four3 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924
d80a02aa83143/f6b603d8b71647f292bab2c60630e953/")) four4 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924
d80a02aa83143/c0c86f3a3d1a40769c2a12aa5a2f35be/")) four5 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924
d80a02aa83143/f086c1007df3424ab53bf2d00a99c1c2/")) four6 =
true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/d7470c41aa40418a924
d80a02aa83143/404d225f5e8d4c9dacd13cb68b944f39/"))
forum++;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/bf45cf48fddd428991c
e42fef854a5f9/8bac6fca45f84f01b501c3c491af5395/"))
final_survey = true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/ff2f49d5145642d093365d60084e95
a7/")) vocab = true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/6a9519eda2b64aae9877bc980889a5
fe/")) transcripts = true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/51784f4703e8410ca86be07a03f0b1
12/")) resources = true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/courseware/bf45cf48fddd428991c
e42fef854a5f9/a6b9067964324db582b890361b33bce4/"))
request_cert = true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/progress")) progress = true;
    if(url.contains("/course-
v1:UNED+MOONLITE_I_001+2018/discussion/forum/")) forum++;
}
public boolean moduleIni(){
    return(ini);
}
public boolean upMenu(){

```

```

return(vocab||request_cert||transcripts||resources||progre
ss);
}
public boolean module1Completo(){
    return(one1&&one2&&one3&&one4&&one5&&one6);
}
public boolean module2Completo(){
    return(two1&&two2&&two3&&two4&&two5&&two6);
}
public boolean module3Completo(){

return(three1&&three2&&three3&&three4&&three5&&three6);
}
public boolean module4Completo(){
    return(four1&&four2&&four3&&four4&&four5&&four6);
}
public boolean module1Acceso(){
    return(one1||one2||one3||one4||one5||one6);
}
public boolean module2Acceso(){
    return(two1||two2||two3||two4||two5||two6);
}
public boolean module3Acceso(){

return(three1||three2||three3||three4||three5||three6);
}
public boolean module4Acceso(){
    return(four1||four2||four3||four4||four5||four6);
}
public boolean allModulesCompleto(){
    return(module1Completo()&& module2Completo()&&
module3Completo()&&
module4Completo()&&moduleIni()&&homepage&&welcome&&final_s
urvey&&resources&&request_cert&&progress&&transcripts&&voc
ab&&forum>0);
}
}

package com.lmooc;

import java.io.BufferedWriter;
import java.io.FileNotFoundException;
import java.io.FileWriter;
import java.io.IOException;
import java.util.ArrayList;
import java.util.Calendar;
import java.util.Date;
import java.util.StringTokenizer;
import java.util.concurrent.TimeUnit;

```

```

public class Utils {

    public static boolean compareTimes(Calendar
reference1, Calendar reference2, Calendar cal1, Calendar
cal2){
        boolean solapan1 = true;
        boolean solapan2 = true;

        int hour1 = reference1.get(Calendar.HOUR_OF_DAY);
        int min1 = reference1.get(Calendar.MINUTE);
        int hour2 = cal1.get(Calendar.HOUR_OF_DAY); //
gets hour in 24h format
        int min2 = cal1.get(Calendar.MINUTE);
        int suma1 = hour1*60+min1;
        int suma2 = hour2*60+min2;
        if(suma1>=suma2){
            if(suma1-suma2>90)solapan1 = false;
        }else if(suma2-suma1>90)solapan1 = false;

        int hour3 = reference2.get(Calendar.HOUR_OF_DAY);
        int min3 = reference2.get(Calendar.MINUTE);
        int hour4 = cal2.get(Calendar.HOUR_OF_DAY); //
gets hour in 24h format
        int min4 = cal2.get(Calendar.MINUTE);
        int suma3 = hour3*60+min3;
        int suma4 = hour4*60+min4;
        if(suma3>=suma4){
            if(suma3-suma4>90)solapan2 = false;
        }else if(suma4-suma3>90)solapan2 = false;

        return (solapan1||solapan2);
    }

    public static long getDateDiff(Date date1, Date date2,
TimeUnit timeUnit) {

        long diffInMillies = date2.getTime() -
date1.getTime();
        return
timeUnit.convert(diffInMillies,TimeUnit.MILLISECONDS);
    }

    public static User generateStatistics(ArrayList<User>
users, String excluded) {

        try {
            FileWriter fileWriter0 = new
FileWriter("LM00C.csv");

```

```

        BufferedWriter bufferedWriter0 = new
BufferedWriter(fileWriter0);

        bufferedWriter0.write("id,time,mark,disp,sessions,for
um" );
        bufferedWriter0.newLine();

        for (User user : users) {
            StringTokenizer st = new
StringTokenizer(excluded,";");
            boolean skip = false;
            while (st.hasMoreTokens()) {
                String token = st.nextToken();

                if(user.getId() == null
|| user.getId().equals("") || user.getId().equals(token))
skip = true;

            }

            if (!skip) {
                if(user.grade>0) {

                    bufferedWriter0.write(user.getId() + "," +
Long.toString(user.duration)+ "," +user.grade+ ","
+user.MAD+ "," +user.sessions.size() +"," +user.forum);
                    bufferedWriter0.newLine();
                }
            }
        }

        bufferedWriter0.close();

    } catch (FileNotFoundException ex) {
        System.out.println("Unable to open file");
    } catch (IOException ex) {
        System.out.println("Error reading file");
    }
    return null;
}
}

```