Exploración de técnicas cuantitativas de modelado temporal orientadas a la mejora del rendimiento académico en educación superior *online*

Juan Antonio Martínez-Carrascal

Departamento de Informática, Multimedia y Telecomunicaciones

Universitat Oberta de Catalunya

Doctorado en Educación y TIC (*eLearning*)

Tesis por compendio de publicaciones

Universitat Oberta de Catalunya 2023

A mis padres, por ser el exponente máximo del amor y la dedicación a sus hijos. Nada de lo que haga podrá llegar nunca a compararse con lo que han hecho por mí.

A Emma, por los momentos compartidos, que espero se alarguen hasta el infinito.

A mis hijos, Álex y Daniel, por ser uno de los motores de mi vida, y por conseguir que una de las cosas que más quiero sea verlos crecer día a día.

A mis hermanas, Carmen e Inma, o Inma y Carmen. Por lo mucho que suponen para mí.

Declaración previa

Por la presente declaro que el contenido de esta tesis es original y fruto del resultado de mi propio trabajo. En aquellos casos en que se otros autores han contribuido en el proceso de investigación se realiza cita expresa de dicha contribución. Considerando que la presente tesis se realiza por compendio de publicaciones, esto aplica en particular a los artículos cuya publicación ha derivado de este trabajo de investigación. Igualmente, los trabajos previos de otros autores que sustentan el trabajo se han citado en la medida en que han sido utilizados.

La presente tesis se presenta para completar el Doctorado en Educación y TIC (*eLearning*) en la Universitat Oberta de Catalunya y no ha sido enviada ni está en consideración para ningún otro título o calificación en esta u otra Universidad.

Juan Antonio Martínez-Carrascal

2023

Agradecimientos

Completar una tesis es un trabajo arduo, largo, y a menudo descorazonador. Sería por ello injusto atribuirme de forma individual el mérito de haber llegado hasta aquí.

No puedo empezar sin agradecer a mis padres los esfuerzos de estos ya muchos años, sin esperar retorno alguno. Su ahínco para que tuviera aquellas oportunidades de las cuales ellos no pudieron disponer. Por darme – darnos – siempre todo lo necesario. Gracias por anteponer nuestro bienestar y nuestra educación a cualquier otra cosa, costara lo que costara. Gracias por ser mis primeros profesores. Gracias por estar siempre a mi lado. Gracias, por todo.

En lo estrictamente vinculado a este trabajo, quiero agradecer a la Dra. Dolors Márquez su ayuda en los aspectos estadísticos. A Inma Martínez y Carmen Martínez, sus aportaciones valiosísimas en aspectos de modelización. Al Dr. Martin Hlosta y el Dr. Jorge Muñoz-Gama, la posibilidad de trabajar con ellos. Un lujo en ambos casos. Al Dr. Fernando de la Torre, sus consejos, antes durante, y espero que después de este trabajo. A Óscar González, por ser más que un amigo, y por su ayuda incondicional siempre que los aspectos prácticos lo requieren. Y al grupo de investigadores de SNOLA, por la aportación de valiosas ideas durante estos años.

Y necesariamente, debo y quiero concluir agradeciendo la dedicación y confianza de dos personas, sin las cuales este trabajo hubiera sido imposible. Quiero agradecer a la Dra. Elena Valderrama el empujarme a iniciar este camino. Sin ella, este trabajo simplemente no existiría. Gracias, Elena, por la confianza que siempre me has mostrado. Por tu trabajo infatigable siempre que hemos coincidido. Por aportarme la satisfacción de trabajar con alguien como tú.

Y por supuesto, a la Dra. Teresa Sancho. Gracias, Teresa, por ayudarme a completarlo. Por las muchas horas dedicadas. Por la capacidad de tener siempre un concepto que ofrecer, una línea que abrir, una idea, un problema, y una guía de solución. Gracias por ofrecerme siempre un reto, y por hacer fácil lo difícil. Gracias por animarme a hacer que los números hablen.

Gracias a las dos, porque os considero las personas más adecuadas para dirigir no este, sino cualquier trabajo en que se plantee una posible mejora académica. Ambas sentís la docencia como una vocación, en la que ofrecéis lo máximo por aptitud y actitud. Nadie mejor que vosotras para transmitir qué es y cómo debe entenderse la docencia.

Abstract

Mejorar el rendimiento académico es uno de los retos más relevantes del mundo de la educación. Un bajo rendimiento frena el avance social, tanto a nivel individual como colectivo, mientras que su mejora, contribuye al progreso. Focalizando en la educación superior, incrementar el rendimiento académico aumenta la competitividad laboral y reduce el desempleo, además de optimizar el uso de los recursos públicos destinados a la educación.

El presente estudio explora técnicas cuantitativas que, profundizando en la comprensión de las razones subyacentes a un bajo rendimiento, puedan contribuir a la mejora académica, incluyendo tanto la reducción del abandono como a la mejora del logro en términos generales. El foco en la comprensión, y no simplemente en la predicción comporta entender el aprendizaje como un continuo, y empuja a centrarse en técnicas que puedan modelar la evolución temporal de los procesos analizados. La exploración de técnicas irá siempre acompañada de su aplicación práctica, que se realizará en cursos de educación superior con una componente *online.*

Tras una prospección preliminar, dos técnicas han centrado nuestra investigación. Se trata de técnicas de uso poco común en problemas vinculados a la mejora del rendimiento académico, pese a ser habituales en otras disciplinas científicas. En primer lugar, la aplicación del análisis de supervivencia al estudio del abandono a nivel de curso orientado a su reducción. En segundo, el uso de técnicas de minería de procesos para la modelización del proceso de aprendizaje aportando la visión de qué hace – y que no hace – el estudiante, qué caminos de aprendizaje se demuestran más convenientes en términos de resultados, y hasta qué punto el estudiante se separa de ellos.

La investigación realizada se materializa en ocho artículos, donde se realiza la conceptualización de técnicas, se revelan causas tras el bajo rendimiento y se aportan propuestas para mejorarlo. Más allá de la contribución metodológica, queremos destacar algunos resultados de su aplicación práctica. En el caso del análisis de supervivencia, ha permitido detectar colectivos vulnerables, factores de riesgo asociados al abandono, y cuantificar el impacto en ambos casos. Por lo que respecta a la metodología para la modelización de caminos de aprendizaje, ha permitido cuantificar el seguimiento que hace el estudiante del camino pautado, y la relación entre este seguimiento y su resultado académico.

Consideramos esta última aportación metodológica, basada en la minería de procesos, y muy específicamente en el uso de *skeletons*, como particularmente relevante, tanto por lo novedoso de la aproximación como por su potencial de aplicación en cualquier entorno educativo, y muy específicamente en entornos *online*.

El trabajo abre además interesantes líneas de investigación futura. En primer lugar, sugerimos el uso de las metodologías planteadas más allá de los cursos de educación superior. En particular, animamos a la investigación en otros niveles educativos (por ejemplo, pre-universitarios) y más allá del nivel de curso (por ejemplo, etapa educativa o titulación en el caso de estudios superiores). En segundo, la modelización sugerida de caminos de aprendizaje abre interesantes líneas de estudio en aspectos como su integración en el diseño de intervenciones, o la determinación de aquellos caminos más convenientes en términos de rendimiento, contribuyendo en ambos casos a la mejora académica, objetivo último tras la prospección de técnicas que realiza este trabajo.

Contenido

Lista de figuras

Lista de tablas

Acrónimos

Lista de acrónimos usados en el presente trabajo:

AUC: *Area Under Curve*

EDM: *Educational Data Mining*

EPM: *Educational Process Mining*

GPA: *Gradded Point Average*

GBT: *Gradient Boosted Tree*

k-NN: *k-Nearest Neighbors*

GSP: *Generalized Sequential Pattern*

LMS: *Learning Management System*

MOOC: *Massive Open Online Course*

ROC: *Receiver Operating Characteristics*

SPM: *Sequential Pattern Mining*

SVM: Support Vector Machine

Capítulo 1 Introducción

"Educad a los niños y no será necesario castigar a los hombres."

Pitágoras (569 a.C.-475 a.C)

Mejorar los resultados académicos es una de las principales preocupaciones de todos los actores vinculados a la educación, y finalmente, de la sociedad en general. Desde una óptica macro, hay una correlación entre logro académico y prosperidad ciudadana (Blundell et al., 2005; Brunello & Comi, 2004; Card, 1999; Tamborini et al., 2015). Desde una perspectiva puramente académica, los resultados son un indicador clave en la medida de calidad de los estudios. Esta razón justifica el interés de las instituciones de educación superior en mejorar el diagnóstico y la comprensión de aspectos como el abandono académico o el bajo rendimiento y en diseñar intervenciones orientadas a mejorar los indicadores asociados.

La investigación educativa ha abordado este tema con particular interés en los últimos años. El análisis de la literatura muestra que la aproximación más frecuente pone el foco en la predicción de factores como el abandono o un bajo resultado, más que en la interpretación de las causas subyacentes. Hay un interés creciente por explorar nuevas técnicas y posibles algoritmos que incrementen las ratios de predicción. En paralelo, y dada la creciente disponibilidad de datos -particularmente en entornos *online*-, se ha trabajado en la identificación de variables relevantes para alimentar modelos predictivos.

Este foco en la predicción comporta a menudo que la interpretación educativa quede en un segundo término. A modo de ejemplo, métodos como las redes neuronales pueden aportar altas tasas de predicción, pero adolecen a menudo de capacidad interpretativa. En cuanto a las variables consideradas relevantes, con frecuencia no son propiamente la causa, sino que enmascaran otras situaciones. Es deseable la exploración de técnicas que focalicen en la interpretabilidad de los resultados y faciliten su discusión en profundidad. El correcto análisis de estos resultados debe permitir el diseño de intervenciones orientadas a la mejora académica.

Como factor adicional, buena parte de estos trabajos predictivos (Conijn et al., 2017) focalizan en visiones estáticas, pese a la naturaleza dinámica del aprendizaje. Los trabajos que consideran el efecto temporal (Ameri et al., 2016; Emmert-Streib & Dehmer, 2019) reflejan la conveniencia y el factor diferencial que puede aportar el análisis temporal. Por este motivo, nuestra exploración partirá del estudio de métodos predictivos más comunes en la literatura (Martínez et al., 2019; Martínez-Carrascal, Márquez Cebrián, et al., 2020), pero evolucionará hacia técnicas donde la consideración temporal es un aspecto clave.

El presente estudio se centra en **explorar técnicas que permitan modelizar el proceso de aprendizaje a nivel de curso, contribuyendo a la mejora académica.** Tras la fase de exploración metodológica, abordaremos dos ejes de mejora. El primero, la reducción del abandono, problema que por su especificidad abordaremos mediante técnicas de análisis de supervivencia. El segundo, la mejora global de resultados. Para este aspecto, tendremos en cuenta la relevancia de los caminos de aprendizaje y abordaremos su modelización a nivel de curso y la cuantificación de su seguimiento por parte del estudiante. La aproximación en ambos casos será cuantitativa, y con un foco en la aplicabilidad, cosa que comporta tener en mente el diseño de intervenciones. El objeto de estas intervenciones será la mejora de resultados académicos en cursos reglados de educación superior, y específicamente en entornos que incluyen una componente virtual.

Estamos por tanto ante un trabajo que entronca claramente con la analítica de aprendizaje (que usaremos habitualmente en este trabajo por su nombre inglés, *learning analytics),* disciplina que busca mediante el análisis de datos contribuir a la comprensión y mejora del proceso de aprendizaje. Pese a que el análisis podría realizarse en cualquier etapa educativa, el foco del presente trabajo se situará en la educación superior, concretamente, en aquellas modalidades con una componente en línea significativa.

El interés por el ámbito de la educación superior *online* se justifica por las mayores tasas de fracaso respecto a la educación presencial que demuestran las cifras y que reflejan diferentes estudios (Bawa, 2016; Simpson, 2010, 2013). Añadir que el foco específico en los estudios *online*, se justifica por varias causas. En primer lugar, por la creciente penetración de este tipo de estudios en la educación superior. Actualmente, constituyen no solo una opción preferente para una parte de la sociedad, sino la única alternativa para personas que no tienen acceso a educación superior presencial por razones de horario o distancia, o inluso por tener algún tipo de discapacidad.

En segundo lugar, la pandemia derivada del SARS-CoV-2 obligó a las instituciones a ofrecer alternativas basadas en la no presencialidad para seguir manteniendo la actividad formativa. Actualmente, y con la experiencia a raíz de la pandemia, los modelos híbridos son utilizados en muchos programas.

Finalmente, conviene también remarcar que el impacto de aspectos como el abandono es incluso mayor en estudios *online* que en otras modalidades (Bawa, 2016; Nistor & Neubauer, 2010), lo que comporta que es un ámbito particularmente adecuado para este tipo de estudios.

Desde un punto de vista metodológico, las técnicas cuantitativas basadas en evidencias son particularmente apropiadas en entornos *online*, donde es posible registrar la actividad de estudiantes y profesorado a través de las plataformas virtuales. En estos entornos, la perspectiva de las analíticas de aprendizaje (*learning analytics*) para la comprensión y mejora de los procesos de enseñanza y aprendizaje es especialmente conveniente.

Focalizando aún más, y tras una exploración preliminar, consideramos que un análisis mediante técnicas que incorporan el factor tiempo como eje vertebrador del proceso de enseñanza y aprendizaje es especialmente adecuado. El bajo rendimiento no obedece a un hecho abrupto y puntual (Ameri et al., 2016) sino que deriva de una evolución temporal. Estudios recientes (Emmert-Streib & Dehmer, 2019) subrayan la relevancia de la aproximación longitudinal al problema. Algunos autores consideran que este tipo de aproximaciones superan a otras en cuanto a interpretación de resultados (Ameri et al., 2016).

Fruto de este foco temporal, emerge la idea del aprendizaje como proceso. El aprendizaje se realiza a lo largo del tiempo a través de una secuencia de actividades. En nuestro trabajo, la consideración temporal se materializa primero en los estudios del análisis de supervivencia, orientado a entender cuándo tiene lugar el abandono a lo largo del curso y las razones subyacentes. Seguidamente, ampliamos esta visión conceptualizando el aprendizaje como proceso. A través de la minería de procesos (*process mining*) se analizarán los caminos de aprendizaje (*learning paths*) evaluando la conveniencia de seguir dicho camino, y permitiendo la detección de desviaciones en el comportamiento de los estudiantes. Conocido el momento en que se producen hipotéticas desviaciones, será posible establecer intervenciones orientadas a reconducir la situación.

Como ya hemos anticipado, la mejora de resultados a nivel de curso se concreta en dos ámbitos: la reducción del abandono, y el incremento de las ratios de superación del curso. En la reducción del abandono jugará un papel relevante el análisis de supervivencia. En lo que respecta a la mejora de resultados en términos generales, nos centraremos en el análisis de los caminos de aprendizaje, que modelizaremos mediante el análisis de procesos – y muy específicamente, mediante el uso novedoso de *process skeletons* –. Ambas técnicas comparten una **visión dinámica del proceso de aprendizaje y la relevancia del tiempo como factor clave** en dicho proceso.

El desarrollo de este trabajo de investigación se ha realizado mediante un compendio de publicaciones. De cara a contextualizarlos, este capítulo esboza las bases previas que permiten entender el hilo argumental que trenza los trabajos presentados (8 en total). Se expone en primer lugar de forma sintética el **problema que aborda la tesis** (Sección 1.1). Una vez focalizado se detalla el **concepto de logro académico** (Sección 1.2), para abordar seguidamente los **modelos explicativos** existentes (Sección 1.3). A continuación, se detalla cómo los conceptos reflejados en estos modelos se traducen en variables específicas (Sección 1.4). En este punto, estaremos en condiciones de exponer las **técnicas habituales de la minería de datos para su análisis** (Sección 1.5). Realizado este análisis se definirán las **preguntas de investigación** que aborda el presente estudio (Sección 1.6) así como el orden en que se presentarán las publicaciones en los capítulos sucesivos. El capítulo se cierra con la sección 1.7. donde se presenta una síntesis de los principales aspectos abordados.

1.1. Problema abordado, foco y relevancia de las aportaciones

Informes recientes de organismos internacionales, y específicamente aquellos centrados en la medida del rendimiento, muestran que existe un problema no resuelto en lo que respecta a la mejora del logro académico (Antoninis et al., 2020; European Commission, 2020; Organisation for Economic Co-Operation and Development (OECD), 2020). Muy específicamente, tanto la reducción del abandono como del fracaso escolar son objeto de investigaciones y esfuerzos de la comunidad educativa.

Estos problemas se manifiestan en las diferentes etapas educativas. Aun así, el presente trabajo focalizará en la educación superior, y específicamente en los entornos *online*. La primera razón que justifica esta elección es la relevancia de los estudios superiores en el progreso social. A nivel europeo, la Unión Europea lo remarca específicamente, indicando el impacto de su actuación tanto en la recuperación económica como en la resiliencia futura (*Higher Education Initiatives | European Education Area*, n.d.).

Dentro de la educación superior, los entornos *online* son particularmente relevantes. Hoy en día están presentes en la práctica totalidad de etapas educativas, y han permitido dar continuidad en la pandemia derivada del SARS-CoV-2, aportando soluciones incluso en entornos educativos fundamentados en la presencialidad. De nuevo, a nivel europeo, esto ya quedaba reflejado en la comunicación de la Comisión al Parlamento Europeo de mayo de 2017 (Commission, 2017), donde se remarca el papel de la tecnología en la educación y la relevancia específica de modelos abiertos y del uso de sistemas *online* y *blended*. Esta comunicación indica el que sería un punto de partida conceptual del presente trabajo:

> *"Early detection of problems is crucial for identifying the support that students need. Flexible study options (part-time or online) (...). Strategies to help disadvantaged students to access and go on to complete higher education are a promising way to achieve these objectives."*

Será éste el punto de partida de la investigación: **aportar técnicas que basándose en el análisis de la evolución del aprendizaje permitan modelar esta evolución, con el objetivo final de contribuir a la mejora del rendimiento académico en los estudios de educación superior, y muy específicamente en aquellos sustentados en entornos *online***. Con aún mayor precisión, se focalizará en el estudio a nivel de curso, por ser un componente clave sobre el que se construye el currículum académico. La relevancia del factor *online* deriva tanto de su presencia en la educación actual, como de los datos que aporta sobre la evolución del proceso de aprendizaje. El análisis realizado parte de técnicas de minería de datos (*data mining*) pero buscando siempre la interpretación en términos educativos (*learning analytics*). El foco del estudio no será pues el procedimiento o algoritmo utilizado, sino la comprensión del proceso de aprendizaje que hay detrás de los datos que explotan las diferentes técnicas planteadas.

Como muestra la literatura, y como se verá en muchos de los trabajos referenciados a lo largo de este compendio, aplicar técnicas descontextualizadas del proceso de aprendizaje comporta la pérdida de visión de una parte muy relevante del problema. La presente adopta la perspectiva de *learning analytics* para la comprensión de escenarios concretos de enseñanza y aprendizaje. Por lo tanto, el diseño o exploración de técnicas *per se* no será el foco, sino siempre su modelización del comportamiento del estudiante para comprender por qué se producen determinadas situaciones.

Desde esta perspectiva, el trabajo que se presenta, explora y adecua herramientas que permiten el diseño de intervenciones académicas orientadas a mejorar el logro académico en educación superior *online* centrándose en la comprensión del proceso y las condiciones del aprendizaje. Para ello, se utilizarán técnicas que – aun estando ya presentes en otras disciplinas – son poco habituales en la resolución de este tipo de problema.

En particular, dos serán las técnicas que resultan de interés a partir de la investigación realizada: el análisis de supervivencia y la minería de procesos. La relevancia del tiempo y del proceso seguido en el logro académico apuntan a la pertinencia de ambas técnicas, de uso común en otras disciplinas. Ambas demostrarán capacidad interpretativa y gran potencial en el diseño de intervenciones.

Los trabajos publicados pueden contribuir por tanto a **mejorar el rendimiento académico a partir de la comprensión** de cómo evoluciona el estudiante y de dónde y/o cuando puede presentarse un problema académico. Comprender el problema y las causas subyacentes permite la **intervención adecuada** – tanto en cuanto a destinatarios de la intervención como en cuanto a momento de ésta – en aquellos casos en que el análisis muestre un **riesgo potencial** en términos de rendimiento esperado a nivel de curso. Como aspecto colateral, la búsqueda de nuevas aproximaciones llevará a la translación de aspectos metodológicos vinculados originalmente a otras disciplinas científicas, pero que como se verá proporcionan resultados relevantes en el ámbito educativo.

Específicamente, los trabajos mostrarán cómo detectar factores de riesgo basados en características preexistentes, o en aspectos específicos del modelo de evaluación. Dotarán también de una visión temporal que permita establecer momentos específicos en que es conveniente plantear una intervención. Las técnicas aportarán visión sobre **cuándo intervenir**, pero – y más importante – **por qué intervenir**. Más allá de los escenarios estudiados en los trabajos publicados, el hecho de tener como referente el proceso de aprendizaje hará que los métodos aportados puedan proporcionar nuevos enfoques en el análisis futuro de problemáticas educativas.

1.2. Logro académico: concepto, medida y relevancia

El logro académico (*academic achievement*) es el término general usado para referirse a los resultados de un proceso de aprendizaje llevado a cabo en un entorno académico (Spinath, 2012). A nivel de curso, (Steinmayr et al., 2014) notan que "se relaciona comúnmente con aspectos de rendimiento que indican hasta qué punto un estudiante cumple metas específicas vinculadas a una actividad académica en una institución de educación".

Plantear un trabajo que analice su mejora requiere – más allá de la definición previa – establecer formas de medida. El concepto de logro admite aproximaciones cualitativas. Diferentes estudios, y específicamente los vinculados a la psicología educativa, presentan este tipo de aproximación. Las visiones cuantitativas, y en particular las que derivan de las analíticas de aprendizaje, buscan explicar las razones existentes tras los conceptos cualitativos. Estas aproximaciones requieren establecer claramente cómo medir el logro. Las métricas habituales son el uso de calificaciones – bien sean vinculadas a actividades específicas o a nivel de titulaciones completadas – y el uso de *tests* específicos orientados a su medida.

En entornos de educación superior, el uso de calificaciones es la forma común de evaluación de logro. A nivel universitario, el cálculo de una calificación media en particular, el *Grade Point Average* – o GPA por sus siglas en inglés – es la medida más habitual. A efectos de la presente tesis, consideraremos las calificaciones del estudiante en el curso objeto de análisis como indicadoras de la consecución del logro académico.

Esta decisión está justificada ya que diferentes estudios concluyen que la calificación es un indicador del rendimiento del estudiante, y específicamente, del logro (Cassidy, 2012; Gough, 1953; Setiawan et al., n.d.; Spinath, 2012; Tsiakmaki et al., 2019). Adicionalmente, a nivel de curso – ámbito en que se centra este trabajo –, el logro comporta necesariamente una evaluación curricular. Conviene remarcar que una vez ingresado en el sistema universitario, el método común para considerar superado un curso pasa por la evaluación del profesor que imparte la asignatura. Esta evaluación será expresada mediante una escala cuantitativa (numérica o literal) que refleja hasta qué punto el estudiante ha superado el curso. A efectos de análisis, siempre habrá también una distinción entre "superación" y "no superación". Dichos estados podrán adicionalmente desglosarse para obtener información más relevante. En particular, y dentro de la no-superación, se pondrá especial interés en el análisis de los casos en que la no superación se debe a un abandono prematuro del curso.

1.3. Modelos del logro académico

Las teorías educativas han analizado el logro académico desde hace más de 50 años. El modelo de Carroll (1963) puede considerarse el primer modelo formal. Dicho estudio no indica de forma explícita el ámbito de educación (aun cuando habla de *school learning*) pero los ejemplos presentes en dicho artículo hacen pensar en modelos orientados a las etapas iniciales de la educación. Por descontado, no contemplan entornos *online*.

Carroll focaliza en la aptitud del estudiante y el tiempo dedicado. La aptitud se traduce en un mayor o menor logro dependiendo del tiempo aplicado en el estudio, la perseverancia (que intrínsecamente conlleva también un aspecto de dedicación temporal), la capacidad de comprensión y la calidad de los eventos de aprendizaje. Gráficamente, el modelo se muestra en la Figura 1.1.

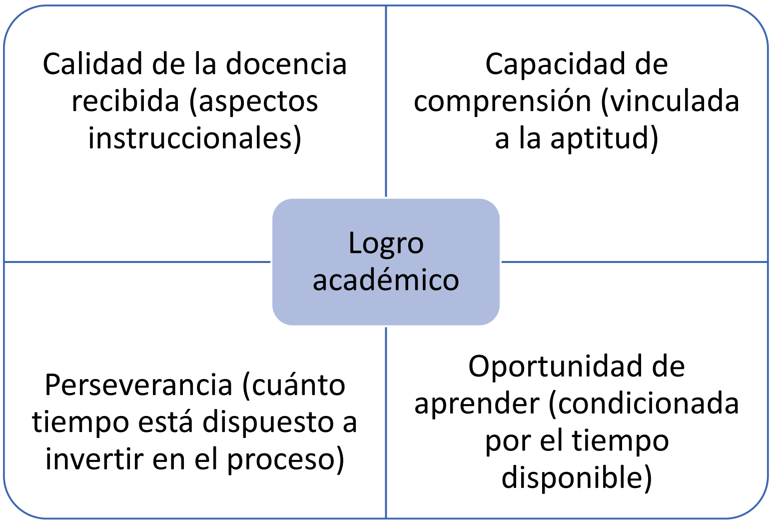


Figura 1.1– Factores condicionantes del logro académico en el modelo de Carroll

En el modelo de Carroll (pensado implícitamente para entornos presenciales) el tiempo constituye un parámetro fundamental. El tiempo dedicado efectivamente es un factor clave, y deriva tanto de la distribución temporal establecida para las diferentes asignaturas (marcado por la institución educativa) como de la dedicación temporal que realiza el estudiante fuera del aula. Ambos factores totalizan el tiempo de dedicación, fundamental para determinar el logro. Un mismo tiempo de dedicación no garantiza un mismo resultado para diferentes estudiantes, ya que el impacto del tiempo dedicado depende de la aptitud, los conocimientos previos y la calidad de la docencia que recibe. Como ejemplo, a mayor aptitud se requerirá menor tiempo de dedicación, pero a la vez, una aptitud sin dedicación no garantiza la consecución del logro.

Con posterioridad, Bloom (1974) profundiza en el modelo del logro y extiende los trabajos de Carroll y los suyos propios (Bloom & others, 1971). De hecho, se trata de un modelo global del proceso de aprendizaje, pero que permite adentrarse en el concepto de logro. En particular, segmenta las características del estudiante, diferenciando entre características cognitivas y características afectivas. Bloom muestra así que más allá del concepto de aptitud de Carroll - fuertemente vinculado a la inteligencia como tal - da cabida a factores anímicos que modelos posteriores considerarán clave. Adicionalmente, Bloom remarca la calidad de la docencia e introduce como relevante el proceso de aprendizaje. En cuanto a los resultados, el logro se segmenta en diferentes componentes, dando idea de la complejidad que subyace bajo el concepto en sí. La Figura 1.2. muestra la síntesis del modelo de aprendizaje de Bloom:

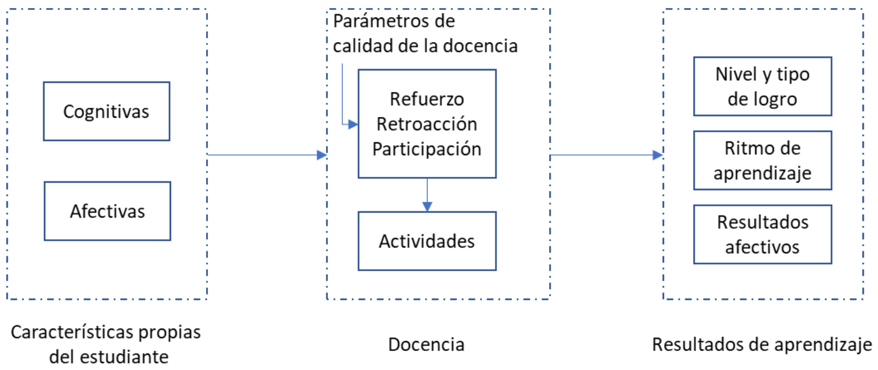


Figura 1.2–Síntesis gráfica del modelo de Bloom (basada en (Seel, 2012))

Basado en estos estudios, Walberg (1984) adopta también una visión que incluye el logro como un resultado del proceso de aprendizaje. Walberg considera que existe una productividad educativa, que se basa en nueve factores, entre los que aparecen aspectos no sólo vinculados al estudiante. El modelo da relevancia al centro educativo y a factores extracurriculares y psicosociales vinculados al entorno del estudiante. En una línea similar en lo que respecta específicamente al entorno, Proctor (1984) considera el clima social del entorno de aprendizaje como elemento clave. Bajo este paraguas se incluyen conceptos abstractos, como normas, prejuicios, ambiente escolar. También puede enmascarar variables demográficas – como el nivel de renta, o aspectos raciales –.

Notar que estas referencias siguen hablando genéricamente del aprendizaje en 'escuelas' y no específicamente en entornos de educación superior. En su artículo sobre la mejora del modelo en estos centros, Proctor remarca la importancia de las características propias del estudiante (Proctor, 1984). Sin embargo, estas características pasan por el filtro del clima escolar y el entorno directo de aprendizaje (con relevancia específica del papel del profesor). A partir de ahí, se establecen interacciones que producen oportunidades de aprendizaje y modelan las propias expectativas del estudiante en el proceso de aprendizaje. Fruto de todo ello resulta finalmente el logro académico. Gráficamente el modelo puede verse en la Figura 1.3.

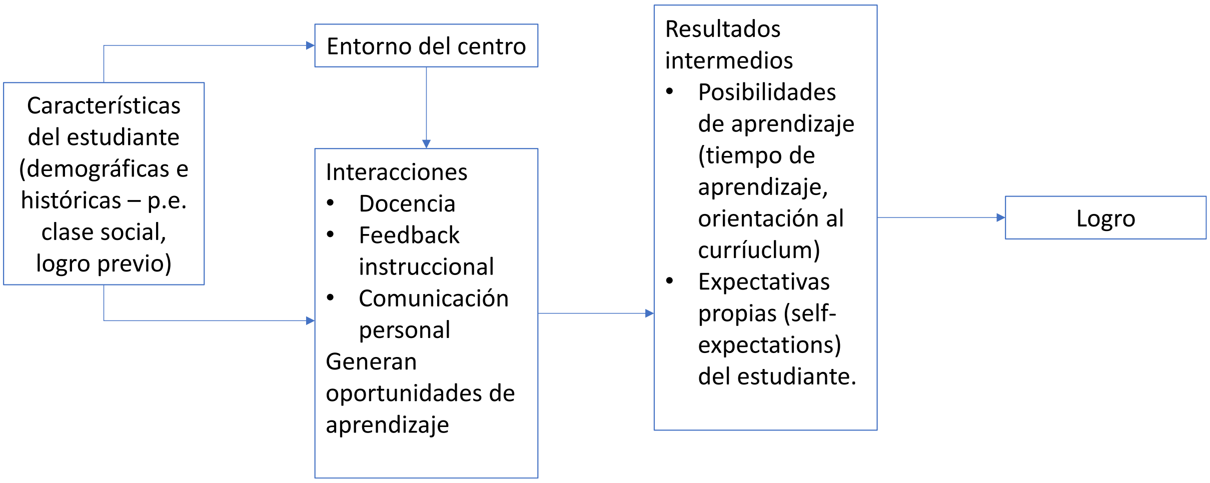


Figura 1.3–Factores que influencian el logro académico en el modelo de Proctor (basado en la reflexión sobre el modelo realizada en (Huitt, 2000))

Modelos posteriores, como el de Schereens (1990), apuntan ya claramente a la visión del aprendizaje como proceso. Este modelo, propio de la educación básica, es ya trasladable a cualquier nivel educativo (incluida la educación superior). Gráficamente, el modelo puede verse en la Figura 1.4. La idea subyacente de proceso, y de interrelaciones está presente en modelos posteriores, como el de Gage y Berliner (1992). El resultado de estos modelos se traduce en un conjunto de recomendaciones docentes que incluyen la definición clara de los objetivos o la existencia de una evaluación formal, que se asume como punto final del proceso. Si la evaluación no produce el resultado esperado, se considera que se debe volver a analizar el proceso, con el objetivo final de mejorar la docencia y con ello los resultados de aprendizaje.

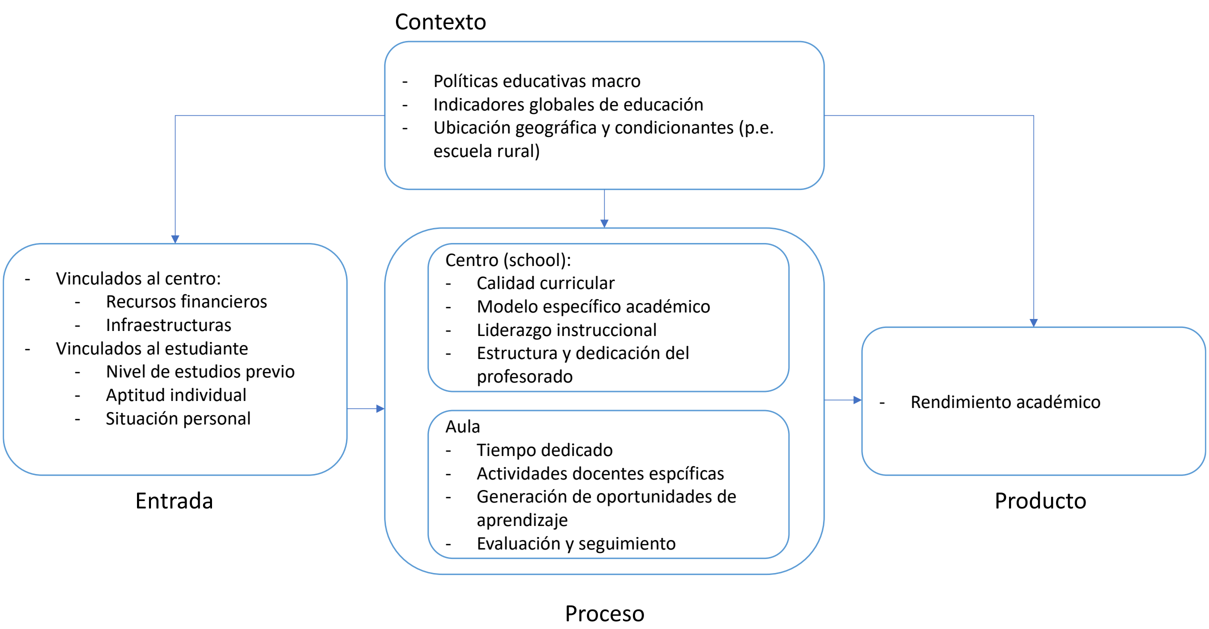


Figura 1.4– Visión gráfica del modelo contexto-entrada-proceso-producto de Scheerens

McIlrath & Huitt (1995) introducen ya la componente predictiva en el propio modelo. Consideran el aprendizaje como un proceso, que tiene lugar en un contexto determinado, recibe unas entradas y produce unos resultados. El contexto engloba aspectos como la propia escuela o la familia. Las entradas al sistema serían las aptitudes de los propios estudiantes, así como las actitudes del profesorado. Esto daría lugar a procesos de aprendizaje – con particular foco en las dinámicas de clase – y darían como resultado el aprendizaje en sí.

En la parte predictiva, el modelo de Huitt ya determina que el tiempo dedicado al estudio (*academic learning time*) es uno de los mejores predictores del éxito académico. Este concepto no es atómico, puesto que en realidad engloba diferentes ejes: su distribución, la calidad del tiempo de aprendizaje activo, las oportunidades de aprendizaje a largo plazo que tiene el estudiante y la gestión del tiempo y la planificación que realiza.

Huitt focaliza aún más en los aspectos temporales, puesto que la planificación académica de alto nivel también impacta en los aspectos temporales. En concreto, considera claves los días de que dispone el curso académico, los que realmente el estudiante asiste a clase, y las horas que el estudiante puede dedicar al día. Es remarcable que ninguno de estos aspectos figuraba en los modelos iniciales de evaluación del rendimiento.

Eccles y Wigfiled (2002) proponen la teoría del valor de las expectativas. Ambos consideran que más allá de la aptitud y el tiempo, las expectativas tienen un peso particularmente relevante. El modelo considera que la motivación es fundamental para el logro, y divide las influencias en el logro académico entre aspectos distales y aspectos proximales. Los factores distales tienen un impacto indirecto en el logro e incluyen factores sociales externos, ciertas características individuales (como el género) o experiencias previas. Estos aspectos pasan por un proceso cognitivo consciente en el estudiante, e impactan sobre los factores proximales (en particular, por ejemplo, sus expectativas de éxito). Son estos aspectos proximales, modulados, los que impactan finalmente en el logro. Se trata de un sistema con interrelaciones complejas, en el que explícitamente se considera la realimentación temporal. Gráficamente una visión simplificada del modelo se puede ver en la Figura 1.5.

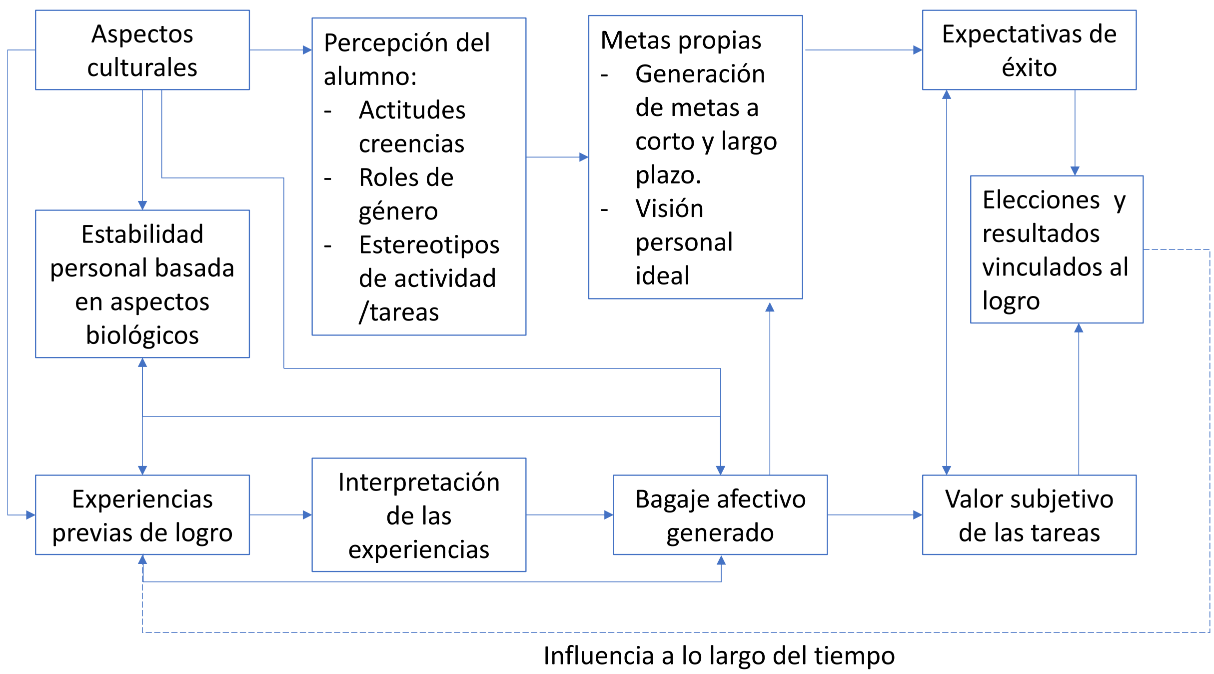


Figura 1.5– Modelo de Eccles y Wigfield de logro (basado en (Eccles & Wigfield, 2002))

El modelo de Eccles y Wigfield no se ha materializado en aplicaciones específicas, pero algunos de sus aspectos más relevantes se han trasladado a trabajos posteriores, como los de

Byrnes y Miller (2007). Estos autores proponen el modelo de oportunidad-propensión (*opportunity-propensity*) en que el logro académico depende esencialmente de la posibilidad de acceso a oportunidades de aprendizaje y su capacidad individual de aprovechar estas oportunidades.

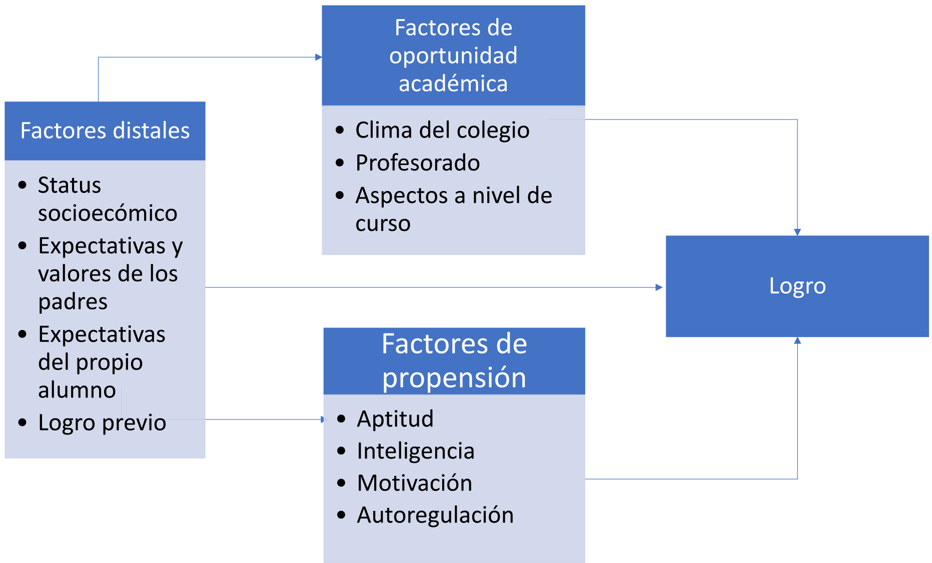


Figura 1.6– Modelo *opportunity - propensity* de Byrnes y Miller (Byrnes & Miller, 2007)

Aun cuando los modelos previos tratan el logro de forma general, el abandono ha sido tratado mediante modelos específicos. Como se verá en el capítulo 3 la definición de abandono dista de ser trivial. En particular, cuando se considera a nivel de curso la mayoría de artículos que lo analizan, no lo definen formalmente (Xavier & Meneses, 2020). A efectos de nuestra investigación, se considerará abandono el hecho de cesar en un curso sin la intención de retomarlo. Las elevadas ratios de abandono han fomentado estudios específicos tal y como reflejan diferentes compilaciones, tanto para estudios *online* como presenciales (Hachey et al., 2022; Larsen et al., 2013; Lee & Choi, 2011; Muljana & Luo, 2019).

Focalizando en el abandono, el modelo de Tinto (Tinto, 1975)- conocido como '*student attrition model'* - se considera el primer modelo formal. Según este modelo, la persistencia académica es una función de la propia motivación del estudiante, y el ajuste del propio estudiante a las características socio académicas de la institución donde realiza el estudio.

Con posterioridad a este modelo, Bean (1985) introdujo nuevos factores en el *'student attrition model'* que añade factores académicos, de entorno y de socialización. La evolución de estas teorías acaba dando lugar al modelo de Rovai (2003), que conjuga la visión del rendimiento académico y del abandono. Según esta teoría, las grandes categorías de factores influyentes son el propio estudiante (incluyendo sus características y habilidades), factores externos, y factores vinculados a la propia institución académica. Este modelo, conocido como *Composite Persistence Model* (CPM), es considerado el referente para estudios *online* (Lee & Choi, 2011).

En el caso específico de los aprendizajes virtuales, la percepción del control sobre el aprendizaje y la autorregulación se muestran como factores particularmente relevantes en este tipo de estudios (Cassidy, 2012; Findley & Cooper, 1983; Nota et al., 2004). Cho y Shen (2013) estudian específicamente el papel de la autorregulación en entornos virtuales de aprendizaje. La visión clara de la meta del estudio por parte del estudiante, así como su autoeficacia son aspectos que influyen de forma determinante en el rendimiento. Aun así, cabe decir que estos aspectos figuraban ya en algunos estudios previos no focalizados en estudios *online* (Zimmerman & Pons, 1986).

Los trabajos recientes, buscan trasladar los conceptos cuyo impacto es considerado relevante en el logro (o específicamente en el abandono) a variables concretas. Estas variables pueden extraerse bien sea de encuestas específicas – para evaluar aspectos como las propias expectativas – o bien directamente de trazas que el estudiante deja en su proceso de aprendizaje – p.e. notas intermedias o registros de asistencia a clase –. La sección 1.4. analiza algunas de las variables más comúnmente utilizadas en los sistemas de predicción del logro académico.

1.4. Variables explicativas del logro académico

La búsqueda de variables influyentes en el logro académico es incluso anterior al desarrollo de los modelos formales. Ya en 1964, Watley y Merwin (1964) analizaron cómo factores vinculados a la capacidad intelectual y a otros aspectos psicológicos influyen en el logro. Los autores apuntan a la relevancia del ranking de los estudiantes en estudios previos, así como a los resultados de *tests* específicos (en particular, los valores de las pruebas SAT en matemáticas y lengua) como predictores del logro.

Los trabajos vinculados a la predicción del logro continuaron, vinculados principalmente a estudios presenciales, y basando sus resultados en indicadores extraídos con el objeto de la predicción (por ejemplo mediante la realización de pruebas como el SAT o a través de variables que reflejen el comportamiento en clase (Hoge & Luce, 1979)). El número de estudios es ingente, como reflejan compilaciones de trabajos vinculados (por citar algunos (Agudo-Peregrina et al., 2014; Conijn et al., 2017; Hellas et al., 2018; Khan & Ghosh, 2021; Moreno-Marcos et al., 2019; You, 2016)).

Gran parte de los trabajos se centran en determinar la capacidad predictiva de variables individuales (Bowers et al., 2013; Holder, 2007; McGregor & Elliot, 2002) – una o un conjunto– pero sin relación con un modelo específico. Más que asentarse en los modelos existentes se buscan las variables que puedan predecir el logro o fracaso académico. Muchas variables son específicas (horas dedicadas, asistencia a clase…) y otras de más difícil cuantificación. A modo de ejemplo, von Stumm et al. (2011) consideran que la inteligencia es el predictor principal, y que esta inteligencia puede reflejarse en ciertos aspectos de la personalidad.

El uso masivo de sistemas de *elearning* - bien como entorno principal de docencia, bien como entorno de soporte - junto con el creciente interés por las analíticas de datos masivos, ha llevado a la búsqueda de indicadores clave en los sistemas de gestión de aprendizaje (LMS por sus siglas en inglés). Cualquier acción que realiza el estudiante queda registrada, facilitando la extracción de indicadores y la creación de modelos explicativos basados en ellos, así como la reflexión posterior a partir de los resultados. Esta modelización y análisis caracterizan la perspectiva denominada *learning analytics*. Con la eclosión de los LMS, han crecido también exponencialmente el número de trabajos que buscan mejorar el aprendizaje a partir de estos indicadores, y específicamente predecir el resultado académico basándose en ellos (Cerezo et al., 2016; Conijn et al., 2017; Macfadyen & Dawson, 2010; You, 2016).

La mayoría de los trabajos versan sobre la predicción en un único curso. Conijn et al. (2017) realizan un trabajo transversal, analizando un total de 17 cursos en los cuales analizan las potencialidades predictivas de diferentes indicadores. Las conclusiones de dicho trabajo muestran que los resultados no siempre son trasladables entre cursos. Hace falta un trabajo de contextualización para dar significado a determinadas variables, y la portabilidad entre cursos no es siempre realizable, en particular, si no se contextualiza y define adecuadamente el significado pedagógico de las variables en que se basan los modelos.

Basándose en la literatura preexistente, el estudio analiza las variables consideradas más relevantes en los trabajos previos. Se incluyen variables cuantitativas vinculadas a la actividad (número de clics, número de sesiones *online*, tiempo total en la plataforma), aspectos de periodicidad (regularidad del tiempo de estudio y de los intervalos entre sesiones), tipología de contenidos (contribuciones en foros, cuestionarios realizados, visión de las respuestas) y grados intermedios. El estudio concluye que la capacidad predictiva no es tan alta como puede parecer a priori, y que solo la calificación de las pruebas intermedias - a nivel de curso - es un indicador válido independientemente del curso analizado.

Como reflexión adicional, los autores de la compilación citada (Conijn et al., 2017) consideran que *"more elaborate theoretical reasoning is needed in learning analytics to achieve generalizable result"*. En otras palabras, se ha profundizado en la evaluación cuantitativa de los datos y en su impacto estadístico a nivel predictivo, pero no en su significado y en cómo encajan las variables seleccionadas en la literatura con los modelos previos. Se trata de análisis *'data driven'* más que *'concept driven'*. Enlazando esto con nuestra propuesta de investigación, creemos que la visión longitudinal del curso, el análisis del aprendizaje como proceso y la evaluación de los caminos de aprendizaje, aportan un valor conceptual del que adolecen las variables descontextualizadas.

Abordaremos seguidamente las técnicas que harán uso de estas variables. Sin embargo, antes queremos indicar que nuestras primeras publicaciones – reflejadas en el Capítulo 2 – se centrarán en el análisis específico de variables y métodos en un contexto de análisis del impacto de dichas variables en el logro académico (en términos de superación o no de una asignatura).

1.5. Técnicas habituales para el análisis cuantitativo de datos educativos

Las variables consideradas relevantes serán el punto de partida necesario para alimentar las diferentes técnicas usadas, bien para explicar el pasado, bien para predecir aspectos futuros. En el caso de los escenarios de *learning analytics*, pueden encontrarse referencias a gran variedad de técnicas con diferentes aproximaciones (Castro et al., 2007). La clasificación se realiza tanto en base al tipo de problema – regresión, clasificación o generación de grupos (*clustering*) – como a nivel de técnica específica.

Entre los trabajos que realizan aproximaciones predictivas, los métodos de clasificación y regresión son los más habituales. Este tipo de método es utilizado cuando se busca establecer una relación entre un conjunto de variables explicativas y una variable de salida, bien sea cuantitativa (p.e. predecir la nota final del curso) o cualitativa (supera o no el curso, abandona o no). En lo que respecta a técnicas específicas encontramos el uso de árboles de decisión (Hai-Jew, 2017; Rizvi et al., 2019), técnicas bayesianas (Payandeh Najafabadi et al., 2013) o aproximaciones basadas en lógica difusas (Nebot et al., 2006). Cuando el objetivo es determinar grupos que presentan características similares, se utilizan técnicas de *clustering*. Una revisión del uso de estos algoritmos en problemas vinculados a *elearning* puede encontrarse en los trabajos de Dutt et al. (2015).

Las técnicas de *data mining* utilizadas en problemas de *learning analytics* son diversas. Recientemente, Dutt et al. (2017) realizan una revisión sistemática de la literatura existente, donde se remarca que los algoritmos tradicionales no deberían utilizarse de forma directa en problemas de *learning analytics* y la relevancia del foco en el problema y de la necesidad de un preprocesado de datos. Por este motivo, y como punto previo, el Capítulo 2 explora algunas técnicas existentes. De hecho, dos de nuestros estudios (Martínez et al., 2019; Martínez-Carrascal, Márquez Cebrián, et al., 2020) exploraran diferentes técnicas, orientándonos a poder establecer una comparativa de métodos. Esta comparación no es directa como indican – focalizando en el abandono – Bowers et al. (2013). Su análisis de 36 trabajos predictivos muestra de forma clara que métricas como la precisión o la sensibilidad pueden no ser suficientes para evaluar la bondad de las predicciones. El trabajo propone también una comparativa gráfica que permite una mejor comparación entre métodos, y que utilizaremos para validar nuestros resultados.

Los algoritmos específicos y los indicadores de precisión son sin duda relevantes. Sin embargo, y si se considera como fin último el diseño de una intervención, un aspecto particularmente relevante es el momento en que se realiza la predicción. El tiempo juega un papel muy relevante. Las identificaciones tempranas de situaciones de riesgo son comunes en la literatura (Ameri et al., 2016; Villano et al., 2018). Por ello, el segundo de los trabajos que describimos en el capítulo 2 se orientará al impacto del tiempo en la bondad de la predicción (Martínez et al., 2019). Como veremos, la capacidad predictiva de los algoritmos, con independencia de cuál sea, mejora a medida que se incluye un mayor número de datos temporales.

Pese al gran número de trabajos de predicción temprana, e incluso algunos orientados a comparativas periódicas, ciertos autores han justificado recientemente la necesidad de aproximaciones longitudinales, que tengan en cuenta factores evolutivos y no solo visiones estáticas (Ameri et al., 2016). Ello motiva la búsqueda de métodos menos explorados como son el análisis de supervivencia o el análisis de procesos que centran dos de los capítulos de esta tesis.

Este último trabajo explora ya el uso del análisis de supervivencia para el análisis del abandono en cursos masivos abiertos (*MOOC*) apuntando ya su potencialidad y su comportamiento superior a las técnicas habitualmente exploradas permitiendo descubrir factores que no muestran otras técnicas. Siguiendo esta aproximación, uno de nuestros estudios (Martínez-Carrascal, Hlosta, et al., 2023) ha buscado justamente explicar el abandono, pero en entornos de educación superior reglada. Nuestros resultados confirman el potencial de la herramienta incluso cuando el conjunto de variables disponibles se limita a aspectos demográficos. Como veremos, esto permite la detección de poblaciones con mayor riesgo de abandono en función de sus características personales. Adicionalmente, y como muestra de la interpretabilidad de resultados, el método determina el peso relativo que tienen los diferentes factores en dicho abandono.

Siguiendo con la línea de aproximaciones que tienen en el tiempo un factor clave, y ligando con el enfoque del aprendizaje como proceso, otra de las técnicas poco exploradas que analizamos en el presente trabajo es la minería de procesos. Como se verá en el Capítulo 4, la técnica tiene su origen en el análisis de procesos empresariales. Nuestra aproximación permite detectar hasta qué punto un estudiante se desvía del comportamiento esperado. En este caso el sistema sí que requiere de datos vinculados al propio LMS. La aplicación planteada permite evaluar en cualquier momento del curso hasta qué punto el estudiante se desvía de un camino establecido, permitiendo posibles intervenciones. Adicionalmente, permite la comparación de caminos, indicando si seguir un determinado camino está relacionado con un mayor o menor logro académico a final de curso.

Finalmente, y en la línea apuntada por Mclirath y Huitt (1995), esta aproximación permite precisamente aquello que deben permitir los propios modelos: "*starting points to begin developing more appropriate educational experiences for our society's next generation*". Tanto su interpretabilidad como la aproximación temporal, lo hacen una técnica particularmente interesante para la evaluación, y finalmente la mejora del logro académico.

1.6. Preguntas de investigación y estructura de la presente memoria

El análisis del escenario planteado muestra que el estudio de los procesos de aprendizaje permite explotación adicional mediante el uso de datos obtenidos de los sistemas de aprendizaje virtual mediante la aplicación de técnicas que consideren no solo la instantánea de los datos sino su evolución temporal.

Esta tesis explora técnicas para la modelización temporal del aprendizaje. Una modelización que derive en una mejor comprensión de los factores que subyacen tras un bajo rendimiento académico, con el fin último de facilitar el diseño de posibles intervenciones que deriven en mejoras de rendimiento. La modelización planteada irá siempre acompañada de casos de estudio, que proporcionarán contextos específicos de aplicación y están orientados a resolver las preguntas de investigación que se indican a continuación.

En primer lugar, y con relación al **análisis del abandono académico orientado a su reducción**, se plantea cómo obtener información temprana que pueda permitir el diseño de intervenciones en fases iniciales. Para ello se exponen las cuestiones:

RQ1: ¿Qué factores preexistentes influyen en el abandono académico?

RQ2: ¿Cuál es el impacto específico de dichos factores en el abandono?

RQ3: ¿Cómo influyen los resultados académicos iniciales de un determinado curso en la posibilidad de abandonarlo?

Abordar estas preguntas comporta necesariamente una exploración metodológica orientada a la detección de factores de riesgo y poblaciones más vulnerables en términos de rendimiento académico. La exploración mostrará también aquellos puntos en los que focalizan las líneas de investigación actuales, así como los aspectos considerados no resueltos.

La concreción y cuantificación de los factores de riesgo, debe proporcionar guías para la intervención temprana. Al margen del abandono buscaremos también **detectar cuándo el estudiante abandona el camino de aprendizaje sugerido por el profesorado en entornos *online*, poniendo en riesgo la superación** de los objetivos académicos. Como punto preliminar, será necesario evaluar que el camino sugerido por el equipo docente favorece la consecución de las metas académicas. Este concepto se materializa en las siguientes preguntas:

RQ4: ¿Cómo puede validarse la bondad de un camino de aprendizaje en términos de superación de una asignatura?

RQ5: ¿Cómo puede establecerse una medida cuantitativa de hasta qué punto el estudiante está siguiendo el camino pautado por el docente?

Esta última pregunta tiene como fin último estimar poblaciones de riesgo basándonos en la divergencia de un posible camino de aprendizaje. Dado que el objetivo es la mejora del aprendizaje, la respuesta a las preguntas previas debe dar lugar al planteamiento de un conjunto de acciones orientadas a la reducción del suspenso y el abandono. Por ello, más allá de las preguntas planteadas, se propone reflexionar sobre la siguiente cuestión de enfoque eminentemente práctico:

Una vez realizada la modelización, ¿qué acciones específicas pueden plantearse para mejorar los resultados académicos en cursos de educación superior *online*?

Vistas las preguntas de investigación, y considerando la naturaleza de la presente tesis como compendio de publicaciones, se exponen seguidamente las publicaciones realizadas, vinculándolas a la estructura de capítulos de la presente tesis. Para simplificar su referencia posterior se ha asignado un identificador numérico a la publicación.

Tabla 1.1 – Estructura de la memoria y vínculo con las publicaciones realizadas.

Capítulo	Ámbito	Id. publicación	Detalle de la publicación
Capítulo 2	Exploración metodológica	Artículo 1	(Martínez et al., 2019) Predicting student performance over time. A case study for a blended-learning engineering course.
		Artículo 2	(Martínez-Carrascal, Márquez Cebrián, et al., 2020) Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course.
Capítulo 3	Análisis de supervivencia	Artículo 3	(Martínez-Carrascal & Sancho-Vinuesa, 2021) Does failing the first assessment affect withdrawal decisions at course level? An empirical evidence.
		Artículo 4	(Martínez-Carrascal, Hlosta, et al., 2023) Using survival analysis to identify populations of learners at risk of withdrawal:conceptualization and impact of demographics.
		Artículo 5	(Martínez-Carrascal & Sancho-Vinuesa, 2023) Exploring the impact of assessment characteristics on course withdrawal: a survival analysis approach

Capítulo	Ámbito	Id. publicación	Detalle de la publicación
Capítulo 4	Minería de procesos y análisis de trayectorias	Artículo 6	(Martínez-Carrascal, Valderrama Vallés, et al., 2020) Combining clustering and sequential pattern mining to detect behavioral differences in *log* data: conceptualization and case study.
		Artículo 7	(Martínez-Carrascal & Sancho-Vinuesa, 2022) Using Process mining to determine the relevance and impact of performing optional quizzes before evaluative assessments.
		Artículo 8	(Martínez-Carrascal, Munoz-Gama, et al., 2023) Evaluation of Recommended Learning Paths using Process Mining and Log Skeletons: Conceptualization and Insight into an *Online* Mathematics Course

La Tabla 1.1 muestra la evolución que ha seguido el camino de investigación. Se empieza mediante una exploración de diferentes técnicas (Capítulo 2), que mostrará la **relevancia del tiempo y de un análisis longitudinal**. Eso permitirá proponer dos técnicas donde el tiempo juega un papel clave y que permiten su aplicación en nuestro ámbito de estudio. La primera será el **análisis de supervivencia**, que como veremos será de particular utilidad a la hora de analizar el abandono (Capítulo 3). La segunda, la **minería de procesos (*process mining*)**, técnica comúnmente utilizada en análisis de procesos empresariales, pero de uso residual en la analítica académica (Capítulo 4). Esta técnica nos permitirá explorar los caminos de aprendizaje que siguen los estudiantes, validando si se ajustan a los esperados, cuantificando las desviaciones y pudiendo establecer momentos de intervención. Nos resultará particularmente útil el uso de los llamados *skeletons*, con los cuales planteamos una técnica de modelado de los caminos de aprendizaje. Finalmente, un último capítulo (Capítulo 5) sintetiza los resultados, asociando además las publicaciones planteadas a las diferentes preguntas de investigación. Finalmente, un último capítulo (Capítulo 5) sintetiza los resultados, asociando además las publicaciones planteadas a las diferentes preguntas de investigación.

La relación explícita entre los artículos y las preguntas de investigación se muestra en la Tabla 1.2:

Tabla 1.2 – Relación de las publicaciones con las preguntas de investigación

Artículo	**RQ1**	**RQ2**	**RQ3**	**RQ4**	**RQ5**	**Comentarios**
Artículo 1 (Martínez et al., 2019)						Exploración inicial de aspectos metodológicos
Artículo 2 (Martínez-Carrascal, Márquez Cebrián, et al., 2020)						Profundización en aspectos metodológicos, y bases para abordar el análisis temprano vinculado a RQ3
Artículo 3 (Martínez-Carrascal & Sancho-Vinuesa, 2021)						Constatación de la relevancia del tiempo en el rendimiento, y en particular en el abandono temporales
Artículo 4 (Martínez-Carrascal, Hlosta, et al., 2023)						Explicación de las causas subyacentes en el abandono mediante análisis de supervivencia

Artículo	RQ1	RQ2	RQ3	RQ4	RQ5	Comentarios
Artículo 5 (Martínez-Carrascal & Sancho-Vinuesa, 2023)						Impacto de aspectos de diseño de curso en el abandono, obtenidos mediante análisis temporal
Artículo 6 (Martínez-Carrascal, Valderrama Vallés, et al., 2020)						Aspectos metodológicos de exploración de RQ4/RQ5
Artículo 7 (Martínez-Carrascal & Sancho-Vinuesa, 2022)						Aspectos metodológicos de exploración de RQ4/RQ5
Artículo 8 (Martínez-Carrascal, Munoz-Gama, et al., 2023)						Modelización del aprendizaje como proceso. Propuesta metodológica para modelizar un camino de aprendizaje, y evaluar el seguimiento por parte del estudiante

Respecto a la traslación al terreno práctico planteado en la pregunta referente al diseño de acciones, los diferentes trabajos la contemplan debido a su naturaleza transversal y a que finalmente, pese a no ser una pregunta de investigación formal, es un objetivo último de la presente tesis.

1.7. Conclusiones

El presente capítulo ha puesto de manifiesto que existe un problema no resuelto en lo que respecta a la mejora del rendimiento académico, justificado además la relevancia de explorar técnicas que contribuyan a esta mejora. Será este el **problema abordado** por el presente estudio.

Se han mostrado las bases teóricas de los diferentes modelos existentes explicativos del logro académico junto con una primera prospección de variables predictivas de un bajo rendimiento. El estudio realizado pone de manifiesto la necesidad de una **búsqueda de métodos** que pongan mayor énfasis en la **interpretabilidad** de los resultados, así como la conveniencia de la **aproximación temporal**. Los siguientes capítulos profundizarán en los métodos existentes, para focalizar en el **análisis de supervivencia y minería de procesos**. Como veremos, ambos cumplen las condiciones de interpretabilidad y foco temporal.

De cara a abordar el problema, se han plantado **cinco preguntas de investigación** que serán abordadas en la presente tesis, realizada por compendio de publicaciones. Un total de **ocho artículos** permiten responder a dichas preguntas. Se ha expuesto también la relación entre las preguntas y los artículos, que justifican la **estructura de contenidos de este trabajo**: exploración de técnicas, visión del análisis de supervivencia para atacar el abandono, y análisis de procesos para contribuir al modelado del proceso de aprendizaje. Se estará así en condiciones de concluir con un capítulo final que sintetiza los resultados obtenidos y las principales contribuciones.

Capítulo 2 Exploración de técnicas

"El único conocimiento verdadero es saber que no sabes nada."

Sócrates (470 a.C.- 399 a.C.)

Artículos que se presentan:

Martínez, J. A., Campuzano, J., Sancho-Vinuesa, T., & Valderrama, E. (2019). Predicting student performance over time. A case study for a blended-learning engineering course. *CEUR Workshop Proceedings*, *2415*.

Martínez-Carrascal, J. A., Márquez Cebrián, D., Sancho-Vinuesa, T., & Valderrama, E. (2020). Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course. *Computer Applications in Engineering Education*, *28*(3), 590–605. https://doi.org/10.1002/cae.22229

2.1. Contextualización de los artículos

Tal y como se ha mostrado en la introducción, son muchas las técnicas utilizadas para la resolución de problemas educativos desde la perspectiva de *learning analytics*, así como las variables que se utilizan en estudios predictivos (Castro et al., 2007; Dutt et al., 2017). Como factor adicional a tener en consideración, diferentes estudios miden la eficacia de las técnicas con métricas diferentes lo que complica la comparativa.

La primera parte de la investigación, que presentamos en este capítulo, ha consistido en abordar mediante técnicas de uso habitual en análisis de factores que condicionan el logro académico. Debido a la dispersión de técnicas y variables que refleja la literatura, planteamos una exploración inicial que permita determinar las vías más adecuadas para profundizar en el análisis de la mejora del rendimiento.

Focalizando en la comprensión de razones que justifican el bajo logro académico, y en el diseño de potenciales intervenciones, se ha buscado determinar en primer lugar hasta qué punto pueden hacerse predicciones en base a los datos existentes en los LMS, y la coherencia de estas predicciones en relación con la literatura existente. Recordemos que una de las razones para focalizar nuestro análisis en entornos *online* es justamente la disponibilidad de datos asociada al uso de estos sistemas. Por otro lado, y considerando que el objetivo final es la posible intervención, se ha realizado una evaluación de cómo mejora la capacidad predictiva de los modelos a medida que se dispone de más datos conforme avanza el curso. Esto permitirá la intervención tan pronto como los resultados de la predicción puedan considerarse fiables.

Dos son los artículos que corresponden a esta primera fase. Ambos exploran técnicas de clasificación. El primero explora cómo evoluciona la predicción en el tiempo (Martínez et al., 2019). El segundo, focaliza en la predicción temprana (Martínez-Carrascal, Márquez Cebrián, et al., 2020).

2.2. Revisión de variables y técnicas

Los dos artículos que se presentan en este capítulo han utilizado técnicas de clasificación con el propósito de determinar diferencias en términos de superación o no superación del curso. En esta primera etapa se ha utilizado un curso *blended*. Este modelo de aprendizaje incluye una componente principal presencial y otra que se apoya en sistemas *online*. Por ello, dispondremos de datos recogidos en un LMS, que permitirán analizar los patrones de seguimiento del curso por parte de los estudiantes. A nivel pedagógico, el equipo docente proporciona pautas claras sobre cuándo y qué tareas deben realizarse en cada momento a lo largo del curso.

Las variables utilizadas en estos dos artículos nos informan sobre el comportamiento del estudiante en relación con las tareas establecidas basándonos en la relevancia de la actividad que muestra la literatura previa. No consideramos solo aspectos cuantitativos de su interacción con el LMS – por ejemplo. número de veces que el estudiante hace *login*, o número de páginas visualizadas –, sino que también se ha buscado la contextualización de variables como parte del proceso de aprendizaje. A modo de ejemplo, desde un punto de vista pedagógico, no es equivalente visualizar un determinado contenido en el LMS cuando lo sugiere el equipo docente que hacerlo a posteriori. Siguiendo esta lógica, las variables con las que trabajarán los modelos introducen conceptos como cuántos contenidos son vistos a tiempo, o cuantos se visualizan con posterioridad al período sugerido.

Estas variables nos permiten obtener información – aun cuando de momento de forma muy primitiva – sobre si el estudiante sigue o no el proceso pautado y al ritmo correcto. Este simple análisis pone de manifiesto la relevancia del tiempo (que será factor clave en el análisis de supervivencia que se muestra en el Capítulo 3). Es también un germen – ciertamente primitivo – de las reflexiones que surgirán cuando se evalúan los caminos de aprendizaje (Capítulo 4), donde se muestra la diferencia entre realizar una actividad en el momento pautado o bien con posterioridad, suponiendo un retraso, y, de hecho, una divergencia del camino pautado.

En este estadio de la investigación, conviene destacar que el número de variables potenciales es muy elevado. Por ello será siempre necesario un proceso de reducción de variables, orientado a utilizar sólo aquellas con mayor capacidad explicativa. Considerar mayor número de variables puede permitir reducir el error de clasificación, pero a la vez, aumenta el riesgo de sobreajuste (*overfit*).

Como punto adicional, los dos estudios publicados dejan de lado de forma intencionada aspectos ligados a las calificaciones académicas obtenidas. Según se ha comentado en el capítulo anterior, las variables ligadas a notas previas (p.e. *GPAs*) suponen un indicador muy relevante. El hecho de obviar su uso se explica por nuestro interés en las variables ligadas a la realización de tareas específicas y al foco en el proceso de aprendizaje – qué se hace y cuándo – más allá de la aptitud del estudiante. Ser capaces de predecir incluso sin hacer uso de notas mostrará la capacidad de predicción del resultado académico a partir de la trayectoria de aprendizaje reflejada en las trazas recogidas en el LMS.

Considerando que los artículos focalizan en la predicción de la superación o no superación del curso, se han considerado las técnicas habituales en la resolución de problemas mediante *learning analytics* que tienen por objeto la clasificación de estudiantes en diferentes grupos (Shahiri et al., 2015; Wolter et al., 2014). La lista inicial se ha filtrado para considerar aquellas que según la literatura presentan mejores resultados en términos de capacidad predictiva.

- Técnicas bayesianas.
- Árboles de decisión, incluyendo las versiones de GBT y *'random forest'*.
- Vecinos cercanos (*k-Nearest Neighbors* – KNN –).
- *Support Vector Machine* (SVM).
- Redes neuronales.

La bondad de las técnicas se compara frecuentemente en términos de error de clasificación. Aun así, Bowers et al. (2013) indican que el análisis no puede limitarse simplemente a considerar este error, ya que no se trata de un parámetro que por sí solo permita la comparación de resultados. Siguiendo esta recomendación, nuestros dos estudios contemplan también los valores de sensibilidad y especificidad, proporcionando una visión gráfica más completa. Como sugiere este artículo, la simple consideración de estos valores proporciona una mejor visión del comportamiento del algoritmo. La figura 2.1 muestra el marco de comparación gráfica, basada en uno de los artículos presentados (Martínez et al., 2019):

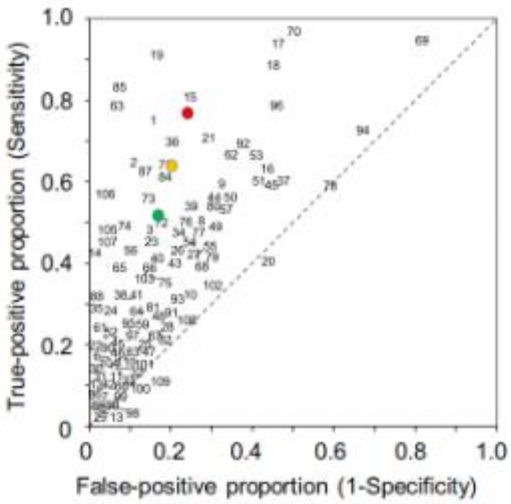


Figura 2.1– Framework propuesto para la comparación de modelos según sensibilidad y especificidad. Imagen de (Martínez et al., 2019)

Bowers et al. (2013) sugieren la comparación de los falsos positivos (valor indicado en el eje de abscisas) con los positivos reales indicados en el eje de ordenadas. Esta comparación sitúa los mejores modelos cerca del punto (0,1). Esta visión gráfica permite detectar qué posibles modelos con bajo ratio de error, pueden estar alejados de ese punto, indicando un comportamiento no óptimo. En particular, la bondad del método se incrementa a medida que la distancia de su representación gráfica en la figura 2.1 se aleja del eje y=x.

Al margen de estas consideraciones, y basados en los estudios de Cohen y Rosset (Cohen, 1960; Rosset, 2004) se han incluido también los valores de AUC (*Area Under Curve*) y el coeficiente *Kappa* como relevantes. El primero de ellos, mide el área bajo la curva operativa característica (ROC por sus siglas en inglés) permitiendo establecer la calidad de un modelo en todos los intervalos de clasificación. Se trata de un valor en el intervalo [0,1] que mide la calidad del clasificador. Valores en torno a 0.5 indicarían una clasificación aleatoria, mientras que valores próximos a 1 indicarían un modelo que clasifica perfectamente.

Respecto al valor Kappa (Cohen, 1960; Landis & Koch, 1977), se trata igualmente de un valor en el intervalo [0,1], pero que en este caso mide hasta qué punto la predicción se ha establecido por puro azar (Kappa=0), o bien está estadísticamente fundada (Kappa=1). Valores por encima de 0.4 son considerados asociados a modelos cuyos resultados son admitidos como no basados en azar.

2.3. El factor tiempo

Como se ha reflejado tanto en el capítulo introductorio como en el apartado anterior, el tiempo es un factor relevante. En este punto de la investigación, esto es así por dos razones. La primera radica en que una potencial intervención debe ser realizada a tiempo (Rienties et al., 2017). Permitir intervenciones más tempranas, puede suponer anticipar el tratamiento del problema, y con ello, un mayor potencial de mejora del logro. El segundo, es que el hecho de considerar el aprendizaje como proceso que se desarrolla en el tiempo, y donde cada acción se realiza en un momento concreto, en condiciones específicas. Ambas consideraciones se han tenido en cuenta en el desarrollo de los trabajos que se presentan en este capítulo.

El primero de ellos (Martínez et al., 2019) focaliza en las diferencias esperables en términos predictivos según el momento en que se realiza la predicción. En particular, se evalúa la capacidad predictiva de los modelos planteados en cuanto a la superación o no del curso. La predicción se realiza en tres momentos a lo largo del mismo. Se evalúa la mejora de la predicción a lo largo del tiempo haciendo uso de los indicadores expuestos en el punto anterior, completándola con una comparativa gráfica con trabajos anteriores basada en la proximidad al punto (0,1) de la Figura 2.1.

Pero la relevancia del tiempo va más allá de la capacidad de intervenir en el momento adecuado. El tiempo es parte relevante del proceso de aprendizaje. No sólo es relevante qué se hace, sino cuando se hace. Por ello, en el segundo de los artículos de este capítulo (Martínez-Carrascal, Márquez Cebrián, et al., 2020), las variables se seleccionan para identificar qué hace el estudiante y cuándo lo hace. Las reflexiones llevadas a cabo durante la realización de dicho trabajo servirán de base para justificar la idoneidad de la minería de procesos (Capítulo 4). Deberá eso sí analizarse cómo esta minería de procesos puede trasladarse a la resolución de problemas de *learning analytics*.

2.4. Principales resultados de las contribuciones del capítulo

Indicábamos en el punto 2.1. que el objetivo de esta primera parte de la investigación, y muy específicamente de los trabajos presentados en este capítulo era la exploración de técnicas de uso habitual en análisis de factores que condicionan el logro académico. Los artículos presentados completan esta fase exploratoria, aportando además líneas de continuidad para la presente investigación.

Se ha conseguido **establecer un entorno de trabajo** que permite el análisis de datos educativos, de forma sistemática. Dejando de lado los aspectos técnicos inherentes al procesado de datos, se ha analizado **qué factores son relevantes, cómo influyen en la superación académica, y hasta qué punto pueden considerarse variables predictivas** en diferentes momentos del curso. Se ha tomado como **marco de comparación** de resultados el expuesto en (Bowers et al., 2013). En términos de capacidad predictiva, el trabajo en (Martínez et al., 2019) muestra que los modelos resultantes para el curso *blended* analizados en dicho artículo son comparables en términos de precisión de la predicción a los publicados por otros autores en diferentes cursos. Validamos así las técnicas utilizadas y comentamos seguidamente cuáles han mostrado mejores resultados en nuestro caso, así como la mejora de la capacidad predictiva a medida que se amplía el horizonte temporal de estudio.

Focalizando en la bondad de ajuste de las técnicas exploradas, la mejor técnica depende del horizonte temporal del estudio. En términos generales, SVM (*Support Vector Machine*) ha dado resultados razonablemente buenos en los diferentes escenarios planteados. Las técnicas bayesianas por su parte se han comportado particularmente bien cuando el escenario temporal es amplio. Se han determinado también las **variables relevantes**: la asistencia a clase y la visualización de contenidos en el momento especificado se muestran como principales variables a considerar. Remitimos al primero de los artículos de este capítulo (Martínez et al., 2019) para profundizar en el detalle de estos hallazgos.

En esta línea, conviene indicar que el estudio realizado en estos artículos apunta ya que – para el caso del curso analizado – trabajar en cada momento del curso los contenidos asociados al mismo ayuda a superar la asignatura (Martínez-Carrascal, Márquez Cebrián, et al., 2020).

Visualizar los contenidos proporcionados por el equipo docente en el momento en que el docente se asocia a mayores tasas de superación de la asignatura respecto a un trabajo tardío – o la ausencia del mismo –. En el caso concreto de este curso –que se imparte con metodología *blended* – no es equivalente ver un contenido en el momento en que se está trabajando, que procrastinar su estudio. Esta idea de la **importancia del momento –** y, de hecho, de la secuencia **–** aparecerá de nuevo, con una aproximación formal más adecuada en el Capítulo 4.

Por otra parte, y aun cuando fuera del foco principal de este capítulo, el modelado ha puesto de manifiesto resultados con impacto docente. Por ejemplo, se han podido visualizar diferencias de comportamiento entre estudiantes repetidores y no repetidores. Este hecho subraya la **importancia de buscar métodos focalizados en la interpretabilidad**, y no focalizar simplemente en el aumento de la capacidad predictiva. Las variables deben contextualizarse. Es decir, debe observarse qué sentido tienen dentro del proceso de aprendizaje y cómo interactúan entre ellas.

En cuanto a la relevancia del horizonte temporal de estudio, la predicción mejora a medida que se amplía la ventana temporal de análisis. En términos estrictamente predictivos, hay un aumento de las ratios de predicción a medida que aumenta el intervalo de tiempo analizado. Este resultado es coherente con la investigación previa. El tiempo juega un papel relevante cuando se considera la predicción del logro académico. Pero más allá de este hecho, constatamos que el tiempo juega también un papel por su importancia en el proceso concreto objeto de estudio. Esta constatación deriva de los hallazgos relativos a la diferencia entre haber cursado la asignatura o no, o del hecho de la diferencia entre trabajar a tiempo o procrastinar. En definitiva, se constata que **el aprendizaje es un proceso, donde el tiempo tiene un rol fundamental**.

Queremos concluir este apartado subrayando que el trabajo realizado ha permitido constatar también que no puede abordarse el análisis de cursos sin tener en cuenta los aspectos pedagógicos intrínsecos, y específicamente los tipos, tiempos y secuencias de actividades planteadas. Esta conclusión puede pasar por alto en el análisis de un curso individual, pero resulta clara cuando el análisis se realiza de forma común a un conjunto de cursos. Este análisis muestra que un determinado factor pueda influir de forma diferente en diferentes cursos. Los resultados en términos de variables específicas, o de tasas de predicción no son directamente trasladables entre cursos, o en todo caso pueden extrapolarse únicamente a cursos de características similares.

Consideramos que este hecho refuerza la diferencia entre el concepto de *learning analytics*, y la minería de datos educativa (*educational data mining*). La imposibilidad de traslación directa de resultados entre cursos revela el aspecto diferencial que supone un análisis en términos de *learning analytics,* aun cuando la aproximación es común en términos de *data mining.* Esto nos hace considerar que, **si se buscan resultados en términos de intervención, los cursos deberán analizarse considerando su diseño instruccional específico**. De otra forma, los potenciales hallazgos adolecerán cuanto menos de falta de ajuste al curso objeto de estudio.

2.5. Siguientes pasos de la investigación

Llegados a este punto, puede decirse que mejorar el logro académico comporta necesariamente mejorar la comprensión del proceso de aprendizaje. En este punto, el tiempo juega un papel particularmente relevante, tanto por el momento en que se realiza la potencial detección de situaciones que puedan anticipar un bajo logro académico, como por su rol inherente a cualquier proceso.

Incluiremos por ello el **tiempo y el proceso como factores clave** en la mejora del logro. Y segmentamos esta mejora en dos aspectos: por un lado, la búsqueda de la reducción del abandono; por otro, la mejora de resultados de estudiantes que no abandonan. Los dos capítulos siguientes abordarán de forma separada cada uno de estos conceptos. El capítulo 3, abordará la reducción de las tasas de abandono mediante análisis de supervivencia, mientras que dejaremos para el Capítulo 4 la visión integral del proceso. Antes de abordarlos, presentamos los dos artículos de exploración metodológica que se han publicado sobre los aspectos cubiertos en este capítulo, y a los que referimos para obtener mayor detalle sobre los aspectos tratados en este capítulo.

Artículo 1: Predicting student performance over time. A case study for a blended-learning engineering course.

Martínez, J.A., Campuzano, J., Sancho-Vinuesa, T. & Valderrama, E. (2019). Predicting student performance over time. A case study for a blended-learning engineering course. CEUR Workshop Proceedings, 2415, 43-55.

Scopus: SJR 2019: 0.177, Category: Computer Science (miscellaneous)

Nota: Las referencias numéricas a figuras y tablas contenidas en el artículo que se presenta a continuación deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

Predicting student performance over time. A case study for a blended-learning engineering course

Juan Antonio Martínez[1] [0000-0002-7696-6050], Joaquim Campuzano[1] [0000-0002-4055-8346], Teresa Sancho-Vinuesa[2] [0000-0002-0642-2912] and Elena Valderrama[1] [0000-0001-7673-2310]

[1] Universitat Autonoma de Barcelona, Edifici D, Campus UAB, 08193 Bellaterra, Spain
[2] Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain
juanan.martinez@uab.cat

Abstract. In recent years, different studies have focused in analyzing whether it is possible to explain and predict performance of students based on information we know about them, and in particular, on that obtained from Learning Management Systems (LMSs). A review of existing literature shows we can still raise no conclusion, and in particular when dealing with face to face (F2F) studies.

In this article, we analyze the performance of a first-year engineering course, offered in a higher education institution (a public university). The course under analysis lasts for 12 weeks and is offered with flipped classroom methodology. Activities that students should follow out of class are scheduled in advance, and communicated to students during the learning period. In addition, there has been a previous effort to align learning activities and learning outcomes.

The goal is to determine if prediction models fed with data gathered during the learning process can provide an accurate estimator of students at risk. This risk evaluation will be done considering as core data those reflecting activity, being of particular relevance, traces stored in LMS as part of the learning process.

Our study demonstrates performance can be estimated based on this data, with increasing accuracy over time. Activity performed by the student is linked to academic result, and this relation is verified even when not taking into account any graded results obtained during the learning process.

Keywords: learning analytics, performance prediction, student modelling.

1 Introduction

General adoption of LMSs has motivated a growing interest for Educational Data Mining and Learning Analytics in general and academic performance prediction in particular. Prediction is one of the most explored areas in both fields, increasing its relative weight over time. Research papers related to prediction were around 30% in 1995-2005 [1], while this number increases to over 40% in more recent studies [2].

This rising interest for prediction is more than justified, due to its potential impact at all educational levels. At the macro level, it can help institutional managers to implement educational policies addressed to reduce failure and increase overall quality. Predicting and understanding the reasons behind prediction results can be a powerful tool to increase global academic performance.

While this macro level is undoubtedly of great interest, we will focus on the teaching impact. We consider prediction a potential lever with different potential applications. If we can provide early prediction, we can redress risk behaviors. This change in behavior can be implemented in a passive way – i.e. just informing students of the potential risk – or in a more active fashion – i.e. implementing specific teaching measures -. If we consider late predictions, they can help to understand causes of failure and redesign pedagogical approaches for forthcoming editions of the same course.

Our study will be carried out in a first course engineering subject in a public on-campus university. Being an on-campus university is relevant, as attending on-campus classes mixes with on-line activities. The subject under analysis has been selected due to its blended-learning methodology, where class attendance mixes with on-line lectures and activities. We aim to estimate probability of success, focusing on behavior of the student regarding the different activities. In particular, considering behavior of the students both in on-campus and on-line activities.

The goal of this research is to evaluate to what extent risk of failure in a flipped face to face (F2F) course can be predicted based on the analysis of the student's behavior. The underlying hypothesis is that student performance (in terms of pass-fail) can be predicted, in good measure, by his/her activity during the course even not taking into account grades gathered in evaluative activities performed during the course. In order to validate the hypothesis we raise two research questions (RQ):

- RQ1: Can student's final performance in terms of "pass-fail" in a course be anticipated by analyzing his/her behavior regarding the fulfillment of the programmed learning activities without taking into account grades obtained in evaluative assessments?
- RQ2: How is this prediction influenced when limiting data to those gathered in the early stages of the course?

Affirmative answer to the first question would suggest activity performance is linked to academic result. This would open a path to better understand the learning process for this particular subject and to suggest potential improvements in pedagogical design. Regarding the second question, early prediction can be useful to redress individual student behavior and reduce overall failure. Answers to both questions would help to improve teaching quality.

2 Theoretical framework

Learning analytics (LA), defined formally by Siemens [3] and Ferguson [4], cover a full set of studies dealing with the extraction of meaningful information from data retrieved in the learning process. Ferguson [4] focuses on "measurement, collection, analysis and reporting of data". Closely related, but with different goals, we find Educational Data Mining (EDM), more commonly accepted definition by Baker and Yacef [2] which tends to focus on techniques.

A core concept in both fields is data, being of particular interest those gathered during the learning process. While the process of data gathering related to the learning process is intrinsic in online universities, most on-campus universities were not born with this idea in mind. In recent years, and with the general adoption of LMS systems, there has been a shift towards data gathering and analysis.

The extraction of information from LMS data is a topic on its own. Agudo-Peregrina, Iglesias-Pradas, Conde-González, and Hernández-García [5] suggest to begin by classifying information around two main axes: interactions based on agent and interactions based on frequency of use. Each of these axes will include a number of specific variables, which depend on the study and the expected output [6].

Conijn, Snijders, Kleingeld, and Matzat [6] compiled pre-existing work and summarized variables considered of potential interest in the literature. The use of variables which are linked to the learning process is common. In particular, we can find number of resources viewed, quizzes started, sessions or total clicks. It is likely to remark that the different studies analyzed provide different impact and influence of variables depending on the course under consideration.

According to [5] there is also no consensus on the influence of a given variable. This fact is also reflected in [6], concluding that in order to get better results "we need to get a better insight into what the LMS data represents".

Both compilations ([5, 6]) show the huge number of different variables that are present in different studies. This is also present in LMS related web sites who focus on information gathered from LMS systems [7]. Whichever the initial variable set is, a selection process will be mandatory, in particular if the number of variables is high in relation to the number of samples.

Before entering the prediction process itself, the nature of the problem must be focused. Failure analysis can be approached as a regression problem (i.e. estimating final graded performance of student) or as a classification problem (i.e. analyzing whether the student will pass or fail). The classification approach is common in the literature, with studies suggesting better performance and potential detection of meaningful patterns [8].

Once variables are selected, and considering we face a classification problem, different studies use different methods for prediction. To have some examples, the range goes from simple decision trees [9], to behavioral clustering [10]. Different research compilations regarding techniques ([11, 12]) show there is no universal method that provides suitable results for all situations. In our case, and considering our study is not focused on the techniques themselves, we will evaluate results with most common methods, without being tied to a particular one, putting the focus on the interpretation of results.

Whichever the technique, evaluating model goodness is the next required step. Due to the classification nature of the problem, area under curve (AUC) is a potential indicator of model goodness. AUC "is a one-number measure of a model's discrimination performance, i.e., the extent to which a model successfully separates the positive and the negative observations" [13]. AUC can also help in cases where we are not dealing with large datasets [14].

Some studies raise concerns about use of AUC as performance indicator [15]. In particular, AUC makes no difference between errors, although when thinking of failure/success classification this can be of potential interest. Depending on the potential application of the prediction, false negative and false positive errors could have different impact, and this information is not contained in AUC.

LA and EDM reseearch does not have unique indicators for evaluating model performance. Due to this fact, comparison of published research works is not straightforward ([16]). Different studies apply different metrics regarding model validation. Bowers, Sprott and Taff ([16]) suggest a framework for comparison. This framework considers global accuracy, but at the same time, includes a graphical view with information related to sensitivity and specificity.

Considering the potential drawback of AUC as a unique indicator, and also the need to compare to previous research, keeping accuracy, sensitivity and specificity besides AUC can help to effectively compare with pre-existing works. AUC can be a general indicator of the overall quality of the classification, while rest of parameters provide additional information and allow to compare with previous research.

This same compilation includes articles with different time scenarios. Best performing models are fed with long-time data (math achievement trajectories from grades 7-12, non-cumulative GPA (Grade Point Average) from grades 9-12, and student engagement trajectories from grades 8-12). All of them include evaluative data as input variables to the model.

The impact of graded activities gathered during the course is common in the literature. At the same time, some studies compiled look for pre-existing variables that could be also of potential interest to performance models. This variables can be external to the learning process, and can include social or economic aspects.

As a final consideration, [16] also concludes that "the predictive utility of many variables is dependent upon course site design and pedagogical goals". It seems clear that while there has been a technical approach to data mining, there has not been such an evolution on seeking the interpretation or generation of relevant information from the data stored in information systems in general and LMS in particular.

3 Methodology

We face a classification task, without being restricted to a particular data mining technique. Models will be fed with activity data. This activity data will also include a time scope in order to evaluate three kind of models according to the time when the data are gathered: early (4 weeks), medium (8 weeks) and late (12 weeks).

3.1 Suitable techniques

While it is not the goal of the paper to discuss about data mining algorithms, we did not want to restrict our study to a particular technique. We selected those present in literature compilation regarding student performance prediction [12]. Selected techniques were naive Bayes, neural networks, decision trees (including both gradient boosted trees – GBT- and random forest –RF-) and Support Vector Machines (SVM). For all of these techniques, we will keep classification error, sensitivity and specificity as parameters to compare with existing literature, and AUC as an additional check.

Regarding models, we define the true positive class as those students likely to fail who actually fail. Those students marked as failing who really pass will be considered False Positives (i.e. Type I errors), while students marked as passing who really fail will be the false negative class (i.e. Type II errors). This approach will permit direct comparison with results in [16].

3.2 Variables

Due to the different and high number of variables present in the literature, and to the fact that they normally include graded activities, we decided to begin from scratch, but keeping in mind lessons learnt from previous compilations ([5, 6]). In particular, we look for meaningful variables linked to the learning process.

We reviewed our course design, and looked for knowledge derived from our teaching experience. We considered three core concepts as fundamental to explain academic results: class attendance, continuous working and flipped behavior. Not all this piece of information was kept as structured data before performing this study.

In particular, we had no information regarding class attendance. Class attendance is not mandatory, and there is no specific control, as students can decide – without academic impact – whether they attend classes or not. The introduction of the flipped classroom methodology made us think about potential non-intrusive techniques to estimate it.

This estimation was performed through the use of a learning engagement tool (Socrative). This tool was introduced as part of the course design to help the detection of areas that need reinforcement. Questions are performed to students during class to evaluate contents that are clear and those that need reinforcement. Questions have no impact in grades. They help instructors to focus on specific areas depending on the answers students provide. Information in the logs allow us to provide an estimation of student attendance to class. We summarize attendance in each of the periods (early, medium and late attendance). It is an estimation – and not an exact value – as the tool is not used in every class.

Continuous working is complex to evaluate and measure. In order to keep simple and at the same time meaningful variables, we opted to keep the volume of information collected in the LMS log file per user and week. We kept one variable for each week of the course that reflects the amount of log lines the LMS. For each of the periods (early, medium, late) we consolidate work in the whole period into a single variable.

We raised concerns regarding activities performed offline. To capture this offline activity, we reinforced the need to use the LMS as part of the pedagogical design. Users can obviously work offline, but video lessons and problem solving require access to the platform. In this way, we can assume users with greater activity levels are those with greater number of log entries.

The above group of variables can reflect continuous work but does not directly link to flipped behavior. The flipped methodology would make advisable to review certain topics before attending class. The list of required activities and due dates is part of the course design. These activities and dates are communicated to students in advance on a per-week basis. So, we included a new set of variables, reflecting for each week the amount of work that was assigned to that week and was effectively performed on time.

The need to get this information requires that all instructors share a common set of activities instructed to students. Each of the activities will have a unique indicator. Once this indicator is located in the log files for a given user, date can be compared to due date for that activity. This approach makes it possible to compute on-time performance of activities for every student, provided that all instructors set the same dates for activity performance.

So far, we have variables reflecting class attendance for each of the periods. We also have a per-week estimation of workload performed based on the log data, and finally the amount of work assigned to each of the weeks performed on time. For this last two datasets we also keep the total work performed in the period.

Due to the high number of variables, a forward selection process will be necessary. This is done to keep the recommended ratio between number of variables and number of samples avoiding potential overfit [17]. This operation will be done for each of the time scopes (early, medium, late) under analysis.

3.3 A word on pre-existing data

Different studies have analyzed pre-existing variables which can condition students' outcomes [16]. We discarded variables without direct link to the learning process. In our case, and after discussion, we kept the grade you get when entering the university, and the fact of being new or repeating student. Variables such as city of residence or family income were not considered due to our focus on activity.

Regarding the grade the student enters the university with, we thought that under the same conditions, students with higher entering grades should be more likely to pass. Regarding the fact of being new or repeating student, our experience shows that repeating students show different behavior than those being enrolled in the subject for the first time.

3.4 Summary

Table 1 summarizes the information we have provided both for methods and variables. Remember the goal will be to classify students based on probability of passing or not for different moments along the course.

Table 1. Summary of classification methods and variables.

Classificiation methods	Variables (common to all methods)
Naive Bayes Neural Networks (NN) Decision trees (DT) – incl. GBT and RF- Support Vector Machines (SVM)	Class attendance (summarized for early, medium and late period) Work performed on a per-week basis, estimated through the LMS log file. Total work in early, medium and late period Work corresponding to the contents covered in class in each specific week (on-time work) Aggregated on-time work for early, medium and late period Number of times the student was enrolled in the subject University access mark

4 Results

Table 2 shows the results for the different methods and time scopes. For clarity, only Random Forest is shown among decision tree techniques, as gradient boosted and simple decision trees provided no better results. Cross-validation has been performed through k-fold cross-validation.

Table 2. Results for different methods and timelines

		Bayes	NN	RF	SVM
Early	Classif error	36.6	37.1	35.1	33.7
(1st block)	(+/-STD)	(+/-12.3)	(+/-3.7)	(+/-5)	(+/-9.3)
	AUC	0.75	0.734	0.724	0.671
	(+/-STD)	(+/-0.105)	(+/-0.11)	(+/-0.09)	(+/-0.13)
	Sensitivity	53.6	92.4	51	74.8
	(+/-STD)	(+/-31.7)	(+/-6.6)	(+/-10.5)	(+/-9.4)
	Specificity	82.6	33.0	79	51.1
	(+/-STD)	(+/-11.1)	(+/-8.5)	(+/-10.9)	(+/-18.9)
Medium	Classif. error	29.9	25.4	28.1	29.5
(2nd block)	(+/-STD)	(+/-3.7)	(+/-6.7)	(+/-2.2)	(+/-11)
	AUC	0.817	0.831	0.8	0.715
	(+/-STD)	(+/-0.081)	(+/-0.05)	(+/-0.08)	(+/-0.1)
	Sensitivity	64.6	66.9	66.7	50.1
	(+/-STD)	(+/-4.9)	(+/-11.7)	(+/-14.4)	(+/-18)
	Specificity	79.4	80.6	77.4	89.8
	(+/-STD)	(+/-11.6)	(+/-7)	(+/-5.1)	(+/-8.5)
Late	Classif. error	24.4	24.7	35.6	29.9
(3rd block)	(+/-STD)	(+/-5.7)	(+/-3.4)	(+/-5.9)	(+/-5.1)
	AUC	0.86	0.827	0.718	0.826
	(+/-STD)	(+/-0.062)	(+/-0.066)	(+/-0.09)	(+/-0.06)
	Sensitivity	77	73	51.9	85.8
	(+/-STD)	(+/-5.3)	(+/-6)	(+/-16.4)	(+/-11.9)
	Specificity	76.4	78.2	78.4	61.2
	(+/-STD)	(+/-12.6)	(+/-3.9)	(+/-11.8)	(+/-8.7)

We have compared results with the compilation in [16], where impact of different variables in published models is shown. For clarity, we have added just our Bayesian models results, as they offer best performance in terms of AUC for early and late activity, and for mid-term is close to maximum. Results are shown in figure 1, marked as "Early prediction", "Mid-term prediction" and "Late prediction". The compilation includes 110 indicators (depicted as numbers in Figure 1) from 36 different prediction works:

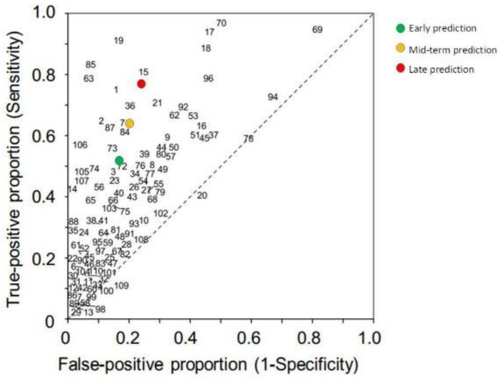


Fig. 1. Comparison of current work and previous research. Adapted from [16]

We also analyzed which were the variables that show relevant in the different time scenarios for the Bayesian method analyzed. Table 3 summarizes the results for the different time scopes:

Table 3. Relevant variables for different periods under analysis

Period under analysis	Selected variables
Early	Attendance to class Work in weeks 1,4 Flipped behavior in weeks 2,3 Access grade to university
Medium	Attendance to class (both during early and medium periods) Work in weeks 1,4,5,7 Overall work in Medium period Flipped behavior in week 3
Late	Attendance to class (all periods) Work in weeks 1,4,7,11 Overall work in Late period Flipped behavior in week 3

5 Discussion

Before answering RQ1 and RQ2, we need to validate to what extent the models outlined in this paper are of potential interest. This question can be answered analyzing Figure 1 and results in Table 2.

Figure 1 shows that final performance of student can be predicted based on the activity data analyzed. We can create models that anticipate success based on this activity data, being those models more accurate the longer the period under consideration. AUC values confirm also models are better the longer the period we analyze.

Prior to comparing with other LA/EDM works, we want to remark our focus has been set on activity. We could obtain potential better models by analyzing partial grades obtained by students, but this could mean losing the focus on the impact of student behavior regarding the subject. This is particularly relevant for us, even more considering we are facing changes in methodology, such as the flipped behavior.

Limiting analysis to early stages provides average results. While we believe a longer time period would be advisable, in particular if the goal is to take actions which can derive costs, performance is similar to other published results. In particular, if we look into the works compiled in [16], we find studies with similar performance, such as [18, 19] – with indicators depicted as 3,72 and 73 in Figure 1–. In the first case, indicators included in the study are socio-economic, while in the second they are related to extra-curricular activities. As a noticeable point, none of them includes grades as predictor variables.

Within our constraints, to get better results it is necessary to broaden the time scope. Doing so – medium and late models – we get results similar to [20, 21] (whose variables are depicted as points 1,15 and 36 in Figure 1). Those studies have also broader time scopes (minimum 1 year) and do not constrain to limit grading data. Best predicting scenarios in [16] include always graded data [22, 23].

Being able to predict results considering whether activities have been performed or not is relevant from a pedagogical point of view and opens potential future lines. While it is not the main goal of this research we also have analyzed results in Table 3 regarding individual variable impact. If we deep into variable details, attending to class, or performing homework makes a difference. Looking this fact from another angle, we can tell students that coming to class and doing what they are instructed to will help them to pass the subject.

Class attendance is present in all cases, independently of time scope. Regarding homework, during the first period, it is relevant your attitude in the first weeks, and just before first partial test. When the period is longer, it becomes more relevant the amount of work performed in the whole period. We believe there is even room for improvement with this same dataset trying to look for other pieces of information that can remain unnoticed inside the huge data volume.

We have compared our findings with results in [5] . In that case, authors consider "there is a relation between some type of interactions and academic performance in online courses, whereas this relation is non-significant in the case of VLE-supported F2F courses". We believe this relation can be found also in F2F or VLE supported studies as long as they include a pedagogical design that requires the use of VLEs. If this is done so, evidences will be gathered in the LMSs and can show differences in behavior.

Consistently with findings by Bowers et al. [16] for dropout flags, and in particular for early prediction, models still lack accuracy (i.e. classification error is high). The potential impact of this error will depend on the purpose of the prediction. If we use them to just raise early alarms, it would not be critical. If deeper pedagogical actions are taken to redress behaviors that can anticipate failure, there would be non-optimal use of resources

Although the models exposed can set the basics for targeted actions, the design of specific policies or pedagogical interventions to reduce failure should take into account not only global accuracy but the impact of false positives and false negatives. For instance, we could consider small group tutoring actions for students classified as likely to fail. The specific design of the action should be made taking into account the false positive rate – i.e. it will affect students who would potentially pass without the action – and at the same time the false negative rate – i.e. there is a group of students marked as passing who will not potentially pass -. We believe this analysis opens a really interesting future line in the field of pedagogical design.

The longer the period, the higher the values for accuracy and sensitivity. In other words, with longer periods we are more certain about final results regarding true positive class. In our case, that means we are more certain behavior of the user could lead to failure.

While this has been a model for an individual subject, we would like to make a reflection about robustness and portability. Computation of data in Table 2 has been done through cross-folding validation, and shows high values of variance in some cases (i.e. sensitivity in early models). To solve this issue it would be advisable to have a higher number of samples (i.e. more students to analyze).

Regarding portability, the process we followed to extract information shows there is a great dependence on course design. Portability of the resulting model itself is not straightforward, but we believe the methodological approach is. The analysis followed can help to obtain models for any flipped classroom course. A pedagogical design that helps gathering evidences from LMS, combined with meaningful variables and, for F2F universities, class attendance should generate models that anticipate potential success based on pure activity data. We believe it will be difficult to generate portable results among different subjects even in same university unless they share a common course design.

Although it was not the goal to establish a comparison among algorithms, Bayesian models have shown good performance related to computational cost. SVM performs also well, but at a higher computational cost. Decision-tree family algorithms can be of interest but would need a higher number on samples to avoid deviations. Finally, deep learning techniques have not provided considerable gain and have higher computation requirements.

To sum up, and going back to the research questions introduced in this paper, final performance of individual students can be anticipated considering only activity data. Relevant aspects for success, considering the course design in our study, include class attendance and different aspects related to homework. Regarding the influence of time, early periods lack accuracy, and would not be optimal if the goal is to set-up actions which involve high costs. As we consider longer periods – medium and late – the models get better. Results for these medium and late models can be of potential help both to redress behavior – in the case of the medium prediction – or – once course is finished – to analyze results and improve course design for future course sessions.

6 Open lines

This paper wants to set the basics for defining specific actions to reduce failure in engineering studies in higher education. Lines of activity include:

- Improve models, in particular in early periods, potentially including new data.
- Deepen into the meaning of the variables selected as more relevant.
- Apply same methodology to other subjects in order to validate and compare results.
- Define actions to reduce failure based on early and medium prediction analysis and to improve pedagogical design based on early, medium and late predictions.

Authors are open to collaboration in previous lines – or to carry out similar research in other environments –. For those interested in carrying out similar research on their own, data processing was done through Python scripts, using Scikit-learn libraries (https://scikit-learn.org/) for modelling algorithms. In particular, sklearn.naive_bayes, sklearn.neural_network, sklearn.tree, sklearn.ensemble (for GBT) and sklearn.svm implementations were relevant among those used [24]. The method does not rely in any particular LMS, but our study was carried out on Moodle platform (https://moodle.org/). Final models were also tested on RapidMiner software (https://rapidminer.com/) to validate results.

References

1. C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications,* vol. 33, no. 1, pp. 135-146, 7 2007.
2. R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions".
3. G. Siemens, "What are Learning Analytics?," 2010. [Online].
4. R. Ferguson, "Learning analytics: drivers, developments and challenges Journal Item Learning analytics: drivers, developments and challenges," *International Journal of Technology Enhanced Learning,* vol. 4, no. 5, pp. 304-317.
5. Á. F. Agudo-Peregrina, S. Iglesias-Pradas, M. Á. Conde-González and Á. Hernández-García, "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning," *Computers in Human Behavior,* vol. 31, pp. 542-550, 2 2014.

6. R. Conijn, C. Snijders, A. Kleingeld and U. Matzat, "Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS," *IEEE Transactions on Learning Technologies,* 2017.
7. "The Indicators Project – Dabbling in analytics," [Online]. Available: https://indicatorsproject.wordpress.com/.
8. P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira and R. Abreu, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance".
9. J. Yoo, S. Yoo, C. Lance, J. Hankins, J. Yoo, S. Yoo, C. Lance and J. Hankins, "Student progress monitoring tool using treeview," in *Proceedings of the 37th SIGCSE technical symposium on Computer science education - SIGCSE '06*, New York, New York, USA, 2006.
10. P. J. H. Y. I.-H. J. Yeonjeong, "Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute," *The Internet and Higher Education,* pp. 1-11, 2016.
11. F. Castro, A. Vellido, À. Nebot and F. Mugica, "Applying Data Mining Techniques to e-Learning Problems".
12. A. M. Shahiri, W. Husain, N. '. Aini and A. Rashid, "ScienceDirect The Third Information Systems International Conference A Review on Predicting Student's Performance using Data Mining Techniques," *Procedia - Procedia Computer Science,* vol. 72, pp. 414-422, 2015.
13. S. Rosset, "Model selection via the AUC.," in *Machine Learning, Proceedings of the 21st International Conference*, 2004.
14. C. Ferri, J. Hernández-Orallo and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognition Letters,* vol. 30, pp. 27-38, 2008.
15. J. M. Lobo, A. Jiménez-Valverde and R. Real, "AUC: a misleading measure of the performance of predictive distribution models," *Global Ecology and Biogeography,* vol. 17, no. 2, pp. 145-151, 1 3 2008.
16. A. J. Bowers, R. Sprott and S. Taff, "Do We Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity and Specificity," vol. 96, no. 2, pp. 77-100, 2013.
17. F. E. J. Harrell, K. L. Lee and D. B. Mark, " "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Stat Med,* pp. 361-387, 1996.
18. A. I. ""Mother got tired of taking care of my baby." A study of dropouts," *Austin, Texas: Austin Independent School District.,* 1982.
19. J. L. Mahoney and R. B. Cairns, "Do extracurricular activities protect against early school dropout?," *Developmental psychology,* vol. 33, no. 2, pp. 241-253, 1997.
20. E. Allensworth and J. Easton, "What matters for staying on-track and graduating in chicago public high schools: A close look at course grades, failures, and attendance in the freshman year," 2007.
21. D. Doss, "Ninth grade course enrollment and dropping out," in *Annual Meeting of the American Educational Research Association*, San Francisco, 1986.

22. A. J. Bowers and R. Sprott, "Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis," *Journal of Educational Research,* vol. 105, no. 3, pp. 176-195, 1 4 2012.

23. B. Muthén, "Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data," in *Handbook of quantitative methodology for the social sciences*, D. Kaplan, Ed., Sage Publications, 2004, pp. 345-368.

24. «API Reference — scikit-learn 0.21.2 documentation» [Online]. Available: https://scikit-learn.org/stable/modules/classes.html.

25. L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an "early warning system" for educators: A proof of concept," *Computers & Education,* vol. 54, no. 2, pp. 588-599, 2 2010.

22. A. J. Bowers and R. Sprott, "Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis," *Journal of Educational Research,* vol. 105, no. 3, pp. 176-195, 1 4 2012.
23. B. Muthén, "Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data," in *Handbook of quantitative methodology for the social sciences*, D. Kaplan, Ed., Sage Publications, 2004, pp. 345-368.
24. «API Reference — scikit-learn 0.21.2 documentation» [Online]. Available: https://scikit-learn.org/stable/modules/classes.html.
25. L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an "early warning system" for educators: A proof of concept," *Computers & Education,* vol. 54, no. 2, pp. 588-599, 2 2010.

Artículo 2: Impact of early activity on flipped classroom performance prediction: a case study for a first-year Engineering course.

Martínez-Carrascal, J. A., Márquez Cebrián, D., Sancho-Vinuesa, T., & Valderrama, E. (2020). Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course. Computer Applications in Engineering Education, 28(3), 590-605. https://doi.org/10.1002/cae.22229 ISSN 1061-3773

Scopus: SJR 2020: 0.478 Q2 Computer Science (miscellaneous) Q2 Engineering (miscellaneous) Q2 Education

WoS: JCR 2020: 1.532 Q3 Education, Scientific Disciplines Q3 Engineering, Multidisciplinary Q4 Computer Science, Interdisciplinary Applications

Nota: la versión que se adjunta del artículo corresponde a la última enviada al editor para su publicación. No se incluyen los cambios de maquetación propios de la revista. Para la versión impresa del artículo consultar https://doi.org/10.1002/cae.22229. Adicionalmente, indicar que las referencias numéricas a figuras y tablas contenidas en el mismo deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

1

Impact of early activity on flipped classroom performance prediction: a case-study for a first-year Engineering course

Juan Antonio Martínez-Carrascal [1,3,*], Dolors Márquez Cebrián[2], Teresa Sancho-Vinuesa [1] and Elena Valderrama[3]

1 Open University of Catalunya, 08018 Barcelona, Spain; jmartinezcarra@uoc.edu; tsancho@uoc.edu;
2 Autonomous University of Barcelona, Department of Economics and Economic History, 08193 Bellaterra, Barcelona, Spain ; MariaDolores.Marquez@uab.cat;
3 Autonomous University of Barcelona, Engineering School, 08193 Bellaterra, Barcelona, Spain ; juanan.martinez@uab.cat; elena.valderrama@uab.cat;
* Correspondence: jmartinezcarra@uoc.edu;

***Abstract*— This article addresses the problem of performance prediction in a flipped learning course on digital systems. The course is a part of a first-year Engineering degree offered at a public university. The extent to which failure at the end of the course can be anticipated is analyzed by examining the students' behavior related to academic activities. This prediction will be made by considering whether students perform different activities as instructed. The timeframe is also limited to the early stages of the course. Although anticipating results at early stages on the basis of behavioral aspects makes it possible to redress the behavior of at-risk students, at the same time, it also introduces limitations for potential models. As there are no assessments during the early period, no course activity grades will be available for mod- eling purposes. In this study, different data mining techniques will be used, which will offer better results. Results will be analyzed in terms of the performance of at-risk students and compared with previous research studies. Models will also provide information about factors of relevance for the academic success of the course..**

***Index Terms*—. blended learning, failure prediction, flip classroom, learning analytics, student modeling.**

I. INTRODUCTION

Performance prediction is a recurrent topic in learning analytics (LA) and educational data mining (EDM) literature. Different studies have addressed the problem of anticipating academic results by analyzing either information we have about students, traces gathered during the learning process, or both [1, 2, 3]. These studies show that the best results are generally obtained when using graded data and long-term periods. In these scenarios, results can help to explain why students succeed or fail, but they rarely help to redress behavior.

The rising interest in LA and EDM, and specifically in prediction, is happening at the same time a shift in education policies is taking place. New policies are being designed with student workload in mind, in particular in Europe [4] and have led to an evolution in classical pedagogical designs. Traditional presential lectures change their role. Pre-recorded lessons cover most theoretical concepts. Students review these lessons before attending classes, often performing other activities -such as specific problem sets- in advance. Face-to-face classes are targeted to solve doubts generated and focus on harder or more challenging areas [5]. Teachers can anticipate any points that might need special attention through data analysis, hence evolving towards adaptive learning.

For on-campus universities, being able to offer campus benefits, while at the same time providing the flexibility and possibilities of flipped methodologies, can be a distinguishing factor. The use of technology, and in particular of virtual learning environments (VLEs), provides extremely useful information to better understand the learning process: data. Data reflects behavior, and different behavior should provide different results.

In this context, we believe this difference in results can be anticipated by analyzing the activity done during the first stages of a course, even when we have no information about intermediate grades obtained in assessment activities. To verify our hypothesis, we raise two research questions that will be covered in this paper:

RQ1: Can potential success (in pass-fail terms) be anticipated by analyzing student behavior at the early stages of a course?

RQ2: If RQ1 is affirmative, which factors are particularly relevant to success?

The first question (RQ1) directly relates to validation of the hypothesis while the second (RQ2) lays the foundations for future analysis and potential actions to impact performance. If we can find factors that can be influenced, we could provide mechanisms to increase overall student performance by detecting and acting to mitigate risk behavior.

II. THEORETICAL FRAMEWORK

Learning analytics (LA), defined formally by Siemens [6] and Ferguson [7], cover a full set of studies dealing with the extraction of meaningful information from data retrieved in the learning process. Ferguson [7] focuses on "measurement, collection, analysis and reporting of data". Closely related, but with different goals, we find Educational Data Mining (EDM), whose most commonly accepted definition is that by Baker and Yacef [8] which tends to focus on techniques.

A core concept in both fields is data, being of particular interest those gathered during the learning process. While data gathering related to learning processes is intrinsic to online universities, most on-campus universities were not born with this idea in mind. In recent years, and with the general adoption of learning management systems (LMS), there has been a shift towards data gathering and analysis.

Different authors have compiled existing work either as a general state of the art overview [9] or focusing on specific topics, and in particular on prediction [10]. These compilations show research has analyzed different stages of education – from early stages to University studies-, focused on different influencing factors – from pure academic to socio-economic – and even data-mining techniques – from decision trees to neural networks -. These analyses can also be approached from a global view (i.e. dropping out university or not), or looking at specific subjects (i.e. analyzing a set of subjects or even a specific one).

Regardless of the way authors deal with problems, the initial stage of the research involves the extraction of information. LMS systems are a common source of data. Agudo-Peregrina et al. [11] suggest beginning by classifying information around two main axes: interactions based on agents and interactions based on frequency of use. Each of these axes will include a number of specific variables, which depend on the study and the expected output [10].

Compilation in [10] analyzed pre-existing work and summarized variables considered of potential interest in the literature. The use of variables that are linked to the learning process is common. In particular, we can find the number of resources viewed, quizzes started, sessions or total clicks. It is commonly observed that the different studies analyzed provide different impact and influence of variables depending on the course under consideration. The impact of a specific variable is also dependent on the course under analysis. This is also reflected in [11].

Both compilations [10, 11] show the huge number of different variables that are present in different studies. This is also present in LMS related websites, which focus on information gathered from LMS systems [12]. Whichever the initial variable set, a selection process will be mandatory, in particular if the number of variables is high in relation to the number of samples.

Before entering the prediction process itself, the nature of the problem must be focused. Failure analysis can be approached as a regression problem (i.e. estimating student's final grades) or as a classification problem (i.e. analyzing whether students will pass or fail). The classification approach is common in the literature, with studies suggesting better performance and potential detection of meaningful patterns [13].

Regarding classification methods, different compilations [1, 14] show that the range of techniques goes from simple decision trees [15], to behavioral clustering [16] or artificial neural networks [17] , being decision trees, k-nearest neighbors and SVM among the most used [18]. These compilations also show that there is no universal method that performs best for all situations [19]. Besides performance parameters, meaningful interpretation of results is also relevant [20]. Interpretation is more straightforward in the case of decision trees or Bayesian models, while it is not so in some others such as neural networks.

Although accuracy is often the main performance indicator for models, it cannot be analyzed on its own, as it may be influenced by class distribution and does not allow for direct comparison among studies [2]. Authors suggest a framework for comparison that considers global accuracy, but at the same time, includes a graphical view with information related to sensitivity and specificity.

As an additional check, and considering a cross-folding process is involved, models can also be tested against agreement by chance. Different statistical indicators can help to determine interrater reliability, Kappa being one of the most commonly used [21]. Kappa values above 0.4 are advisable to consider models as moderately good [22].

The aforementioned compilations include articles with different time scenarios. Best performing models are fed with long-time data (math achievement trajectories from grades 7-12, non-cumulative GPA (Grade Point Average) from grades 9-12, and student engagement trajectories from grades 8-12). All of these include evaluative data as input variables to the model. Comparable results can be found in compilation in [1]. Papers compiled show accuracy values ranging from 50% to as high as 98%. Looking more closely at these studies, higher accuracy values are obtained when considering assessment grades as part of the input set and not constrained to early stages, and no Kappa check is performed in these cases.

The impact of graded activities gathered during the course is common in the literature. Shahiri et al. [1] selected the most commonly used variables when predicting student performance. Cumulative grade point average (GPA) is the most used source: around 30% of papers compiled in this study use this source of data. In descending order of usage we can find student demographics, individual assessment, extracurricular activities and psychometric factors. As the authors outline, this last category is less used due to its qualitative behavior. Pre-existing variables that can be of potential interest to models are also analyzed in the compilations shown. Pure activity work is not common, as activity data is usually accompanied by graded data.

The compilations above, and in particular [10, 11] show the diversity and large number of variables that are present in different studies. This is also present in LMS-related websites which focus on information gathered from LMS systems [14]. A feature selection phase is a required pre-processing step [23], which becomes even more relevant if the number of variables is high in relation to the number of samples. Correlation-based Feature Selection (CFS) [24] is commonly used as selection algorithm, and in particular when mining academic data [25, 26] . Studies also show that removing irrelevant variables can lead to models with improved performance [27, 28].

Regarding timeframe, Macfadyen and Dawson [29] analyze early prediction, and conclude that "pedagogically meaningful information can be extracted from LMS-generated student tracking data," but there has been no particular evolution in terms of seeking interpretation or generation of relevant information from the data gathered in the LMS. Results are consistent with compilations in [10, 11].

As a final consideration, [2] also concludes that "the predictive utility of many variables is dependent upon course site design and pedagogical goals". It seems clear that while there has been a technical approach to data mining, there has not been such an evolution in terms of seeking interpretation or generation of relevant information from the data stored in information systems in general and LMS in particular. The compiled studies show that more examples are needed to better understand the learning process. In this scenario, we present a study of early prediction based on activity data in a flipped learning environment.

III. Research Path

A. Global scenario and constraints

The course under analysis is a 1st year Engineering course, which teaches the basics of Electronics. The course must be covered by any student enrolled for the degree, and around 400 students – including some who did not pass it in previous years – are evaluated every year.

With the aim of providing a better learning experience that could lead to better academic results, a flipped learning methodology was introduced. We shifted from the classical lecture system to a system where most contents are provided in advance to students, and classes provide case studies, delve deeper into the topic, and/or discuss and solve potential doubts.

The course under analysis is split into three main blocks, each covering 4 weeks (combinational circuits, sequential circuits and block-level design). For the combinational and sequential blocks, lectures are recorded in advance. To enable students to check the skills acquired in the lectures, a self-assessment environment was implemented so that they could not only follow explanations at home but also check their recently acquired skills. There is an assessment after each block, and also a mandatory final exam at the end of the term.

Methodological changes are not a shift towards a 'full online' course design. Students must also attend on-campus classes. These classes have two goals: (1) check whether students have really gathered key concepts, detecting those that are unclear, and (2) provide personalized explanations and problems based on these results. At the end of each class, students are advised to prepare activities for the next class (either by watching lectures or performing additional activities).

Online activities should provide a better learning experience. At the same time, we gather data from student interactions. As stated in the introduction, we wanted to limit data analysis to the early period. Taking longer periods could potentially lead to better predictions, but RQ1 focuses on the early stages as its timeframe. After discussion among the research team, we decided to define early detection as "detection before the first exam took place". That means analyzing data gathered during the first four weeks of the course, and implies not having any graded data for the course.

RQ2 must be taken into account throughout the whole process. The inclusion of relevant information, and its mapping into variables, must be done while bearing in mind its pedagogical meaning. Otherwise, the model could potentially predict, but no pedagogical interpretation of results would be possible.

B. Data sources

1) Initial considerations

Due to the flipped methodology and the blended learning approach, a great deal of activity data can be extracted from the LMS. This will be the main data source, but we will also need additional information in order to fully reflect students' activity, and in particular that not reflected in virtual environments.

There is an obvious difference between data related to online and data related to on-campus activities. The former is normally structured, and is derived from structured sources, the LMS being the most relevant. For on-campus activity, data gathering is normally neither structured nor automated. Needless to say, potential behavior modeling is more complex.

We decided to simplify, and only try to determine the most critical behavioral factor for us: attending classes. We introduced a classroom engagement tool to the pedagogical design, which helps us to know the extent to which topics covered out of class have been understood. We use this tool in class to raise questions that reflect comprehension of critical aspects. There is no impact on grades derived from tool usage, but we gain valuable information about who is attending classes.

In order to be consistent with the literature and previous research, we introduced a third group of data: pre-existing data. This data should help to differentiate cases where the same activity does not produce the same results. Previous research suggests the influence of different aspects, such as family finance, parental level of education or even geographical constraints [30].

We analyze each group of data in detail.

2) Data related to online activities

The path from data to meaningful information regarding online activity begins by transforming VLE logs into student-related activity. Our approach will be to summarize the data contained in around 500,000 log lines associating the information they contain to the different users. Although there could be other approaches [31] – i.e. summarize by data or by item – this one seems the most suitable for this case, as the final goal will be to classify students.

We opted for coarse-grain log processing, as we want to have a global view of the learning process, rather than having a detailed view of a particular item or activity in the course. This is actually considered the normal methodology in data mining processes when dealing with logs [32], and more specifically in Educational Data Mining [31] where different levels of granularity always coexist. In addition, the fine-grain approach could lead to overfitting in the model, as the number of variables could be too high in relation to the number of samples.

An online activity data model should reflect *who* performed *what kind* of activity and *when*. Dates are important, as different attitude towards the learning process can be inferred depending on them. In flipped designs, certain activities are instructed to be performed before attending specific classes. From a pedagogical point of view, covering contents before due date, after, or not covering at all are different approaches. Therefore, item access dates must be compared to the due date for items assigned as homework.

Two additional parameters are kept for each user. The number of different login days during the period under analysis and overall volume of activity reflected in the log file. This activity does not relate to a particular item or kind of item, but is just a global view of the user's activity as reflected by the log. As the level of log verbosity is the same for all users, the amount of information the log keeps about each of them can be considered an indicator of the volume of work in the LMS.

We should note that detailed information about each item in particular could be useful for potential research in specific areas, but could also take our focus away from the high-level problem we are considering.

3) Data related to on-campus activities

Keeping and analyzing only the above data would be as much as assuming that user behavior can be entirely modeled based on online activity. We believe that on-campus activity matters – i.e. coming to class helps you to acquire knowledge –, even more in the case under study when we are talking about a flipped learning methodology where class should provide truly valuable knowledge.

We limited information about class attendance to that which is most directly related to the subject under analysis. We considered extensions, but found no structured data source that could provide relevant information for the goals of our research, and relied on the classroom engagement tool as our data source. Regarding usage of the tool, students were clearly informed that the questions it asks serve the sole purpose of enabling students to check their knowledge gathering while also helping the teacher to detect – from a global perspective – areas that need reinforcement. In particular, there is no penalty for wrongly answering comprehension questions about blended lectures.

4) Pre-existing data

Some of the studies analyzed consider pre-existing variables that can condition outcomes [33] or even drop-out from studies ([34, 30] among others). After considering those references, we kept four main factors as those potentially explaining differences in performance, which could not be shown in pure activity:

- Form of access to university: under Spanish regulations, University can be accessed in different ways. The most common form is through specifically university-oriented studies. Access can also be through studies with less theoretical background and more focused on the labor market (vocational courses). The profiles of the two kinds of students can be potentially different. There are also other ways to get into university, such as specific tests for people over 25.
- University entrance grade. This grade is kept with the idea that under the same conditions, students with higher entrance grades should be more likely to pass.
- Priority indicated by the student when entering university. Again, in the Spanish system, students are assigned based on previous grades. Each student fills in a list of preferences. Assignment is based on availability, grades and preferences. If a student does not meet the minimum grade to take their first choice, they will end up studying something that was not their first option. Again students who take Computer Science as their first choice at our University would be expected to be more likely to pass.
- Number of times the user has been enrolled for the subject. Being enrolled more than once means not having passed the subject

5

in an earlier academic year.

The selection process leads to other potentially interesting variables that were discarded either due to lack of pedagogical interest, being to difficult to gather or to avoid excessively increasing the number of variables.

Also with the idea of focusing on activity, we discarded personal data such as gender, estimated family income or personal situation. We believe their potential influence on failure (reflected for instance in [2, 34]), will be reflected in the logged activity, as activity is finally conditioned – among other things – by personal circumstances.

5) Course structure: the glue

So far, we have structured information on student access to contents, and time – or times - when this happens. We still lack a view of whether the student is behaving as flipped. We need to know whether the student is doing activities when he/she should. As a collateral effect, we need to know what kind of activities he/she is doing.

One of our concerns is whether student behavior fits this flipping. That means answering two questions: what kinds of activities are being performed (watch lectures, do quizzes, read materials…) and when.

For example, in a certain class we encourage students to watch videos or complete quizzes in the LMS before a due date. With the data collected (and structured) thus far it is not possible to know which content is a lecture or a quiz (all of them are ultimately IDs without additional information). Even for the same type of content, some content can be assigned to the student as a recommended activity and another might not.

In order to be able to extract this information, the course needs to be clearly prepared. We agreed in advance to a recommended schedule of 'flipped activities'. This information is structured according to Table 1

TABLE 1: SAMPLE ROWS FOR COURSE STRUCTURE

Class #	Id of item	Action Expected	Due date
1	W1-VID5	WATCH	Feb'15
1	W1-VID6	WATCH	Feb'15
1	W1-QUI4	ANSWER	Feb'15
(…)			

Each row in Table 1 contains information about a suggested activity for a specific class. For instance, in class 1 (Class # column), we tell the student to watch a lecture, which we identify as W1-VID5 (Id of Item column). He/she should also watch W1-VID6, and do two questionnaires. Although they are similar, we distinguish between what we call just 'quizzes' (designed to get the user to check their knowledge, but which do not affect the final grade) and 'evaluative assessments', which have an impact on the continuous evaluation of the student. Expected student behavior is clarified in the column called 'Action expected' – which can be WATCH for lectures, ANSWER for non-evaluative items and EVAL for items which have an impact on evaluation - Due dates for each activity are shown in the last column in Table 1.

C. Data modeling techniques

The world of data modeling is full of different data mining techniques [1, 14], each of them suitable for different situations. We began by selecting those more present in general data mining literature [18], and focusing after that on educational data [13]. We also added not so common techniques such as artificial neural networks [17] or fuzzy models [35]. In order to narrow the list, we considered expected performance of algorithms according to literature and potential interpretation of results.

We finally kept Decision Trees (DT), Support Vector Classification (SVC) and K-Nearest Neighbors (KNN). All of them are among those most used and can provide information for interpretation of results. It is important to note that from a data-mining perspective and limiting to educational problems, [13] compares algorithms and concludes there is "no statistical evidence that the differences in performance between the algorithms are significant". The final common agreement is that only testing for the particular case can tell which algorithm performs better for a given example.

With this rationale in mind, we have not limited our study to a particular algorithm, but we have checked different options considered in previous research. For this reason, and whichever results we get, we will need a comparison framework. Comparison will be made by including our results among those compiled in [2]. That means keeping sensitivity, specificity and error. We also evaluated Kappa, keeping models that present values above 0.3 in initial tests.

D. Preliminary tests

Before performing any modeling techniques, we analyzed the extent to which results are influenced by the variables selected. We carried out statistical tests (t-tests) to determine whether both groups – passing and failing students – showed different values for the variables under analysis. Average values for both groups were computed and the results are shown in Table 2:

TABLE 2: MEAN AND STANDARD DEVIATION VALUES FOR THE DIFFERENT VARIABLES BASED ON GROUP (F: FAIL, P: PASS)

	n		*quiz_ontime*	*quiz_total*	*lectures_ontime*	*lectures_total*	*log_lines*	*login_days*	*attendance*	*access_grade*	*preference*
F	138	(Mean)	1.08	2.23	3.67	8.30	246.9	8.65	53.23	7.18	1.23
		(STD)	0.77	0.90	3.68	5.73	188.1	5.40	35.53	0.92	0.57
P	199	(Mean)	1.53	2.74	7.70	12.73	337.3	12.41	73.83	7.69	1.25
		(STD)	0.83	0.58	5.43	4.88	182.2	4.64	30.07	1.31	0.76
		t(335)	-5.04	-6.32	-7.59	-7.63	-4.42	-6.78	-5.74	-3.95	-0.26
		Signif. level†	***	***	***	***	***	***	***	***	-

† *** : significant level below 0.01 – : not significant

Table 2 shows that the parameters work as expected: users who pass do (in mean values) more activities on time- either exams or video –. This is also true for general log activity and login days. Attendance results are also higher for the passing group.

Following the results from Table 2, we discarded the preference variable from this analysis as there was no initial difference between both groups. For the rest of the variables, and in order to validate statistical significance, a normal distribution was considered for number of logins (*login_days*) and access grade (*access_grade*). Analysis did not make this assumption possible for the remaining variables. Assuming these distributions, we carried out a t-student test for both *login_days* and *access_grade* and a Mann-Withney U-Test (two sided, null hypothesis shift $\mu_0=0$, confidence interval 95%) for the rest. Analysis provided p values $p< 0.0001$ for all cases. As an additional check, we tested the discarded variable (preference) and results showed a p-value of 0.3185 (Mann-Withney U-Test, so not statistically significant, confirming our hypothesis).

In a similar fashion, the means of university entrance was analyzed. Although this may be thought to make a significant difference, we could not appreciate any significant difference between the pass rate for students coming from high school (58% pass rate) vs other means of access (61%).

To summarize, the activity variables and data chosen can show a difference in performance for both groups (pass or fail) and could be the basis for a modeling process.

E. Summary

Table 3 summarizes selected variables that will feed the different techniques:

TABLE 3: SELECTED VARIABLES RELATED TO ACTIVITY USED TO FEED ALL ALGORITHMS

Scope of data	Concept	Variable name
on-line	Lectures watched on time	lectures_ontime
	Total lectures watched (early period)	lectures_total
	Quizzes done on-time	quiz_ontime
	Total quizzes done any time (early period)	quiz_total
	Number of login days	login_days
	Number of user log lines	log_lines
off-line	% of attendance to class	attendance
Pre-existing	Number of times previously enrolled	num_mat

The data in Table 3 will feed all classification algorithms under analysis (DT, SVC and KNN). Data will be picked only for the period already defined as 'early' (4 weeks, from beginning of the course to first exam). We refer to activities performed 'on-time' to those performed before their due date according to Table 1 and 'any time' to those performed 'any time during the early period' (i.e. not necessarily 'on-time').

IV. RESULTS

A. Initial modeling

As stated above, we did not limit our analysis to a single algorithm, and did not initially limit it to a subset of variables. We split the dataset to make different iterations, and also analyzed performance with different variable sets. Initial tests were performed with just two variables that were selected based on correlation with the final grade. Those two variables were *lectures_total* and *quiz_ontime*.

New variables were added, also based on correlation to the final grade. They were introduced based on a correlation-based feature selection algorithm [36]. Iterations with different data subsets were performed for each combination of algorithm and variable set. Table 4 shows the results of this iteration process with different subsets and different numbers of variables (bearing in mind that we always begin with the most relevant – total lectures – and sequentially add new ones. The results include accuracy and Kappa values for the different iterations.

TABLE 4: ALGORITHM ACCURACY AND KAPPA FOR DIFFERENT ITERATIONS (MAX, MEAN AND STD VALUES)

		Accuracy				Kappa			
METHOD	VARS	MIN	MAX	MEAN	STD	MIN	MAX	MEAN	STD
KNN	5	0,60	0,78	0,69	0,04	0,15	0,55	0,34	0,07
KNN	4	0,59	0,80	0,69	0,04	0,15	0,58	0,34	0,08
SVC	2	0,55	0,77	0,68	0,04	0,04	0,49	0,30	0,08
KNN	3	0,59	0,75	0,67	0,04	0,14	0,47	0,30	0,08
SVC	5	0,54	0,75	0,67	0,04	0,02	0,42	0,27	0,09
DT	3	0,52	0,75	0,66	0,04	0,03	0,49	0,29	0,09
SVC	3	0,58	0,75	0,66	0,04	0,03	0,45	0,26	0,08
SVC	4	0,58	0,77	0,66	0,04	0,13	0,47	0,26	0,08
DT	2	0,56	0,75	0,66	0,04	0,05	0,43	0,26	0,08
DT	4	0,55	0,76	0,66	0,04	0,08	0,48	0,27	0,08
DT	5	0,50	0,75	0,65	0,05	0,03	0,47	0,27	0,09

The results shown in the table above, which is sorted by mean accuracy value KNN and SVC, seem to offer better performance, in particular if we limit models of potential interest to those with Kappa values above 0.3.

Despite their poorer performance, Decision Trees were also analyzed. We checked whether it was possible to gather semantic information (for instance, if class attendance higher than a certain value was biased towards passing). No conclusive information could be extracted. We believe that decision trees could be a potentially good method, but would need a much higher number of students to produce relevant results.

For comparison with existing literature, we also extracted the confusion matrix for the best performing algorithm in these tests (KNN including all variables). Results are shown in Table 5:

TABLE 5: RESULTS FOR KNN MODELING WITH ALL VARIABLES

	True0	True 1	Class precision
Pred. 0	75	42	64.10
Pred.1	63	157	71.36
class recall(%)	54.35	78.89	
Accuracy:	0.688	Kappa:	0.34

With the process followed so far, we have different methods that provide accuracy values slightly under 70%, with Kappa values on the boundary of being considered of potential interest.

B. Improvements

Some of the models in the previous subsection (KNN with 3,4 or 5 variables and SVC with just 2) can be considered of potential interest, keeping in mind both accuracy and Kappa values. Looking to improve the model, we decided to explore whether we could conceptually explain the behavior needed to pass the course.

Recalling the recommendations in [10] we focused on the learning process, trying to gain deeper insight that could lead to meaningful information. In the case of on-campus classes prior to VLE systems, some studies show that class attendance was a relevant factor [37]. This study finds an interesting fact. There is a group of students that can be considered particularly motivated, who in fact attend due to their interest in the materials provided in class. This motivation is translated into behavior and correlated to pedagogical achievements.

We translated this analysis into our blended learning course, whose nature modifies the meaning of classical attendance, as part of the contents have been provided online. In order to include both concepts, we combined class attendance and previous lecture-watching activity. In fact, this variable reflects not only class attendance but also maximization of the potential benefits for the student, as the content of face-to-face classes is based on what was worked on previously at home.

Going one step further in the analysis, if the content itself is a motivation, those students who were previously enrolled for the course lack part of the motivation, as they already have all the materials. This same concept can be viewed from another angle: a repeating student that – despite having all contents – attends classes shows extra motivation. And finally, from the point of view of a potential model, the same behavior for a repeating student and a non-repeating student would have different meaning. For this reason, we decided to perform a group segmentation, looking for different models for repeating and non-repeating students.

It can be argued that transfer of the work from pure face-to-face learning to a blended learning scenario is not straightforward. For this reason, we decided to carry out an additional validation process to check the extent to which these two groups behaved differently considering class attendance and off-line lecture-watching activity. Validation was performed by comparing performance of both groups (repeaters and non-repeaters) based on lecture-watching activity and attendance. For lecture-watching activity, we tested both on-time watching and overall lecture-watching activity. We finally kept the latter, as it may be more directly linked to the process of content compilation suggested in [37].

For purposes of comparison, we plotted lecture-watching activity against final course grade. While this grade is not directly used for modeling purposes, it helps to understand whether watching activity has any influence on grade (which ultimately involves passing or not).

The results of this process are shown in Figures 1 and 2. The plots show the normalized value of total lectures watched vs student's final grade in the exam evaluating theoretical concepts that should be gathered through lectures and explanations in class. Class attendance (*attendance* variable) is plotted on a color scale. Plots are shown separately for first-time students and repeaters.

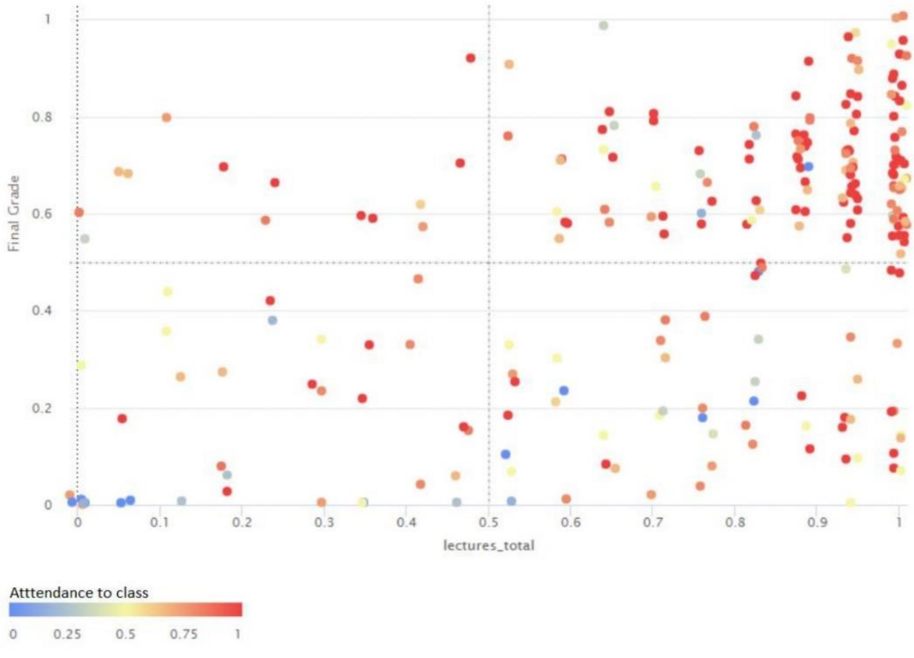


Figure 1: Final grade vs lecture-watching scatter plot for first-time students. Colors indicate class attendance (variables normalized to [0,1] range).

9

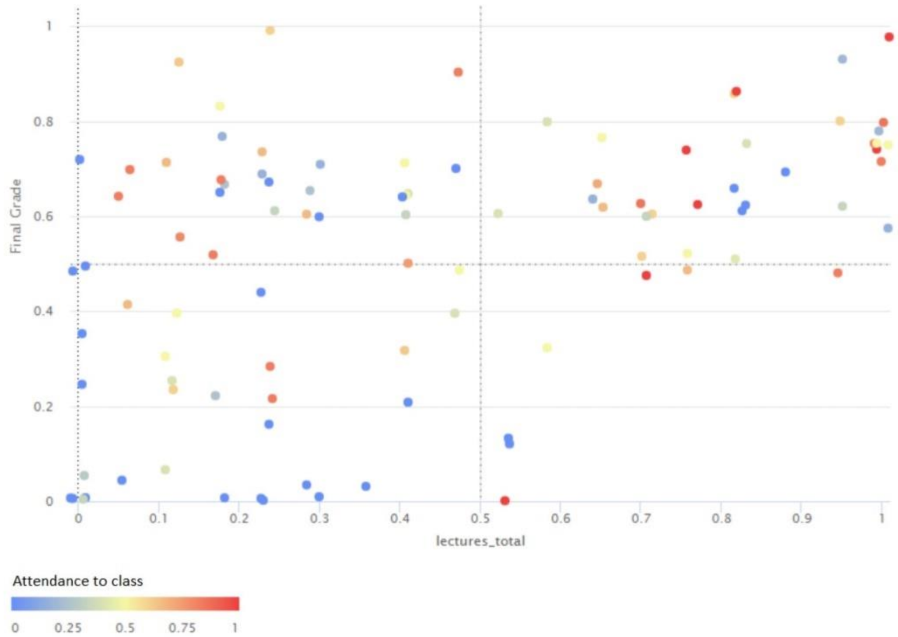


Figure 2: Final grade vs lecture-watching scatter plot for repeating students. Colors indicate class attendance (variables normalized to [0,1] range).

Figures 1 and 2 provide relevant information. First, both groups really behave differently. Second, influence of activity is different for both groups. For first-time students, those with relevant video-watching activity and at the same time relevant attendance (red color in the graph) are very likely to pass (y axis greater than 0 in the normalized values). For repeating students, watching lectures is more relevant, and is a key factor even without high attendance rates.

These findings lead us to implement two changes: (1) we decided to implement separate models for repeaters and non-repeaters and (2) we generated a new variable (named *video_assist*), which includes class attendance and at the same time lecture-watching activity. According to Figures 1 and 2, this variable could be particularly useful for detecting failure in first-time students. The new variable was computed as the product of the normalized values of total videos and assistance.

In order to proceed systematically, we separated our analyses for both groups and performed a forward-stepwise mechanism for variable selection. For first-time students, initial variable selection led to *video_assist, quiz_total, access_grade* and *quiz_ontime*. As in the previous case, CFS was applied to select relevant features for the different techniques under analysis. In terms of performance, SVC provided the best results. The model included just two variables: the one just introduced, which includes lecture-watching and attending class and the university entrance grade. Global accuracy rises to 73.46%, with the Kappa value also increasing to 0.432. The confusion matrix for this technique and variable set is shown in Table 6.

TABLE 6: RESULTS FOR FIRST-TIME STUDENTS, BASED ON COMBINATION OF LECTURE-WATCHING AND CLASS ATTENDANCE AND UNIVERSITY ENTRANCE GRADE

	True0	True 1	Class precision
Pred. 0	57	27	67.86
Pred.1	36	117	76.46
class recall(%)	61.29	81.25	
Accuracy:	0.7346	Kappa:	0.432

For repeating students, the same feature selection process leads to models that will handle *quiz_total, login_days, attendance* and *access_grade* as potentially interesting variables. With those variables, best performance is achieved using the KNN algorithm and keeping all but class attendance. Accuracy rises to 72% and Kappa value to 0.427. The confusion matrix is shown in Table 7.

TABLE 7: RESULTS FOR REPEATING STUDENTS, BASED ON QUIZ_TOTAL, NUMBER OF LOGINS AND UNIVERSITY ENTRANCE GRADE

	True0	True 1	Class precision
Pred. 0	31	14	68.89
Pred.1	14	41	74.55
class recall(%)	68.89	74.55	
Accuracy:	0.7200	Kappa:	0.427

This same combination of variables does not lead to relevant improvement if SVC is used. Also note that the selection mechanism led us to discard some of the variables that are more directly linked to activity, and in particular lecture-watching, which – according to previous rationale- could improve potential models.

Keeping in mind the relevance of watching lectures for repeaters – graphically supported in Figure 2- we decided to keep the lecture-watching activity variables and perform an iterative process with activity-related variables. The test performed showed that lecture-watching (on time), self-assessment quizzes (both total and on-time) and the number of log lines in the LMS, SVC increased performance to 74% providing a Kappa value of 0.472. The confusion matrix is shown in Table 8:

TABLE 8: RESULTS FOR REPEATERS, BASED ON ON-TIME LECTURE-WATCHING, EXAMS (ON TIME AND TOTAL), AND LMS ACTIVITY – MEASURED FROM LOG LINES –

	True0	True 1	Class precision
Pred. 0	31	12	72.09
Pred.1	14	43	75.44
class recall(%)	68.89	78.18	
Accuracy:	0.7400	Kappa:	0.472

As an additional proof of the difference in behavior for both groups, we tested the model generated for first-time students against repeaters. The results are conclusive: accuracy drops to 59% and Kappa value to 0.157, indicating the model for first-time students is not suitable for repeaters.

C. Comparison to previous work

In order to be able to draw conclusions, we compared our results with previous findings in the literature. We plotted data derived from the confusion matrixes obtained in Sections IV.B and IV.C. The plot compares the proportion of false positives (computed as *1- specificity*) in the x-axis to the proportion of true positives (sensitivity) for the y-axis. The best models are those closer to the (0,1) point in the graph. At the same time, methods can be considered better as their distance from the reference line $y=x$ increases. The following models are plotted in Figure 3:

- 5-variable KNN for all students (results summarized in Table 5)
- 2-variable SVC for first-time students (results summarized in Table 6)
- 3 variables KNN for repeaters (results summarized in Table 7)
- 4-variables SVC for repeaters (results summarized in Table 8)

11

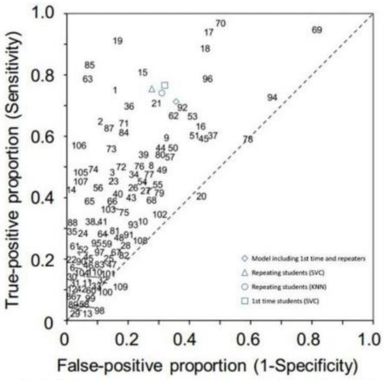


Figure 3: Performance comparison of classification methods in the paper compared to a compilation of research studies (compiled by Bowers et al. [2]) regarding drop-out

V. CONCLUSIONS

Two research questions were introduced in this paper. The first (RQ1) was intended to determine whether success can be anticipated if we focus on student activity in the early stages. A short answer to this question would be affirmative with nuances. While the first model analyzed in table 5 provides neither objectively high accuracy values nor a desirable Kappa value, the results get better for both indicators if we segregate models for first-time students and repeating students. Accuracy values rise to over 70% (72%-74%) and Kappa values are also above 0.4, indicating that these models may be of potential interest. This is reinforced if we compare the results to the existing literature summarized in Figure 3.

From a data mining perspective, pure interpretation of the Kappa value indicates that the model based on activity does not provide results by chance. In other words, and considering the model is fed with activity data, this activity data is effectively translated into academic results. We can therefore conclude that work on the subject is translated into academic performance. This is true even considering we have only analyzed the early stages of the course.

At the same time, Table 3 shows only small difference in best-performant scenarios for the different algorithms compared. These results are consistent with [13]. Results depicted in Figure 3, show that accuracy values are in the range of previous research, even considering that ours focuses on early prediction and does not include graded activities. Most comparable indicators in [2] would be those depicted as 92 – slightly under-performing with respect to the model including first-timers and repeaters – and 15,21 – with slightly higher performance in specific models for repeating and non-repeating -. It is noticeable that indicators depicted as 92 correspond to social situation flags, while both 15 and 21 correspond to graded indicators (consistently showing better performance). We should also note that the best predicting scenarios in [2] always include graded data.

We can also compare our findings with considerations in [11] related to online and face-to-face (F2F) courses. In that case, the authors consider "there is a relation between some type of interactions and academic performance in online courses, whereas this relation is non-significant in the case of VLE-supported F2F courses". Our study indicates this link is also present in our flipped-classroom approach. We conclude that this relation can also be found in F2F or VLE supported studies as long as they include a pedagogical design that requires the use of VLEs. If this is done, different behaviors can be detected through traces in the LMSs and this is translated into academic results.

Our 'affirmative with nuances' answer is consistent with findings of Bowers et al. [2] for dropout flags, and in particular for early prediction: models still lack accuracy (i.e. classification error is high). The potential impact of this error will depend on the purpose of the prediction. If we use them merely to raise early alarms, it would not be critical. If more intense pedagogical actions are taken to redress behaviors that can anticipate failure, there would be non-optimal use of resources. For instance, if we use early alarms to generate small group tutoring actions, we would be offering additional support to students that could potentially pass without it, while not offering it to students who would need it.

Summarizing the results for RQ1, prediction is possible, and even above average in relation to other indicators compiled in [2]. Initial tests provide performances around 68.7% with Kappa values slightly above 0.3, which can be considered poor. Proper selection of variables, group segmentation and new variable generation provide more solid estimators, with Kappa values above 0.4 and accuracy also increases in the 72-74% range.

But the different techniques should not be just analyzed in terms of performance. Romero et al. [20] remark the importance of analyzing pedagogical implications. In our case, these are addressed in RQ2. The different models show that there are general determining factors (lecture-watching is always present) that prove relevant for success. At the same time, relevance of individual indicators can depend on students' situations regarding the course. Once RQ1 shows that prediction is possible and results are not obtained by chance, the results in section IV suggest that the answer depends on whether we are considering a first-time student or a repeater. Repeating a course conditions behavior, and impact of variables is different depending on whether a student is a repeater or not. For first-time students, simultaneously attending classes and doing homework is shown to be relevant. Class attendance is not so critical for repeating students.

It is also noticeable that the models improve if we consider repeaters and non-repeaters separately. In other words, they either behave differently or certain activities are more relevant for each of the groups. This information is not only present in the model itself, but also in Figure 2. Whichever the group, video-watching is a relevant activity, but for first-time students, coming to class makes a difference. We keep this result for a while, as it will contribute to the discussion on RQ2.

First-time students are more likely to pass if they really follow the flipped model – i.e. preparing lectures and attending classes. The second relevant factor for this segment is a student's university entrance grade. Once a student becomes a repeater, entrance grade is not so relevant. The combination of factors, and in particular reviewing lectures and doing homework, is shown to be crucial for passing the subject. Repeaters show lower rates of attendance in comparison to first-time students.

Two interesting factors draw our attention regarding these results. First, they are consistent with [10]: to get better results, it has been necessary to delve deeper into the learning process. Second, and regarding meaning of variables, the results agree with the findings in [10, 11]: influence of variables depends on course under study, and taking that one step further, in our case, it even depends on the type of students we are considering. As shown, a good model for first-time students is useless for repeaters, even when considering the same course. Our results expand on the results in [11], as influence of a given variable does not only depend on the course under study, but also on the situation of the student regarding the subject.

Besides models, and from a pedagogical point of view, we find that engagement – in any form, either class attendance or quiz completion – is a critical factor for success. In this sense, instructors should look for pedagogical designs and methods oriented to motivate student commitment. This can potentially be done in different matters. In the field of engineering we can find for instance the creation of multidisciplinary projects [38] or the use of gamification techniques [39]. These motivational designs are not exclusive to the engineering field. Similar examples can be found for other disciplines ([40, 41, 42] just to cite some examples). These results are consistent with previous pedagogical research, indicating the direct link between motivation and academic result [43]

While the goal was not to compare algorithms, SVC provides good results for both kinds of student, while – in our analysis – KNN did not perform particularly well for 1st time students. This is consistent with the findings in [14]. We can verify that methods that split students between repeaters and non-repeaters work best in terms of accuracy. Regarding segmented models, the one for repeating students is slightly better than that for first-timers. Regarding algorithms, SVC provides good results for both kinds of student, while – in our analysis – KNN did not perform particularly well for 1st time students.

As a final conclusion, activity matters: we can recommend that first-time students watch lectures and attend classes and that repeaters watch lectures and do homework. We believe our results extend on the previous work by Romer [37], which showed the relevance of class attendance for on-campus classes. According to our results, this also holds for flipped-learning environments. For an on-campus university, this is of particular importance, as attendance of classes is shown to be relevant even when contents are given online.

VI. FUTURE WORK

This paper lays the foundations for defining targeted actions to reduce failure in engineering studies. Lines of activity include three axes.

First, gathering new data that will help to improve the results of the model. This new data include social data, trying to detect whether relations also influence results, or results for other subjects at the University.

Secondly, expanding the study to other subjects: while this should be easy in theory, previous work to blend the subject is required. We do believe the results could be promising, but at the same time much effort is required.

Finally, implementing an early-warning system based on the work exposed. As Pistilli and Arnold [44] remark "often students do not understand how well they are performing in a class until it is too late for those performing poorly to change their trajectory". An early-warning predictor can help to make students aware of their performance. If this warning is based on activity, it can help to redress it, ultimately helping to reduce failure.

VII. A WORD ON REPRODUCIBILITY

For those who might be interested in carrying out a similar model, this experiment was conducted on the basis of data gathered from a course offered by the UAB and publicly available at Coursera, part of which is used as blended content for our Campus students. The data classified as LMS data can be obtained directly from the platform. The other datasets were obtained from internal academic management systems.

All this initial data has been treated with Python scripts, using Scikit-learn libraries [45]. These same libraries have provided the basics for modeling algorithms. Final models were also tested on RapidMiner software [46] to validate the results.

13

REFERENCES

[1] A. M. Shahiri, W. Husain, N. '. Aini and A. Rashid, "ScienceDirect The Third Information Systems International Conference A Review on Predicting Student's Performance using Data Mining Techniques," *Procedia - Procedia Computer Science,* vol. 72, pp. 414-422, 2015.

[2] A. J. Bowers, R. Sprott and S. Taff, "Do We Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity and Specificity," vol. 96, no. 2, pp. 77-100, 2013.

[3] C. Romero, P. G. Espejo, A. Zafra, J. R. Romero and S. Ventura, "Web usage mining for predicting final marks of students that use Moodle courses," *Computer Applications in Engineering Education,* vol. 21, no. 1, pp. 135-146, 1 3 2013.

[4] S. Rodrigues, "Overview on Student-Centred Learning in Higher Education in Europe".

[5] A. Taylor, "Flipping Great or Flipping Useless? A Review of the Flipped Classroom Experiment at Coventry University London Campus," *Journal of Pedagogic Development,* vol. 5, no. 3, 2015.

[6] G. Siemens, "What are Learning Analytics?," 2010. [Online].

[7] R. Ferguson, "Learning analytics: drivers, developments and challenges Journal Item Learning analytics: drivers, developments and challenges," *International Journal of Technology Enhanced Learning,* vol. 4, no. 5, pp. 304-317.

[8] R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions".

[9] C. Romero and S. Ventura, *Educational data mining: A review of the state of the art,* 2010.

[10] R. Conijn, C. Snijders, A. Kleingeld and U. Matzat, "Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS," *IEEE Transactions on Learning Technologies,* 2017.

[11] Á. F. Agudo-Peregrina, S. Iglesias-Pradas, M. Á. Conde-González and Á. Hernández-García, "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning," *Computers in Human Behavior,* vol. 31, pp. 542-550, 2 2014.

[12] "The Indicators Project – Dabbling in analytics," [Online]. Available: https://indicatorsproject.wordpress.com/.

[13] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira and R. Abreu, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance".

[14] F. Castro, A. Vellido, À. Nebot and F. Mugica, "Applying Data Mining Techniques to e-Learning Problems".

[15] J. Yoo, S. Yoo, C. Lance, J. Hankins, J. Yoo, S. Yoo, C. Lance and J. Hankins, "Student progress monitoring tool using treeview," in *Proceedings of the 37th SIGCSE technical symposium on Computer science education - SIGCSE '06,* New York, New York, USA, 2006.

[16] P. J. H. Y. I.-H. J. Yeonjeong, "Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute," *The Internet and Higher Education,* pp. 1-11, 2016.

[17] A. O. Osofisan, O. O. Adeyemo and S. T. Oluwasusi, *Empirical Study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database,* 2014.

[18] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems,* vol. 14, no. 1, pp. 1-37, 4 1 2008.

[19] S. B. Kotsiantis, *Supervised machine learning: A review of classification techniques,* 2007.

[20] C. Romero, S. Ventura, P. G. Espejo and C. Hervás, "Data mining algorithms to classify students," in *Educational Data Mining 2008 - 1st International Conference on Educational Data Mining, Proceedings*, 2008.

[21] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement,* vol. 20, no. 1, pp. 37-46, 1960.

[22] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data.," *Biometrics,* vol. 33, no. 1, pp. 159-74, 3 1977.

[23] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence,* vol. 97, no. 1-2, pp. 245-271, 1 12 1997.

[24] M. A. Hall, "Correlation-based Feature Selection for Machine Learning," 1999.

[25] M. Doshi and S. K. Chaturvedi, "Correlation based feature selection (CFS) technique to predict student performance," *International Journal of Computer Networks & Communications (IJCNC),* vol. 6, no. 3, 2014.

[26] F. Zennifa, S. Ageno, S. Hatano and K. Iramina, “Hybrid system for engagement recognition during cognitive tasks using a CFS + KNN algorithm,” *Sensors (Switzerland),* vol. 18, no. 11, 30 10 2018.

[27] A. E. Raftery and N. Dean, “Variable selection for model-based clustering,” *Journal of the American Statistical Association,* vol. 101, no. 473, pp. 168-178, 3 2006.

[28] M. Zaffar, M. A. Hashmani, K. S. Savita, S. Sajjad and H. Rizvi, “A Study of Feature Selection Algorithms for Predicting Students Academic Performance,” 2018.

[29] L. P. Macfadyen and S. Dawson, “Mining LMS data to develop an “early warning system” for educators: A proof of concept,” *Computers & Education,* vol. 54, no. 2, pp. 588-599, 2 2010.

[30] R. Zotti, “Should I stay or should I go? Dropping out from university: An empirical analysis of students' performances,” in *Youth and the Crisis: Unemployment, Education and Health in Europe*, 2015.

[31] C. Romero, J. R. Romero and S. Ventura, *A survey on pre-processing educational data,* 2014.

[32] J. D. Gobert, M. Sao Pedro, J. Raziuddin and R. S. Baker, “From Log Files to Assessment Metrics: Measuring Students' Science Inquiry Skills Using Educational Data Mining,” *Journal of the Learning Sciences,* 2013.

[33] Z. A. Pardos, R. S. J. Baker, M. O. C. Z. S. Pedro, S. M. Gowda and S. M. Gowda, “Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes,” in *Learning Analytics and Knowledge*, 2013.

[34] S. C. Wolter, A. Diem and D. Messer, “Drop-outs from Swiss Universities: An empirical analysis of data on all students between 1975 and 2008,” *European Journal of Education,* 2014.

[35] A. Nebot, F. Castro, F. Mugica and A. Vellido, “Identification of fuzzy models to predict students performance in an E-learning environment,” in *Proceedings of the Fifth IASTED International Conference on Web-based Education*, 2006.

[36] M. Hall and L. Smith, “Practical feature subset selection for machine learning,” *Computer Science,* 1998.

[37] D. Romer, “Do Students Go to Class? Should They?,” 1993.

[38] E. E. Cruz-Miguel, J. Rodríguez-Reséndiz, J. R. García-Martínez, K. A. Camarillo-Gómez and G. I. Pérez-Soto, “Field-programmable gate array-based laboratory oriented to control theory courses,” *Computer Applications in Engineering Education,* vol. 27, no. 5, pp. 1253-1266, 7 9 2019.

[39] A. Sánchez-Carmona, S. Robles and J. Pons, “A gamification experience to improve engineering students' performance through motivation,” *Journal of Technology and Science Education,* vol. 7, no. 2, p. 150, 7 6 2017.

[40] J. E. McLaughlin, L. M. Griffin, D. A. Esserman, C. A. Davidson, D. M. Glatt, M. T. Roth, N. Gharkholonarehe and R. J. Mumper, “Pharmacy student engagement, performance, and perception in a flipped satellite classroom.,” *American journal of pharmaceutical education,* vol. 77, no. 9, p. 196, 12 11 2013.

[41] E. Ushida, *The Role of Students' Attitudes and Motivation in Second Language Learning in Online Language Courses,* vol. 23, Equinox Publishing Ltd., 2005, pp. 49-78.

[42] J. Huizenga, W. Admiraal, S. Akkerman and G. t. Dam, “Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game,” *Journal of Computer Assisted Learning,* vol. 25, no. 4, pp. 332-344, 6 7 2009.

[43] J.-S. Lee, “The Relationship Between Student Engagement and Academic Performance: Is It a Myth or Reality?,” *The Journal of Educational Research,* vol. 107, no. 3, pp. 177-185, 4 5 2014.

[44] M. D. Pistilli and K. E. Arnold, “In practice: Purdue Signals: Mining real-time academic data to enhance student success,” *About Campus,* vol. 15, no. 3, pp. 22-24, 1 7 2010.

[45] “scikit-learn: Machine Learning in Python,” [Online]. Available: http://scikit-learn.org/. [Accessed 20 12 2018].

[46] “Rapidminer,” [Online]. Available: https://rapidminer.com/. [Accessed 20 12 2018].

Capítulo 3 Análisis de supervivencia

“Un hombre sabio será siempre un aprendiz de maestro.”

Platón (427- 347 a.C.)

Artículos que se presentan:

Martínez-Carrascal, J. A., & Sancho-Vinuesa, T. (2021). Does failing the first assessment affect withdrawal decisions at course level? An empirical evidence. Proc. of Learning Analytics Summer Institute Spain 2021, Barcelona, Spain, July 7-9, 2021, CEUR-WS.org, online CEUR-WS.orgVol-3029/paper01.pdf

Martínez-Carrascal, J. A., Hlosta, M., & Sancho-Vinuesa, T. (2023). Using Survival Analysis to Identify Populations of Learners at Risk of Withdrawal: Conceptualization and Impact of Demographics. *The International Review of Research in Open and Distributed Learning*, *24*(1), 1–21. https://doi.org/10.19173/irrodl.v24i1.6589

Martínez-Carrascal, J. A., & Sancho-Vinuesa, T. (2023). Exploring the impact of assessment characteristics on course withdrawal: a survival analysis approach. (en proceso de publicación de los *Procedings* de LASI Spain 2023)

3.1. Contextualización de los artículos

Dentro del logro académico, una de las preocupaciones nucleares, ya explicitadas anteriormente, es el abandono. Aun cuando el abandono puede analizarse a distintos niveles, nuestro foco de estudio será el abandono a nivel de curso. Reducir las tasas de abandono mediante intervenciones apropiadas es una de las mejores palancas en la mejora de resultados académicos.

Tal y como hemos planteado en los capítulos previos, nuestra idea no es abordar el problema buscando maximizar la capacidad predictiva – y en particular la precisión –, sino en términos de interpretabilidad de resultados. Esta búsqueda de la interpretabilidad y del análisis de las razones subyacentes es común en otras disciplinas. Por ello hemos buscado realizar una prospección de técnicas más allá de los contextos educativos.

Analizando específicamente el abandono, y según se desprende del capítulo previo, debemos tener en consideración la relevancia del factor tiempo. La interpretabilidad, unida a la relevancia del tiempo, nos lleva a considerar la aplicación del análisis de supervivencia. Este tipo de análisis es particularmente utilizado en entornos médicos, cuando el objeto es determinar poblaciones de riesgo – bien sea por características preexistentes, por indicios exploratorios o por analíticas de control –, cuantificar el riesgo subyacente y/o determinar la bondad de un determinado tratamiento.

Como veremos, esta técnica admite una traslación directa al ámbito educativo. Haciendo un paralelismo, consideraremos el abandono a nivel de curso como una enfermedad a tratar, donde es muy relevante determinar factores de riesgo, su impacto individual o combinado, y los mejores *tratamientos* (que serán intervenciones desde la perspectiva de *learning analytics*). El análisis de supervivencia nos permitirá detectar qué situaciones pueden ser posibles factores de riesgo. Podrá cuantificarse su efecto, y en base al impacto esperado, diseñar intervenciones específicas. Podrá también evaluarse el efecto que tiene una intervención, o comparar diferentes intervenciones sobre un mismo escenario (lo que comúnmente se conoce como *A/B testing*).

El objetivo último de la aplicación de la técnica que se plantea a entornos académicos es la mejora del logro académico a nivel de curso, que deriva de la reducción potencial del abandono. Nuestra exploración de esta vía de análisis se ha traducido en tres publicaciones. Una preliminar exploratoria (Martínez-Carrascal & Sancho-Vinuesa, 2021), que valida una línea de investigación previa (Figueroa-Cañas & Sancho-Vinuesa, 2021) y donde se establece el impacto que tiene en el abandono superar la primera de las pruebas del curso. La segunda (Martínez-Carrascal, Hlosta, et al., 2023), donde sentamos las bases de la modelización, aplicándolo a la detección de poblaciones de riesgo basándonos en la información disponible al iniciar un determinado estudio. Finalmente, el último artículo de este bloque (Martínez-Carrascal & Sancho-Vinuesa, 2023) aplica el mismo concepto para establecer el impacto que los parámetros de las pruebas evaluativas tienen en el abandono.

Como introducción a estos artículos, expondremos las bases teóricas del análisis de supervivencia (Sección 3.2). Seguidamente, expondremos cómo el análisis de supervivencia permite profundizar en el análisis, comprensión y finalmente reducción del abandono a nivel de curso (Sección 3.3.). Un breve resumen de las conclusiones que se desprenden de los artículos presentados en este capítulo deja paso a los artículos publicados (Sección 3.4).

Indicar que los tres trabajos de este bloque se han realizado basándonos en una base de datos abierta, proporcionada por la Open University. Este hecho supone un valor añadido, ya que al trabajar con datos abiertos, se permite la validación, reproducción y extensión de los resultados presentados en este capítulo.

3.2. Fundamentos del análisis de supervivencia

Los fundamentos del análisis de supervivencia se exponen en los artículos de Clark et al. (Bradburn et al., 2003a, 2003b; Clark et al., 2003a, 2003b) Igualmente, uno de nuestros estudios expone la base de su funcionamiento y su portabilidad al análisis del abandono académico (Martínez-Carrascal, Hlosta, et al., 2023). Se trata de una técnica de particular utilidad cuando el aspecto más relevante es el tiempo transcurrido hasta la materialización de un determinado evento. Es comúnmente aplicado para el análisis del período de tiempo hasta la recaída – o incluso la muerte – en ciertas patologías, cómo ciertas características influyen en él, y cómo este tiempo puede verse alterado con tratamientos específicos.

Para su uso resulta fundamental establecer tres parámetros: **el evento objeto de estudio** – el abandono en nuestro caso –, el **período de observación** – que en nuestro caso irá desde el momento de matriculación hasta la finalización del curso – y la "**supervivencia**" – que en el estudio del abandono corresponde al tiempo que el estudiante permanece formalmente activo a nivel académico –.

La cuantificación del impacto de un factor o un conjunto de factores combinados se puede determinar mediante métodos paramétricos o no paramétricos. Los métodos no paramétricos presentan dos ventajas. Por uno, la simplicidad. Por otro, no requieren conocer a priori la distribución estadística de la supervivencia. Estas razones justifican en nuestro caso el uso de un método no paramétrico. En particular, proponemos las estimaciones de Kaplan-Meier, fundamentalmente por no requerir el conocimiento previo de la distribución estadística. Esto no supone pérdida de aplicabilidad, pero cabe indicar que en sistemas que presentan distribuciones conocidas y estables en lo que respecta a patrones del abandono, métodos paramétricos más complejos pueden proporcionar resultados más ajustados.

Para el detalle tanto de los parámetros clave como de las técnicas de cuantificación nos remitimos a la sección *Survival analysis* de nuestra cuarta contribución según la Tabla 1.1. y que se presenta en este capítulo (Martínez-Carrascal, Hlosta, et al., 2023).

3.3. Portabilidad a escenarios educativos del análisis del abandono

El concepto introducido en la sección 3.2. admite una traslación directa a problemas de *learning analytics* vinculados al análisis del abandono. El uso del método en entornos educativos es poco frecuente si se compara con las técnicas analizadas en el capítulo anterior. Como se muestra en uno de los artículos de este bloque, no existen referencias de uso en cursos reglados *online* de educación superior (Martínez-Carrascal, Hlosta, et al., 2023).

Los artículos que se presentan en este bloque demuestran que el concepto es de aplicación al tratamiento del abandono. En particular la aplicación del método permitirá establecer qué factores aumentan el riesgo y qué poblaciones son más susceptibles. Finalmente, la idea detrás de la aplicación es tratar el abandono tal y como lo hace el ámbito médico para determinadas enfermedades. Por ello, será también adecuado para escoger qué tratamiento presenta mejor pronóstico para el estudiante. Dicho de otra forma, será una herramienta útil para enfocar el diseño de intervenciones.

Aun con la idea en mente de esta traslación, conviene remarcar la necesidad de adaptar los lenguajes y marcos conceptuales utilizados en las diferentes disciplinas. El uso de análisis de supervivencia debe entenderse – una vez trasladado el método – como una herramienta más al servicio de la mejora académica. Es además una herramienta que no anula ninguna de las existentes, sino que las complementa aportando una interesante visión evolutiva.

Nuestro modelo, y tal como se recoge específicamente en la Sección titulada '*Porting survival analysis to withdrawal analysis*' en (Martínez-Carrascal, Hlosta, et al., 2023), considera que el momento del abandono es un factor clave. Si consideramos que el estudiante tiene un período de vida en el curso, nuestro objetivo inicial es que ese período de vida se extienda durante toda la duración del curso.

En entornos *online* de educación superior, donde el abandono es común, proponemos un análisis, bien sea de factores personales (Martínez-Carrascal, Hlosta, et al., 2023) o bien de aspectos vinculados características propias del curso, como la realización de determinadas tipologías de pruebas o su frecuencia (Martínez-Carrascal & Sancho-Vinuesa, 2023). En ambos casos comprobaremos que podremos determinar el incremento o reducción del riesgo de abandono en función de los factores objeto de análisis.

El análisis tanto de qué factores se muestran relevantes como de cuál es su impacto en el abandono permite el diseño y la justificación de intervenciones. A modo de ejemplo, permite demostrar que los alumnos de niveles bajos de renta son particularmente sensibles al período de retorno de matrícula (Martínez-Carrascal, Hlosta, et al., 2023). Esto hace que muchos de ellos opten por abandonar en las etapas tempranas de curso. El impacto se demuestra aún mayor si se trata de estudiantes que declaran discapacidad. A partir de este hecho, se puede sugerir alargar el período de abandono sin penalización como medida inicial para estos colectivos. Para un mayor detalle, se sugiere la revisión detallada de las secciones *Implications* y *Conclusions* del artículo referido (Martínez-Carrascal, Hlosta, et al., 2023) así como la sección *Discussion* del último de nuestros artículos de este capítulo (Martínez-Carrascal & Sancho-Vinuesa, 2023).

En conjunto, y de la misma forma que en el ámbito médico se sugiere uno u otro tipo de tratamiento, la portabilidad de la herramienta sugiere la conveniencia de uno u otro tipo de intervención según la causa subyacente y su impacto específico. Más allá de la cuantificación, el foco no estará en la eficacia y la evaluación de la precisión de los algoritmos, sino en la búsqueda de las causas subyacentes tras el abandono con el propósito de reducirlo.

3.4. Conclusiones del capítulo

En este capítulo se ha mostrado como la portabilidad del análisis de supervivencia permite establecer los factores que pueden condicionar el abandono, aportando información de las causas que subyacen tras él, cuantificando factores de riesgo, y estimando la bondad de intervenciones específicas. Es por tanto una herramienta de utilidad indiscutible para abordar el problema planteado en la presente tesis: mejorar el rendimiento académico a partir de la exploración de técnicas.

Los resultados que derivan de la aplicación del método suponen una ayuda valiosa en el diseño de intervenciones. Esta ayuda incluye establecer qué acciones pueden impactar en mayor medida en la reducción, pero también qué colectivos son destinatarios prioritarios de estas actuaciones.

Queremos indicar también que más allá del colectivo destinatario, la modelización planteada permite estimar el momento adecuado para la intervención, así como aspectos no necesariamente vinculados al estudiante, sino al diseño pedagógico del curso. Los artículos que figuran seguidamente proporcionan los detalles específicos de estos hallazgos.

Aun así, es conveniente indicar que, pese a su aproximación temporal continua, el análisis no constituye una modelización completa del proceso. Para ello, serán necesarias otras técnicas, que se expondrán en el siguiente capítulo.

Artículo 3: Does failing the first assessment affect withdrawal decisions at course level? An empirical evidence.

Martínez-Carrascal, J. A., Sancho-Vinuesa, T. In Proc. of Learning Analytics Summer Institute Spain 2021, Barcelona, Spain, July 7-9, 2021, CEUR-WS.org, *online* https://ceur-ws.org/Vol-3029/paper01.pdf

Scopus: SJR 2021: 0.228, Category: Computer Science (miscellaneous)

Nota: Las referencias numéricas a figuras y tablas contenidas en el artículo que se presenta a continuación deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

Does failing the first assessment affect withdrawal decisions at course level? An empirical evidence

Juan Antonio Martínez-Carrascal [1] [0000-0002-7696-6050] and Teresa Sancho-Vinuesa[1] [0000-0002-0642-2912]

[1] Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain

[jmartinezcarra, tsancho]@uoc.edu

Abstract. Classical evaluation models were based on measuring the grades of a test or set of tests performing along a specific course. The method had drawbacks, and continuous evaluation is nowadays a preferred method.

Continuous evaluation performs a continuous measure of the progress the student fulfils to achieve learning outcomes. However, and in practice, continuous evaluation systems usually consist – or at least, include – a set of assessments along the course which are graded and computed.

In this article, we demonstrate the relevance of the first test grade in terms of withdrawal. We make use of a quite uncommon method in learning analytics, such as survival curves. The method is commonly used in other fields, and in particular in medicine, to detect differences among populations under study.

Practical implementation will be carried out by analyzing a dataset containing higher education online university courses. Results show that the students failing the first grade show not only higher withdrawal rates, but also disengage earlier from the course. This evidence should reinforce the need to design actions targeted to this group, but also to use an initial test as a measure of engagement.

Keywords: withdrawal, assessment failure, student grades, survival analysis, learning analytics.

1 Introduction

Dropout in general, and withdrawal in particular, is one of the core problems of higher education institutions. Dropout means inefficient use of resources, and at the same time, implies frustration for students who leave either the system as a whole or a specific subject in particular.

This fact was one of the topics the Bologna process would aim to resolve, in a general framework of trans-European coordination [1]. As a specific factor, continuous evaluation was encouraged. Instead of the classical approach, where the students had a reduced set of tests to evaluate performance, competencies were considered the key, and continuous evaluation was encouraged.

From a practical point of view, this continuous evaluation would require "the evaluation of a subject through daily classwork, course-related projects, and/or practical work(s) instead of a final examination system." [2]. In practice, however, tests are still being performed and competence acquisition is often measured through grades gathered in a set of tests.

In this scenario, we hypothesize that continuous evaluation is conditioned in some way by the first of the evaluative assessments. Although its relative weight can be minor compared to the whole set of activities, a low grade can have a discouraging effect on the student. To validate our hypothesis, we raise the following research question:

- RQ: To what extent does failing the first evaluative assessment condition withdrawal decision of online students?

As it can be seen we look for an answer that goes beyond a pure affirmative question, indicating that there can be an influence. We aim to compare withdrawal depending on this first test, quantifying the specific impact.

To provide additional significance, we will analyze an assorted set of courses from an open database provided by the Open University. It includes data about a set of courses that have not been specifically designed for our research, including 22 editions of 7 different courses with over 30.000 total enrollments. Results will show that the students failing this first test show not only a higher withdrawal ratio but also tend to abandon the course earlier.

2 Theoretical framework

The theoretical models behind dropout were established around 1975. Works by Tinto [3] establish the first model on the topic. Tinto's model was known as the student integration model. It included both academic factors related to the student herself and factors related to the institution. As a whole, the model considered a set of interactions that conditioned the decision to drop out.

After this initial model, we can find different works that rely on this theory. [4] introduced the 'student attrition model' which relies on the concept of behavioural intention, where dropout is conditioned by a mixture of factors, which include academic, social-psychological, environmental, and socialization factors.

None of these theories considers specifically online studies. [5] performs an analysis, based on the previous theories, and conforms the 'composite persistence model'. In this model, academic performance and dropout are finally a combination of student characteristics, student skills, external factors, and internal factors. The three models abovementioned are the most cited references for the study of dropout but are not unique [6]. We can also cite models by Kember [7] and Lee and Choi [8].

The core of these models is centred on university dropout, which could be defined as 'leaving the university study in which they have enrolled before they have obtained a formal degree'[9]. However, this phenomenon can also be analyzed at a micro-level. In particular, considering the fact of a student leaving a course she is enrolled in. In this case, the term withdrawal is preferred, although compilation works show that the formal definition is unclear. 78% of the recent studies do not provide a clear definition of the term [6]. There are also no specific theoretical models for withdrawal.

For the sake of our research, we will consider it as "voluntary or involuntary removal from a course before completion", a consistent definition with references in the literature [10], [11]. It is noticeable that the concept includes not only the decision to abandon the course but also considers the time, as withdrawal is carried out before the end of the course.

Withdrawal analyses are normally set up on specific course analyses. Besides, we can find a mixture of quantitative and qualitative analysis. Early works related to withdrawal in online environments detected that the pressure of work, technical problems, and lack of time were withdrawal determinants[12]. More recently, focuses on family and organizational support, and course satisfaction and relevance.

Despite the relevance of time in withdrawal, time analysis is uncommon. Most studies are limited to a classification problem, aimed to determine variables the influence whether a student withdraws or not. Among those references to studies considering the relevance of time, we can cite [13], [14] which are focused on university dropout. Focused on a specific course, we can find a MOOC case example[15]. It must be pointed out that most studies focus on the institutional level, and not specifically on withdrawal at the course level.

Among those techniques to approach the problem, we can find correlation analysis, classifiers – both Bayesian and different decision trees -, variance analysis, logistic regression, support vector machines, neural networks, or machine learning techniques. These techniques are found both at university and course level and also in traditional university courses and MOOCs [16], [17]. MOOCs are one of the fields where withdrawal has been more analysed due to its higher rates[16].

Despite survival analysis is commoly used in other disciplines[18], references to survival analysis techniques in e-Learning problems are not so common. The basics behind the technique are described in [19]. The interest on it is more than justified, due to its focus on time – which is particularly relevant when analysing withdrawal – but also for providing better results that classical approaches in terms of prediction[20].

[20] suggests that more research should be performed using this approach. Among those works focusing on time, we can cite [13], [21]. Results in [21] indicate that the beginning of the course is a critical moment that concentrates a high number of withdrawals. [13] performs a survival analysis over time from a university-level perspective, with results showing that grade point average at first semester, gender, and location are relevant for determining university dropout.

Whichever method, the relevance of early activity is considered in different studies. Early activity in general is considered a predictor of final course performance[22]–[24]. Assignment grades in particular constitute a strong predictor of the final performance in MOOC courses [25].

In this scenario, we analyze the specific impact of early grades in evaluative assessments on the withdrawal decision of the student.

3 Methodology

From a methodological perspective, two critical factors arise. First, the technique to be used. Second, a specific database to work with. Regarding the method, and due to the relevance of time, we will map our study as a survival analysis problem, as described in Section 3.1. Regarding the data, we will make use of a publically available database created by the Open University. Details of this database are included in Section 3.2..

3.1 Mapping withdrawal as a survival analysis problem

Survival analysis is 'a collection of statistical procedures for data analysis where the outcome variable of interest is time until an event occurs' [19]. The method is commonly used in other disciplines such as medicine, where survival time or time to relapse is under consideration. A really interesting view of the technique with a practical approach can be seen in a series of articles[19], [26]–[28].

References to survival analysis are scarce in the field of education in general and withdrawal in particular. As indicated in Section 2, we can cite a couple of analyses of university dropout [13], [14] and another focused on MOOC courses[15].

Two specific aspects are needed to perform survival analysis, which are the event under consideration, and the time to event. The event under consideration will be the fact of withdrawing, while the time to event will be the number of days the student remains enrolled in the course.

As different courses will be analyzed we will consider as t=0 the initial day of the course. Times above t=0 will be interpreted as the number of days after the course starts. Negative values reflect a withdrawal after enrolling but before the course effectively starts.

As specific tools we will make use of Kaplan-Meier curves to visualize and analyze the relevance of the variable under analysis, looking for statistical significance and clear interpretation[19]. Statistical validation will be performed considering the null hypothesis that different groups generated based on the grade of the first assessment share the same hazard functions. Log-rank test (in particular, Peto's) will be used for being more robust, and also as it provides more weight to earlier events [29]. As a limitation, Kaplan-Meier does not allow quantifying hazard. Hazard can be computed by using a simple non-parametric method, such as Nelson-Aalen[19].

To quantify the impact, and considering we are not segregating populations based on multiple parameters, we can use a non-parametric method. In particular, we will use the Nelson-Aalen method to estimate cumulative hazard. Although non-cumulative hazard at a specific time can also be computed, cumulative estimation is preferred for being more stable.

3.2 Working dataset

The search for a dataset that allows for analysis linked to our RQ has lead us to consider the public dataset offered by the Open University[30]. At the highest level, this dataset provides information about 22 editions – namely presentations in the OU nomenclature – of 7 different courses. All courses present at least two editions. A total of 32,593 students are enrolled in these courses – modules in the OU nomenclature -.

This database includes both personal and academic data of the students under consideration. For the sake of our purpose, specific information to determine withdrawal – and in particular, withdrawal date - is included. Regarding assessments, the dataset also includes the whole set of evaluative activities linked to every course, with its weight and grades obtained by the students in the different editions of the course.

Table 1 includes information about the first evaluative assessment of the different courses under analysis, as well as its weight. We also provide the total course duration.

Table 1. Data regarding course characteristics

Module	Presentation	Presentation length (days)	Date of 1st assessment (days since presentation start)	Weight of 1st assessment (%)
AAA	2013J	268	19.0	10.0
AAA	2014J	269	19.0	10.0
BBB	2013J	268	19.0	5.0
BBB	2014J	262	19.0	0.0
BBB	2013B	240	19.0	5.0
BBB	2014B	234	12.0	5.0
CCC	2014J	269	32.0	9.0
CCC	2014B	241	32.0	9.0
DDD	2013J	261	25.0	10.0
DDD	2014J	262	20.0	5.0
DDD	2013B	240	25.0	7.5
DDD	2014B	241	25.0	10.0
EEE	2013J	268	33.0	16.0
EEE	2014J	269	33.0	16.0
EEE	2014B	241	33.0	16.0
FFF	2013J	268	19.0	12.5
FFF	2014J	269	24.0	12.5
FFF	2013B	240	19.0	12.5
FFF	2014B	241	24.0	12.5
GGG	2013J	261	61.0	0.0
GGG	2014J	269	61.0	0.0
GGG	2014B	241	61.0	0.0

Table 2 shows an overview of the course enrolment and dropout, including percentage of withdrawal, failure and pass. The pass group includes also those students qualified with distinction:

Table 2. Enrolment and ratios linked to dropout analysis for courses in the OU dataset

Course	#Students	Withdrawals	Fail	Pass
AAA	747	16.73%	12.18%	71.08%
BBB	7903	30.18%	22.32%	47.50%
CCC	4434	44.54%	17.61%	37.84%
DDD	6266	35.86%	22.49%	41.65%
EEE	2934	24.61%	19.15%	56.23%
FFF	7758	30.96%	22.02%	47.03%
GGG	2534	11.52%	28.73%	59.75%

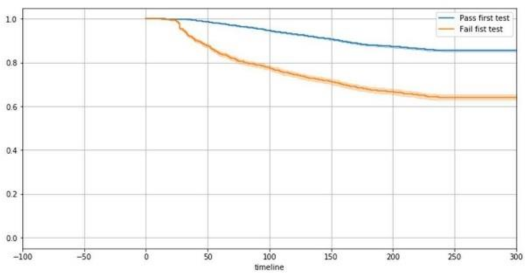
With this information, we can map the problem as a survival analysis problem as indicated in Section 3.1. It is also noticeable that the data used is suitable for performing Kaplan-Meier analysis. The OU has an active policy to manage dropout. Withdrawal time is always recorded. Due to this fact, independence of censoring and survival is guaranteed.

4 Results

4.1 Relevant difference in withdrawal depending on first assessment failure

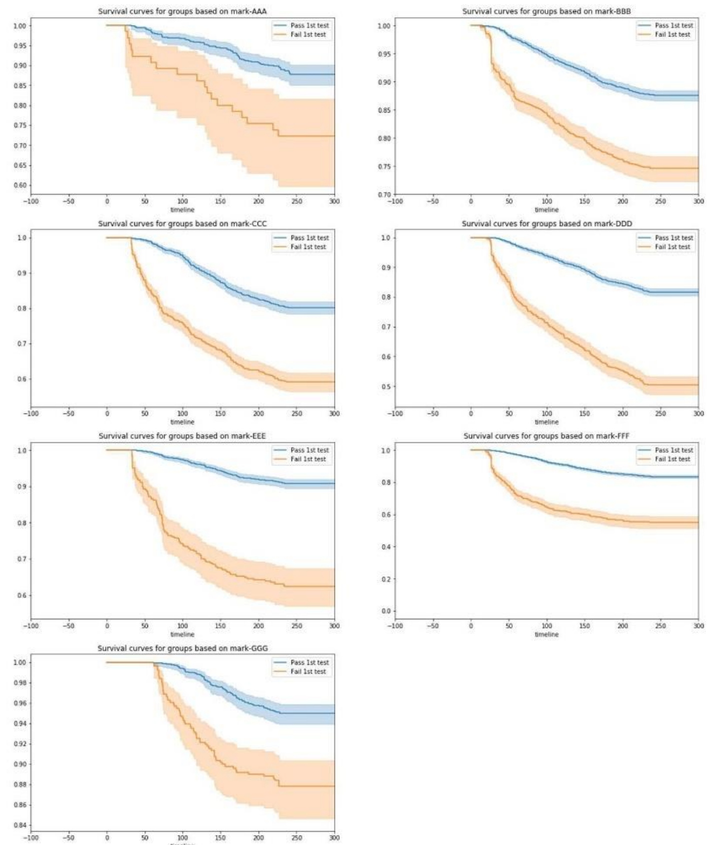
Before quantifying the potential impact, a detailed analysis must be performed to determine whether passing students and failing students of the first test show statistically significant differences in withdrawal patterns. Figure 1 shows the results of the Kaplan-Meier estimates comparing both groups:

Fig. 1. Survival curves for groups generated based on difference in first test



Considering that courses have different withdrawal ratios, we have analyzed also curves on a per-course basis. Results are shown in figure 2, where curves are shown including confidence intervals:

Fig. 2. Survival curves for individual courses



Statistical comparison for all groups results in relevant differences (p<0.005) in all cases, indicating both groups show different patterns regarding withdrawal. Table 3 reflects final withdrawal ratios for the generated groups and different courses, including the increase factor to facilitate comparison.

Table 3. Differences in withdrawal ratios at course end between students who pass and fail first assessment

	Pass 1st test	Fail 1st test	Factor
AAA	0.122	0.277	2.27
BBB	0.125	0.263	2.11
CCC	0.199	0.428	2.15
DDD	0.184	0.510	2.77
EEE	0.092	0.388	4.22
FFF	0.166	0.482	2.91
GGG	0.051	0.126	2.48

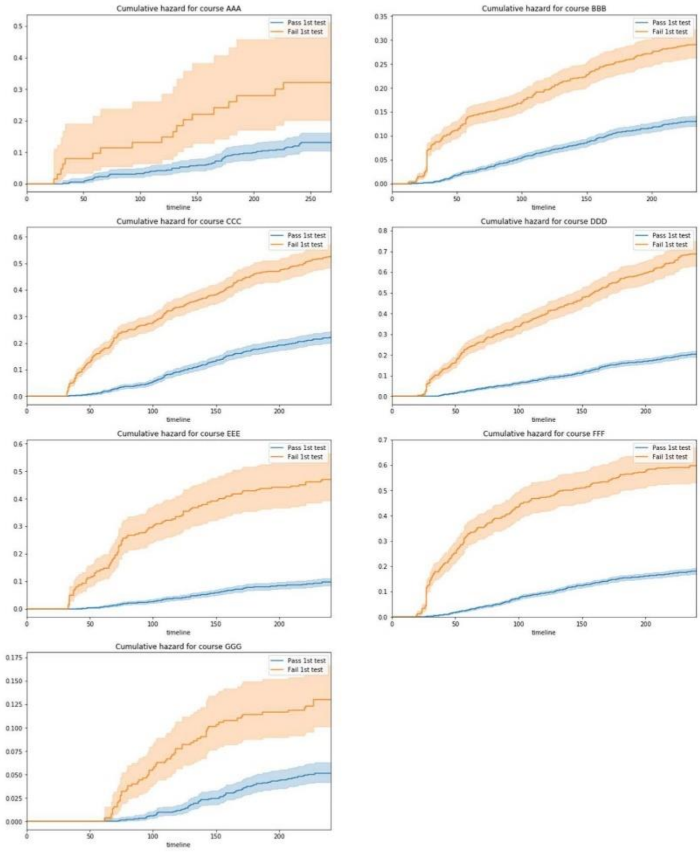
4.2 Higher and earlier withdrawal depending on the result of the first assessment

While survival curves provide group comparison in terms of survival probability, we can still get deeper into analyzing withdrawal hazard. Kaplan-Meier estimates do not provide this information, and we have to make use of specific methods to compute it. In particular, and considering that populations are segmented based on a single covariate, and we have no assumption about distribution, we can use a non-parametric estimator. In particular, we use Nelson-Aalen.

Nelson-Aalen is used to compute the cumulative hazard risk, understood as the probability of a student withdrawing from the course within a small interval of time, assuming she has survived up until the beginning of that interval. Cumulative hazard is preferred to point-wise estimations for being more stable.

In our case study, plots over time for individual courses have been plotted in Figure 3.

Fig. 3. Cumulative hazard curves for individual courses



As it can be seen, those students who fail the first test, show higher withdrawal rates in the long term, but also a relevant increase in early withdrawal.

5 Discussion

The research carried out contributes in two specific ways. First, it demonstrates the utility and potential application of survival analysis applications in e-learning. Second, it provides insight into the relevance of the first assessment in withdrawal.

Focusing on the RQ stated, results in section 4 clearly show that there are relevant differences in withdrawal among those students who fail the first test and those who pass it. This difference is reflected in Figures 2 and 3 for individual courses.

Regardless of the course, global withdrawal ratios are higher for those students failing the first test. As Figure 1 shows, the mean difference in survival at the end of the course is 2.57 higher for those students who pass the first test. When looking at individual courses, the increase in withdrawal at course end can be 4.22 times higher as reflected in Table 3. Besides this difference in final withdrawal, Figure 3 shows also an interesting insight. Much of the above difference is based on a much higher early withdrawal. Before going deeper into this fact, it must be pointed out that these tests are made in an early period when comparing to course duration. Data in Table 1 show that – except for course GGG - the test is performed around the first month of the course, in a course lasting for around 9 months. Also, some of the assessments do not even compute for global course grade and are under 20% in all cases.

With this data in mind, our results would indicate that the first assessment of a course has an interesting predictive power, regardless of its weight and even time. From a learning analytics perspective, the group of students who fail the first test are suitable for targeted interventions aimed to retain them in the course. This result is aligned with the literature reflecting the impact of early activity in a broader sense [22]–[24].

From a methodological perspective, the method exposed fits the suggestion to explore survival analysis in e-learning scenarios[20] with a specific application to withdrawal analysis. It is at least noticeable that a method that is common in medical research shows really few references in e-learning. The authors are open to collaborate in research lines following this idea, and in particular, linked to the analysis of the impact of different factors on withdrawal.

References

[1] A. Veiga, "Researching the bologna process through the lens of the policy cycle," in *European and Latin American Higher Education Between Mirrors*, Sense Publishers, 2014, pp. 91–108.

[2] E. S. Sanz-Pérez, "Students' performance and perceptions on continuous assessment. Redefining a chemical engineering subject in the European higher education area," *Educ. Chem. Eng.*, vol. 28, pp. 13–24, Jul. 2019, doi: 10.1016/j.ece.2019.01.004.

[3] V. Tinto, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Rev. Educ. Res.*, 1975, doi: 10.3102/00346543045001089.

[4] J. P. Bean, "Interaction Effects Based on Class Level in an Explanatory Model of College Student Dropout Syndrome," *Am. Educ. Res. J.*, 1985, doi: 10.3102/00028312022001035.

[5] A. P. Rovai, "In search of higher persistence rates in distance education online programs," *Internet High. Educ.*, vol. 6, no. 1, pp. 1–16, Jan. 2003, doi: 10.1016/S1096-7516(02)00158-6.
[6] M. Xavier and J. Meneses, "Dropout in Online Higher Education: A scoping review from 2014 to 2018," doi: 10.7238/uoc.dropout.factors.2020.
[7] D. Kember, "A Longitudinal-Process Model of Drop-Out from Distance Education," *J. Higher Educ.*, vol. 60, no. 3, pp. 278–301, May 1989, doi: 10.1080/00221546.1989.11775036.
[8] Y. Lee and J. Choi, "A review of online course dropout research: Implications for practice and future research," *Educational Technology Research and Development*, vol. 59, no. 5. pp. 593–618, Oct. 2011, doi: 10.1007/s11423-010-9177-y.
[9] M. S. Larsen, *Dropout Phenomena at Universities: What is Dropout? Why Does Dropout Occur? What Can be Done by the Universities to Prevent Or Reduce It? : a Systematic Review*. Danish Clearinghouse for Educational Research, 2013.
[10] J. M. Lim, "Predicting successful completion using student delay indicators in undergraduate self-paced online courses," *Distance Educ.*, vol. 37, no. 3, pp. 317–332, Sep. 2016, doi: 10.1080/01587919.2016.1233050.
[11] T. J. McClelland, "Why do they leave? an exploration of situational, dispositional, institutional, technological, and epistemological factors on undergraduate student withdrawal from online studies at an institute of technology in New Zealand," Nov. 2014.
[12] G. Packham, P. Jones, C. Miller, and B. Thomas, "E-learning and retention: Key factors influencing student withdrawal," *Educ. + Train.*, vol. 46, pp. 335–342, Aug. 2004, doi: 10.1108/00400910410555240.
[13] J. C. Juajibioy, "Study of University Dropout Reason Based on Survival Model," *Open J. Stat.*, vol. 06, no. 05, pp. 908–916, 2016, doi: 10.4236/ojs.2016.65075.
[14] Y. Min, G. Zhang, R. A. Long, T. J. Anderson, and M. W. Ohland, "Nonparametric survival analysis of the loss rate of undergraduate engineering students," *J. Eng. Educ.*, 2011, doi: 10.1002/j.2168-9830.2011.tb00017.x.
[15] D. Yang, T. Sinha, D. Adamson, and C. Rose, "'Turn on, Tune in, Drop out': Anticipating student dropouts in Massive Open Online Courses," in *Proceedings of the NIPS Workshop on Data Driven Education*, 2013.
[16] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Munoz-Merino, and C. D. Kloos, "Prediction in MOOCs: A Review and Future Research Directions," *IEEE Trans. Learn. Technol.*, 2019, doi: 10.1109/TLT.2018.2856808.
[17] L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, "Predicting Student Dropout in Higher Education," 2016.
[18] F. Emmert-Streib and M. Dehmer, "Introduction to Survival Analysis in Practice," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 3, pp. 1013–1038, Sep. 2019, doi: 10.3390/make1030058.
[19] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, "Survival Analysis Part I: Basic concepts and first analyses," *British Journal of Cancer*. 2003, doi: 10.1038/sj.bjc.6601118.
[20] S. Ameri, M. J. Fard, R. B. Chinnam, and C. K. Reddy, "Survival analysis based framework for early prediction of student dropouts," in *International Conference on*

Information and Knowledge Management, Proceedings, 2016, doi: 10.1145/2983323.2983351.

[21] S. Laato, E. Lipponen, H. Salmento, H. Vilppu, and M. Murtonen, "Minimizing the Number of Dropouts in University Pedagogy Online Courses," doi: 10.5220/0007686005870596.

[22] J. A. Martínez-Carrascal, D. Márquez Cebrián, T. Sancho-Vinuesa, and E. Valderrama, "Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course," *Comput. Appl. Eng. Educ.*, vol. 28, no. 3, pp. 590–605, May 2020, doi: 10.1002/cae.22229.

[23] L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an 'early warning system' for educators: A proof of concept," *Comput. Educ.*, 2010, doi: 10.1016/j.compedu.2009.09.008.

[24] V. A. Nguyen, "The Impact of Online Learning Activities on Student Learning Outcome in Blended Learning Course," *Journal of Information and Knowledge Management*. 2017, doi: 10.1142/S021964921750040X.

[25] S. Jiang, A. E. Williams, K. Schenke, M. Warschauer, and D. O. Dowd, "Predicting MOOC Performance with Week 1 Behavior," *Proc. 7th Int. Conf. Educ. Data Min.*, 2014.

[26] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival Analysis Part II: Multivariate data analysis- An introduction to concepts and methods," *British Journal of Cancer*. 2003, doi: 10.1038/sj.bjc.6601119.

[27] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival Analysis Part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit," *British Journal of Cancer*. 2003, doi: 10.1038/sj.bjc.6601120.

[28] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, "Survival analysis part IV: Further concepts and methods in survival analysis," *British Journal of Cancer*. 2003, doi: 10.1038/sj.bjc.6601117.

[29] S. Yang and R. Prentice, "Improved logrank-type tests for survival data using adaptive weights," *Biometrics*, 2010, doi: 10.1111/j.1541-0420.2009.01243.x.

[30] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Data Descriptor: Open University Learning Analytics dataset," *Sci. Data*, vol. 4, no. 1, pp. 1–8, Nov. 2017, doi: 10.1038/sdata.2017.171.

Artículo 4: Using survival analysis to identify populations of learners at risk of withdrawal: conceptualization and impact of demographics.

Martínez-Carrascal, J. A., Hlosta, M., & Sancho-Vinuesa, T. (2023). Using Survival Analysis to Identify Populations of Learners at Risk of Withdrawal: Conceptualization and Impact of Demographics. The International Review of Research in Open and Distributed Learning, 24(1), 1-21. https://doi.org/10.19173/irrodl.v24i1.6589 ISSN 1492-3831

Scopus: SJR 2021: 1.322 Q1 Education Q1 E-learning

WoS: JFI 2021: 2.770 Q2 Education & Educational Research

Nota: Las referencias numéricas a figuras y tablas contenidas en el artículo que se presenta a continuación deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 24, Number 1

February – 2023

Using Survival Analysis to Identify Populations of Learners at Risk of Withdrawal: Conceptualization and Impact of Demographics

Juan Antonio Martínez-Carrascal[1], Martin Hlosta[2], and Teresa Sancho-Vinuesa[1]

[1]Universitat Oberta de Catalunya; [2]Institute for Research in Open, Distance and eLearning, Swiss Distance University of Applied Sciences

Abstract

High dropout rates constitute a major concern for higher education institutions, due to their economic and academic impact. The problem is particularly relevant for institutions offering *online* courses, where withdrawal ratios are reported to be higher. Both the impact and these high rates motivate the implementation of interventions oriented to reduce course withdrawal and overall institutional dropout. In this paper, we address the identification of populations of learners at risk of withdrawing from higher education *online* courses. This identification is oriented to design interventions and is carried out using survival analysis. We demonstrate that the method's longitudinal approach is particularly suited for this purpose and provides a clear view of risk differences among learner populations. Additionally, the method quantifies the impact of underlying factors, either alone or in combination. Our practical implementation used an open dataset provided by The Open University. It includes data from more than 30,000 students enrolled in different courses. We conclude that low-income students and those who report a disability comprise risk groups and are thus feasible intervention targets. The survival curves also reveal differences among courses and show the detrimental effect of early dropout on low-income students, worsened throughout the course for disabled students. Intervention strategies are proposed as a result of these findings. Extending the entire refund period and giving greater academic support to students who report disability are two proposed strategies for reducing course withdrawal.

Keywords: course withdrawal, demographics, distance education and *online* learning, dropout, intervention design, survival analysis

Using Survival Analysis to Identify Populations of Learners at Risk of Withdrawal: Conceptualization and Impact of Demographics

Academic withdrawal constitutes one of the biggest challenges in education, in particular for *online* higher education (OHE) institutions, where withdrawal ratios are reported to be higher (Bawa, 2016; Simpson, 2010). Aside from its macroeconomic impact, withdrawal causes frustration in terms of expectations, as well as being a waste of time and money from the student's perspective (Lee & Choi, 2013; Simpson, 2010). These facts justify the interest of and motivate these institutions in designing targeted interventions aimed to reduce it.

A critical first step towards a successful intervention is the accurate and reliable identification of learners at risk (Rienties et al., 2016). This identification is mostly understood in terms of prediction. Most research works focus on determining individual risk and on increasing prediction ratios rather than on understanding the reasons behind the risk. While determining if a particular student is at risk can be valuable, the essential issue when considering an intervention is identifying a common risk factor behind a group of learners who may constitute an intervention target.

Furthermore, timely execution is essential. Time plays a particularly relevant role when designing and implementing interventions oriented to reduce course withdrawal and overall university dropout. The moment when a student decides to abandon a course is critical in terms of the intervention design. At the course level, Simpson (2010) showed that 40% of new students at the Open University withdraw from courses before the first assignment. At the university level, Grau-Valldosera et al. (2019) showed that periods of non-enrolment could result in dropout, despite the intention to continue at the time of the break. In both cases, it would be inefficient to implement interventions after the student has effectively dropped out.

When added to the relevance of time, the concept of population at risk—rather than individual at risk—makes us consider survival analysis as a suitable technique. However, a literature review revealed that research using this technique mainly focused on analysing university dropout (Cobre et al., 2019) or attrition in MOOCs (Rizvi et al., 2022; Xing et al., 2019) and was not linked to interventions. Our article focuses on the use of survival analysis as part of the intervention process, detecting populations of learners at risk of withdrawal at the course level in regular OHE courses. The method described will determine the significance and influence of a set of variables on course withdrawal, providing information to select intervention targets and coherent strategies. Additionally, survival curves will provide additional insight which will help the intervention design.

Besides setting the conceptual framework, we performed a practical implementation based on an open dataset from a world-leading *online* university: The Open University Learning Analytics Dataset (OULAD; Kuzilek et al., 2017). This dataset contains data from more than 30,000 students enrolled in 22 *online* course editions from different disciplines, including the withdrawal date for students who abandon different courses. Based on these data, we analysed the impact of students' demographics on withdrawal, determining risk factors and quantifying their impact. Demographics have been identified as some of the causes behind withdrawal (Hachey et al., 2022; Muljana & Luo, 2019) and constitute key features for early dropout prediction in *online* environments (Radovanovic et al., 2021). Nonetheless, the proposed method is applicable to any other variable of interest such as academic performance, background, or psychological features/traits that may impact it.

Literature Review

The Concept of Withdrawal

The analysis of dropout has long been present in educational literature, with 1900–1950 being considered the age of early development, broadening horizons in the 1990s, and showing rising interest in recent years. Compilations can be found linked to higher education (Aljohani, 2016; Behr et al., 2020; Larsen et al., 2013) and specifically to *online* scenarios (Hachey et al., 2022; Lee & Choi, 2011; Muljana & Luo, 2019; Xavier & Meneses, 2020).

Works by Tinto (1975) expounded upon one of the most relevant initial models explaining dropout in traditional education. The core of this theory is the student integration model, where persistence is explained by a student's motivation and ability to match the social and academic characteristics of the institution where she is studying. Years later, Bean (1985) introduced the student attrition model, which relies on the concept of behavioural intention, where dropout is conditioned by a mixture of academic, social-psychological, environmental, and socialisation factors.

These two theories and their combination in Cabrera et al. (1992) and later in Rovai (2003) have been at the core of subsequent studies on the topic. According to Rovai (2003), academic performance and dropout are a combination of student characteristics, student skills, external factors, and internal factors. These four make up the composite persistence model (CPM) and reflect the multivariate nature of dropout.

The term *university dropout* is commonly used to describe situations where students leave the university before obtaining a formal degree (Larsen et al., 2013). Behind this definition lies a complex phenomenon, evidenced by the list of related terms such as dropout, departure, withdrawal, failure, non-continuance or non-completion (Xavier & Meneses, 2020). Dropout is the opposite of retention, defined as "continued student participation in a learning event to completion, which in higher education is a course, program, institution, or system" (Berge & Huang, 2004, p. 3).

At the course level, most papers dealing with withdrawal do not provide a formal definition (77.78% according to a recent scoping review; Xavier & Meneses, 2020). In our research, we used the definition provided by the Open University as "cease studying a module without the intention to resume the study of that module" (Open University, 2022, p. 6).

Approaches for the Identification of Populations at Risk: Survival Analysis

The first stage of a correct intervention design is an accurate and reliable identification of learners at risk (Rienties et al., 2016). Surveys and different data mining techniques are typical approaches used in this identification. Prevalent techniques include decision trees and random forest (Behr et al., 2020), but a whole set of methods can be found in the literature (Xing et al., 2019). However, only a low percentage of studies make use of longitudinal data approaches and, in particular, survival analysis. Ameri et al. (2016) indicated that "there is only a limited attempt at using these methods in student retention problems" (p. 904). Xing et al. (2016) also showed that the performance of classical techniques used to predict dropout could be improved by accommodating temporal modelling approaches.

The use of survival analysis at the course level in the literature is focused on MOOC scenarios (recently Moreno-Marcos et al., 2019; Rizvi et al., 2022; Xing et al., 2019). The existing studies covering survival analysis in OHE all focus on analysing the semesters when students drop out from the university rather than withdrawal from within courses (Ameri et al., 2016; Cobre et al., 2019; Villano et al., 2018). Two of the studies (Ameri et al., 2016; Villano et al., 2018) focused more on comparing the prediction capability of survival methods to existing techniques. On the other hand, Cobre et al. (2019) tried to identify in which semesters students are most likely to drop out, applied in two different academic programmes in Brazil.

Although some studies (Ameri et al., 2016; Villano et al., 2018) highlighted its interpretation of results and its suitability for analysing underlying student issues and helping the design of interventions, none of the studies examined survival analysis itself. Moreover, to the best of our knowledge, none of the studies examined within-course withdrawal. Considering the importance of the moment of withdrawal as well as the method's longitudinal approach and interpretability, we consider it a suitable approach to designing targeted actions oriented to reducing withdrawal.

Influence of Demographics

Rovai's model indicates the relevance of a student's personal factors linked to dropout in *online* studies. Focusing on *online* education, different compilations (Hachey et al., 2022; Lee & Choi, 2011; Muljana & Luo, 2019) investigated the relevance of these factors and showed a lack of consensus among the studies analysed. As noted by Lee and Choi (2011), "findings of many studies were incompatible with one another regarding the relationship between demographics and *online* students' persistence in *online* courses" (p. 603).

In particular, the correlation between gender and course withdrawal is unclear. Some works have indicated a relation, which can even depend on the field of study (Cochran et al., 2014). This work indicated that males showed higher withdrawal rates in courses linked to disciplines such as education or health, but lower in those related to business and math. A large number of studies, however, did not establish a correlation between gender and withdrawal (James et al., 2016; Strang, 2017).

Regarding age, OHE students are older than those in face-to-face learning environments. Once enrolled, older students would have a lower dropout rate (James et al., 2016). Other research, however, did not identify any age-related effects (Strang, 2017).

Prior academic achievement is linked to persistence in *online* learning (Lee & Choi, 2011) and can even be used for prediction (Hachey et al., 2014). Regarding socioeconomic status, it is considered a relevant factor (Hachey et al., 2022). When considering re-enrolment, having a full-time job and cost factors have a negative impact on retention (Grau-Valldosera et al., 2019). Specifically, students requiring financial aid to re-enrol show higher dropout (Cochran et al., 2014).

Few references can be found to the impact of disability. However, in a few studies, disability is cited by some students as a reason for withdrawal (Shah & Cheng, 2019).

Although several research works have used the OULAD dataset, none has been found covering demographics' role in withdrawal. The closest analysis found (Rizvi et al., 2019) considered the impact of these factors on academic outcomes in terms of pass-fail. This study reported that region, neighbourhood poverty level, and prior education constitute strong predictors of failure.

Research Questions

Considering the lack of studies that analyse withdrawal at the course level in regular OHE with a longitudinal approach, the relevance of reliable identification of learners at risk, and the potential of survival analysis, we formulated this research question:

RQ1: How can survival analysis be used to identify populations of learners at risk of withdrawal at the course level, providing insight into the factors behind that withdrawal?

Additionally, considering both the relevance of time and the potential impact of demographics on withdrawal, we posed a second research question, addressing practical implementation:

RQ2: What is the specific impact of demographic factors over time on course withdrawal? Which of these factors impact the withdrawal regardless of the course itself?

Specifically, we decided to analyse the impact of these demographic characteristics based on the OULAD dataset: (a) age, (b) gender, (c) disability, (d) region, (e) previous academic background, and (f) student's economic situation.

As mentioned, the OULAD dataset includes data from 22 editions of 6 different courses. Detailed information on the dataset is provided in the next section.

Method

Survival Analysis

Survival analysis is "a collection of statistical procedures for data analysis where the outcome variable of interest is time until an event occurs" (Clark et al., 2003a, p. 237). The method is particularly used in medical research, where survival time or time to relapse is under consideration (Bradburn et al., 2003a, 2003b; Clark et al., 2003a, 2003b). The portability of the method to other disciplines has been suggested in recent studies (Emmert-Streib & Dehmer, 2019).

Kaplan-Meier (KM) estimates and, specifically, KM curves are common in most survival analyses when the goal is to compare two populations. They are the simplest way to compute survival over time (Clark et al., 2003a). KM estimates help to establish whether life expectancy is different for different populations who have different characteristics, or whether a specific treatment can be more advisable than others. Linked to this estimation, the hazard function indicates the probability of not surviving beyond a certain point in time.

The statistical significance of the resulting curves can be checked with the *log*-rank test (Clark et al., 2003a). This test compares the estimates of the hazard functions of the two groups at each observed event time under the null hypothesis that both groups share the same hazard functions. The original test assigns equal weight to early and late events. Modified versions use weighted functions. In particular, Peto-Peto's *log*-rank test (Peto & Peto, 1972) assigns weights depending on the estimated percentile of the failure time distribution, giving higher weight to earlier events, and is commonly used within this group.

However, KM estimates cannot quantify the impact of a given parameter, particularly when dealing with different variables, i.e., the covariates. When this is required, parametric methods must be used. Fully parametric methods need to assume statistical distribution in the data. If this distribution is known, they can provide more precise models. Semi-parametric methods have the advantage of being able to quantify the impact without assuming a specific distribution. The most used semi-parametric method is the Cox proportional hazards model (Bradburn et al., 2003b). This model is based on a proportional hazard assumption and computes a baseline time-dependent hazard associated with a reference group. This hazard is modified based on the multiplicative effect of the values of the different covariates, whose individual influence is considered constant over time. Once the method is computed, the assumptions need to be checked.

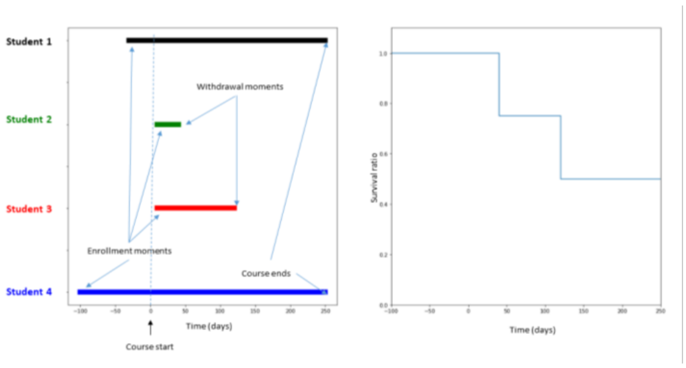
Porting Survival Analysis to Withdrawal Analysis

Approaching a generic problem through survival analysis requires a precise mapping of three concepts: the lifespan, the event under consideration, and the period of observation (Clark et al., 2003a).

In the case of withdrawal, the number of days a student remains enrolled after the specific course starts constitutes the lifespan. The event under consideration is the withdrawal decision. The analysis would also need to monitor on a periodical basis whether the student has withdrawn. To set up a common reference among courses, the course start date would be considered as t = 0. Negative values indicate days before the course starts. Survival curves reflect how a population survives after a certain time. Figure 1 depicts these concepts in a hypothetical course lasting 250 days with four students enrolled.

Figure 1

Graphical View of a Hypothetical Course and Associated Survival Curve



Note. Left panel: Enrolment and withdrawal or completion dates for four students. Right panel: Associated survival curve for this group.

On the left, Figure 1 shows four students enrolling on different dates. The first is Student 4, enrolling 100 days before the course starts. Students 2 and 3 enrol ten days after the course has started. In this example, Student 2 withdraws shortly after enrolment (40 days after the course starts), while Student 3 withdraws 120 days after the course starts. Students 1 and 4 complete the course. The associated survival curve for this group is shown on the right, where we can see that the final survival ratio is 0.5 (2 out of 4 students). The curve provides not only the final ratio but a graphical view of its evolution.

KM plots provide a graphical view of the individual impact of specific covariates. To aggregate and quantify the impact of those found relevant, we used the Cox proportional hazards model, due to its simplicity compared to parametric methods.

Dataset

These concepts were translated into practice using the public dataset offered by The Open University (OU; Kuzilek et al., 2017). This dataset provides information about 22 editions (*presentations* in the dataset nomenclature). A total of 32,593 students are enrolled in these courses. The typical presentation length is around nine months.

Courses included in the dataset were offered via a virtual learning environment (VLE), and each had over 500 students. While part of the OU course portfolio, students without a previous academic background could also enrol. Table 1 summarises a high-level view of enrolment and academic results in the courses included. Academic results are summarised in four categories: withdraw, fail, pass, or distinction.

Table 1

Global View of Enrolment and Academic Results

Indicator	Enrolled	Withdraw	Fail	Pass	Distinction
Number of students	32,593	10,156	7,052	12,361	3,024
Percentage of total (%)	100	31.16	21.64	37.93	9.28

As Table 1 shows, withdrawal constituted 31.16% of the global population enrolled. The distribution of academic results was not homogenous among courses as displayed in Table 2.

Table 2

Enrolment and Academic Results (Per-Course View)

Course	Students	Withdraw	Fail	Pass	Distinction
	n	%	%	%	%
AAA	747	16.73	12.18	65.19	5.89
BBB	7,903	30.18	22.32	38.93	8.57
CCC	4,434	44.54	17.61	26.61	11.23
DDD	6,266	35.86	22.49	35.54	6.11
EEE	2,934	24.61	19.15	44.10	12.13
FFF	7,758	30.96	22.02	38.39	8.64
GGG	2,534	11.52	28.73	44.12	15.63

Note. Courses are identified with anonymised course names (i.e. AAA) in the OULAD dataset.

These high withdrawal ratios may be explained by the fact that they constitute regular OU courses, with high academic standards, but at the same time, require no prior qualification for enrolment. All courses share a common framework for evaluation, including a set of tutor-marked assignments and optionally some computer-marked assignments. Also, there is usually a final exam at the end of each course.

With respect to those students withdrawing, the dataset includes information regarding the date of withdrawal. This date is either the date on which the student notified the university of her withdrawal or the date on which the student's participation in the module ceased, whichever came first. The Open University actively seeks to reduce withdrawal and may monitor *online* student activity to detect it. Students considering withdrawal are advised to contact the module instructor and, if their decision is final, formally report their decision (Open University, 2022).

The dataset also includes some personal information. Table 3 summarises those characteristics in the dataset considered relevant to our study.

Table 3

Characteristics in the OU Dataset Linked to the Research Questions

Scope	Variable	Meaning
Presentation	length	Length in days of the module presentation
Registration	date_registration	The day the student registers for the module presentation
	date_unregistration	The day the student unregisters from the module presentation
Demographic characteristic	gender	Gender of the student (male/female)
	region	The geographic region, where the student lived while taking the presentation
	imd_band	The index of multiple deprivation (IMD) band of the place where the student lived during the module presentation
	highest_education	The highest student education level on entry to the module presentation (5 bands)
	age_band	Age band of the student (3 bands)
	disability	Indicates whether the student has declared a disability

Note. Variable names used match those in the OULAD dataset.

This information was required to approach RQ2. Age, gender, and disability are available directly in the dataset. Previous academic background is expressed as the highest educational level the student achieved before the module started. Region indicates the area where the student lives. Student economic situation is expressed by the index of multiple deprivation (IMD) used in the UK (Kuzilek et al., 2017; Rizvi et al., 2019). The dataset presents IMD figures in bands ranging from 0%-10% to 90%-100%; 0%-10% means that a student lives in the most deprived UK areas, while 90%-100% points to the least deprived areas.

Results

The preceding section identifies two main steps for practical implementation:

1. Use KM estimates to determine populations at risk and the impact of individual covariates on withdrawal.

2. Analyse the combined impact, quantifying the simultaneous effect through Cox proportional hazards model.

The Cox model requires a prior setup of reference values for the covariates. For the categorical variables shown in Table 3, we generated dummy variables and considered the values shown in Table 4 as reference values.

Table 4

Reference Groups for the Computation of the Cox Model

Covariate	Reference group
Gender	Female
Region	North region
Highest education	A level or equivalent
IMD band	50%–60%
Age band	Under 35
Disability	No

Note. IMD = index of multiple deprivation.

When the number of possible values was high, we selected reference values that reflected a more central position (e.g., IMD band = 50%–60%). For the specific IMD scales, we grouped low IMD scales (0%–30%) and high IMD scales (above 80%) to reduce the overall number of values.

Significant Differences Based on IMD Band, Prior Education, and Declared Disability

Covariates to perform KM estimates were extracted from Table 3. Using Peto-Peto *log*-rank tests, we computed *p*-values. Data in Table 2 reflect that different courses show differences in withdrawal ratios. For this reason, we also performed a per-course analysis to determine whether covariates were significant both at the global and individual course levels. The results are shown in Table 5.

Table 5

Statistical Significance of Covariates at the Global and Individual Course Levels

Covariate	Global	Individual course						
		AAA	BBB	CCC	DDD	EEE	FFF	GGG
Gender	ns	*	ns	ns	***	*	*	ns
Region	****	ns	****	ns	ns	**	ns	ns
Highest education	****	ns	****	****	****	***	****	ns
IMD band	****	ns	****	****	****	**	****	ns
Age band	****	ns	**	**	ns	ns	ns	ns
Disability	****	ns	ns	***	****	ns	****	*

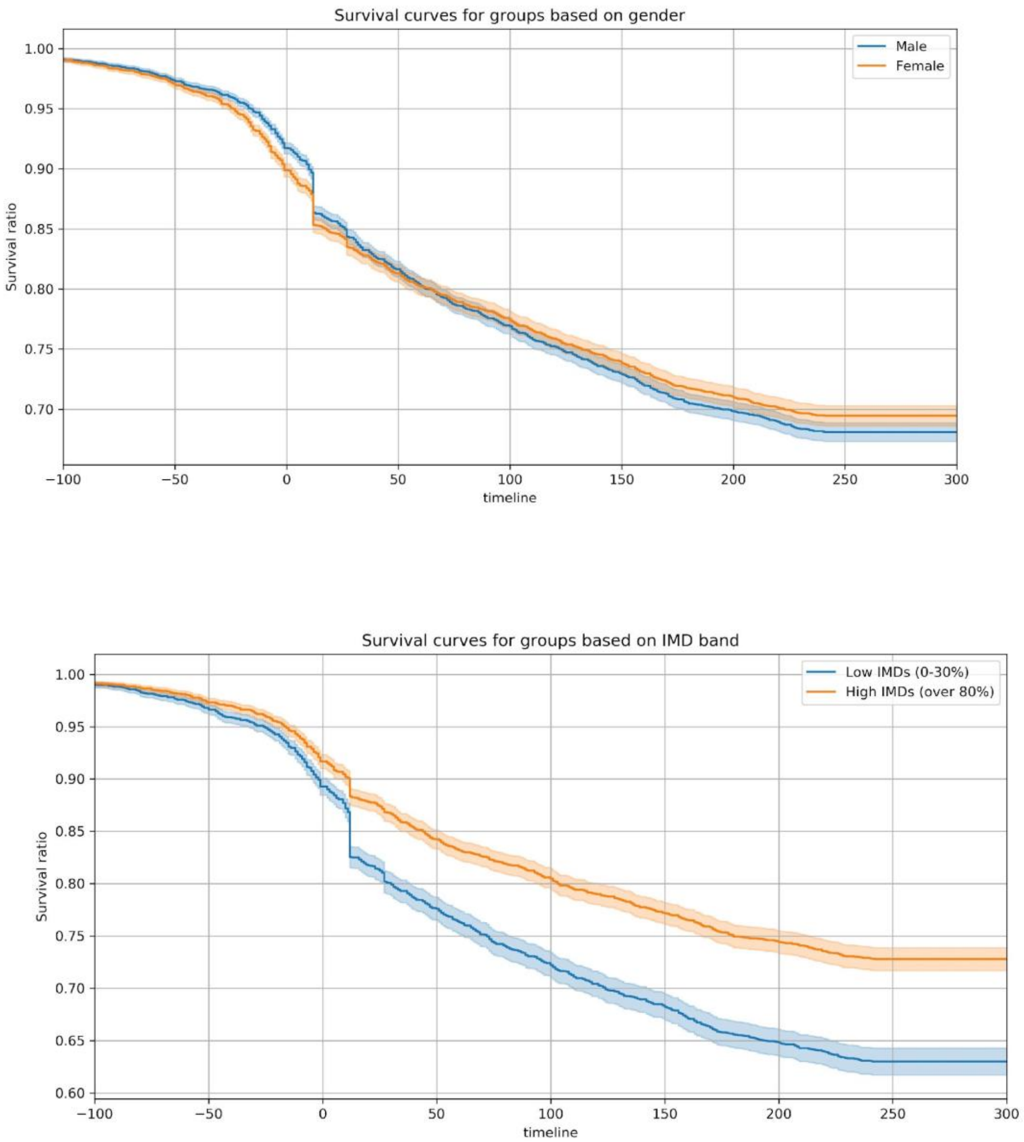
Note: ns = non-significant. IMD = index of multiple deprivation. Courses are identified with anonymised course names (i.e. AAA) in the OULAD dataset. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. **** $p < 0.0001$.

Prior highest education level and IMD band have a clear impact when considering either the global data set or individual courses. At the course level, more data would be needed for course AAA to provide statistically significant results. While age looks relevant globally, its effect disappears in most courses when analysed individually. Thus, no definite conclusion can be extracted at this stage. More data would also be needed, as there is a low ratio of students in one of the scales considered in the dataset.

KM plots help to visualise differences. As an example, we show the global impact of two covariates: gender—not significant according to the test—and the IMD band, which is significant. For clarity, in the case of IMD plots, we compared the high (> 80%) and low (< 30%) groups. The results are shown in Figure 2.

Figure 2

Survival Curves for Different Groups Based on Gender and IMD Band



Note. Top panel: The survival curve for gender. Bottom panel: The survival curve for IMD band. IMD = index of multiple deprivation.

Figure 2 shows minor differences based on gender. Regarding IMD bands, this figure reflects higher withdrawal ratios for the low IMD group, with a higher impact of early withdrawal.

Previous Risk Factors also Present when Considering Simultaneous Effect

We used the Cox model to evaluate and quantify the simultaneous effect of the different covariates. The final Cox models were developed with two strata variables (course and disability) and a set of dummy variables linked to IMD band, region, gender, and previous higher education. Table 6 summarises those variables that appear relevant at either the global or individual course level.

Table 6

Hazard Risk Factor Relative to the Reference Group Based on the Values of Covariates

Covariate	Individual course							Global
	AAA	BBB	CCC	DDD	EEE	FFF	GGG	
Gender: Male	ns	ns	ns	0.83	ns	0.90	ns	0.89
Region: East Midlands	ns	ns	1.23	ns	ns	ns	ns	1.14
Region: London	ns	ns	ns	ns	1.41	1.20	ns	ns
Region: West Midlands	ns	ns	ns	ns	1.39	ns	ns	1.12
Highest education: HE qualification	ns	ns	0.83	ns	ns	ns	ns	0.93
Highest education: Lower than A Level	ns	1.41	1.41	1.30	1.48	1.42	ns	1.38
Highest education: No formal qualifications	ns	1.73	1.48	1.72	2.38	1.38	ns	1.63
Highest education: Post- graduate qualification	ns	ns	0.56	ns	ns	ns	5.51	0.75
IMD band: 0%–30%	ns	1.18	1.33	1.37	1.56	1.44	ns	1.35

IMD band: 30%–40%	ns	ns	ns	1.25	ns	ns	ns	1.14
IMD band: 40%–50%	ns	ns	1.21	1.35	1.75	ns	ns	1.21
IMD band: 80%–100%	ns	ns	ns	ns	ns	ns	1.69	ns

Note: Only covariates significant in at least one course shown. IMD = index of multiple deprivation.

To understand the impact of disability, we compared baseline survival functions for the different strata generated at the course level. Table 7 reflects the withdrawal increase ratio for individual courses when disability was a factor.

Table 7

Impact of Disability on Withdrawal Risk—Individual Course Level

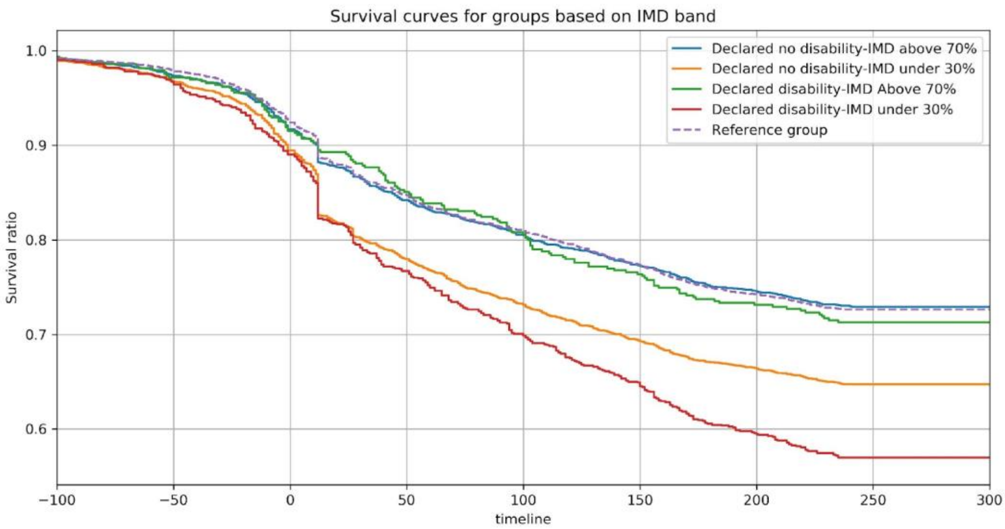
Indicator	Individual course					
	BBB	CCC	DDD	EEE	FFF	GGG
Withdrawal risk increase (declared disability vs declared no disability)	1.14	1.19	1.49	0.99	1.43	1.45

Note. Courses are identified with anonymised course names (i.e. BBB) in the OULAD dataset. Course AAA showed inconclusive results and was omitted from this table.

As a final check, we performed a graphical comparison of withdrawal differences based on the findings above. We generated populations based on the combination of IMD differences—high versus low group—and disability. The results are shown in Figure 3, where a reference group based on data in Table 4 (no disability, IMD band 50%–60%) is also reflected.

Figure 3

Survival Curves for Groups With and Without Declared Disability in High and Low IMD Bands



Note: IMD = index of multiple deprivation.

Figure 3 clearly shows withdrawal risk differences among groups. Besides final survival expectancy—with differences around 35.88% by the end of the course—low-income students drop out earlier. Also, the multiplicative effect of disability and low IMD is clearly displayed. Being in the high IMD group does not significantly reduce withdrawal rates when compared to the reference group. The impact of these findings on potential intervention designs will be addressed in the next section.

Discussion

Two research questions were addressed in this work. The first one, regarding the use of survival analysis, aimed to detect populations at risk of withdrawal and the factors behind it. The second one aimed to translate these concepts into practice, determining the relevance and impact of demographics.

From a methodological perspective, the basics behind the answer to RQ1 are covered in the subsection covering the portability of survival analysis to *learning analytics* scenarios. Identifying students at risk of withdrawal through survival analysis has required the mapping of three concepts: the event under consideration (the withdrawal decision), the period of observation (a course), and the lifespan (the time the student remains in the course). This mapping allows us to identify both at-risk populations and the associated risk factors. Figure 1 concentrates on the basics behind this mapping.

Survival curves provide a graphical insight into the differences among populations based on a set of factors (Figures 2 and 3 offer clear examples). These curves constitute a relevant difference from other data mining techniques. They do not only provide information on final withdrawal ratios, but also show when withdrawal occurs. Statistical relevance of a given factor can also be determined (see Table 5 for examples), and for those factors considered relevant, the impact can also be quantified (Tables 6 and 7 serve as examples for this point). These facts make survival analysis a particularly suitable technique for analysing withdrawal.

All in all, figures 1 (from a theoretical perspective), 2, and 3 (from a practical approach), combined with the data in tables 6 and 7, demonstrate the potential of survival analysis identifying populations of learners at risk.

The second question (RQ2) translates methodology into practice. The application of the method indicates that certain demographic characteristics have an impact on course withdrawal and that this impact is dependent on the course itself. Specifically, three analysed factors increase withdrawal risk: a previous level of education below the reference group (A level), a low IMD band, and a declared disability (see Table 5). As mentioned, the specific impact is different depending on the course (see Table 7). We can compare these findings with previous literature regarding the impact of demographics on withdrawal.

Withdrawal and Demographics: Comparison with the Literature

From a global perspective, the influence of these personal background factors is consistent with the theoretical models (Bean, 1985; Cabrera et al., 1992; Rovai, 2003; Tinto, 1975) and justifies the interest that literature compilations put on them, in particular in OHE (recently, Hachey et al., 2022; Muljana & Luo, 2019).

Our results have shown a different impact for age and gender across the analysed courses, supporting the inconclusive findings reflected in Hachey et al., 2022 and Muljana and Luo, 2019. We have not found that being male reduces risk in some courses, while increasing it in others (as found in Cochran et al., 2014). However, we agree that the relevance and the specific impact of a factor depend on the course under analysis.

Regarding the economic situation, our results at course level are aligned with those indicating the impact of financial hardship at course and university level (Cochran et al., 2014; Grau-Valldosera et al., 2019). Our work indicates a direct relationship between socioeconomic inequality and educational disadvantage, as shown in the lower panel in Figure 2.

The impact of a poor academic background on withdrawal is consistent with earlier research linking lower previous achievement to higher university dropout (Cochran et al., 2014; Lee & Choi, 2011).

Disability is one of the potential reasons behind some dropouts according to Shah and Cheng (2019). Our results confirm this fact and quantify its impact on course withdrawal. Our findings indicate that students with disabilities taking *online* courses would be more likely to withdraw from these courses, in particular those students in low IMD bands. Due to our concerns about equity, we believe more studies on this topic should take specific care of anonymisation and ethical issues.

Implications

The intersectionality and the reliable estimation of risk allow us to identify two points for a potential intervention with targeted populations. First, less affluent students could be contacted, even before the start of the course, and offered options regarding financing. Second, disabled students coming from more deprived areas might benefit from continuous support, which might reduce the slowly increasing difference in withdrawal rates when compared with non-disabled students reporting the same economic condition.

While these demographic factors affect all courses analysed, their impact on dropout is different for each course. This difference needs to be considered when evaluating the outcomes of specific interventions.

We can also find factors that show statistical significance in only some courses. To mention just a couple of examples, gender for course DDD or region for course BBB (see Table 5) warrant investigation. For these cases, we encourage a closer look that considers course-specific details which may explain why.

We also remark on the potential of survival analysis to detect situations that would otherwise remain hidden. Figure 2 (lower panel) and Figure 3 reveal a sudden drop around the second week of the course, particularly affecting low-income students. In fact, this week corresponds to the end of the full-refund period for a given course. A potential intervention aimed at reducing withdrawal would consider extending the period of full refund for low-income students. It is important to highlight that this kind of finding would remain hidden if using techniques which focus only on final ratios and not on temporal evolution.

Our detailed analysis also reveals potential fails in intervention design which do not include a proper identification stage. We can consider for instance prior level of education. It is noticeable that course GGG shows a higher risk of withdrawal for students with a previous higher level of education. While this could be shocking at first glance, this course constitutes a propaedeutic course. Those students who already have this knowledge simply abandon the course. Besides this example, and considering potential interventions, the method reveals that analysis at both a global level and at the level of the individual course is critical to properly identify populations at risk.

Finally, survival analysis provides a clear view of the impact of the different factors analysed. For the case of demographics, IMD band, prior educational level, and declared disability have emerged as the most relevant factors in dropout. It is worth noting that these factors emerge as relevant both at the aggregate level and when considering individual courses.

Conclusion

Survival analysis has proven to be a useful tool to reliably identify populations of learners at risk. The method outlined provides risk quantification, and a clear graphical evolutive view. This view highlights insights that otherwise could remain hidden.

Its use has been particularly suited for the analysis of course withdrawal, due to the relevance of time in dropout. We encourage the use of survival analysis as the first stage in the design of interventions aimed at reducing academic dropout. It can also be of interest in *learning analytics* scenarios where time plays an important role, such as engagement analysis.

Finally, considering the multivariate nature of withdrawal, we advise expanding this research beyond demographics. While focusing on them has shown the method's potential and provided valuable insight, it also constitutes a limiting factor. We encourage the use of the methodology exposed to address the impact of other aspects, such as previous knowledge, activity reflected in VLEs, or course instructional design. Future work should focus on incorporating these dimensions into the analysis to better understand students' behaviour and improve learning experience and academic performance.

Acknowledgements

We would like to thank The Open University for working on the creation of the dataset, and in particular, Professor Bart Rienties from the Institute of Educational Technology for his readiness to provide additional insight to better understand the dataset. Also, Professor Per Bergamin from the Swiss Distance University of Applied Sciences offered insightful feedback during the review process.

References

Aljohani, O. (2016). A comprehensive review of the major studies and theoretical models of student retention in higher education. *Higher Education Studies*, *6*(2), 1–18. https://doi.org/10.5539/hes.v6n2p1

Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016). Survival analysis based framework for early prediction of student dropouts. In S. Mukhopadhyay & C. X. Zhai (Chairs), *CIKM '16: Proceedings of the International Conference on Information and Knowledge Management* (pp. 903–912). ACM. https://doi.org/10.1145/2983323.2983351

Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—A literature review. *SAGE Open*, *6*(1), 1–11. https://doi.org/10.1177/2158244015621777

Bean, J. P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American Educational Research Journal*, *22*(1), 35–64. https://doi.org/10.3102/00028312022001035

Behr, A., Giese, M., Teguim Kamdjou, H. D., & Theune, K. (2020). Dropping out of university: A literature review. *Review of Education*, *8*(2), 614–652. https://doi.org/10.1002/rev3.3202

Berge, Z. L., & Huang, Y.-P. (2004). A model for sustainable student retention: A holistic perspective on the student dropout problem with special attention to e-Learning. *Deosnews*, *13*(5), 1–26.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003a). Survival analysis part II: Multivariate data analysis—An introduction to concepts and methods. *British Journal of Cancer*, *89*, 431–436. https://doi.org/10.1038/sj.bjc.6601119

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003b). Survival analysis part III: Multivariate data analysis—Choosing a model and assessing its adequacy and fit. *British Journal of Cancer*, *89*, 605–611. https://doi.org/10.1038/sj.bjc.6601120

Cabrera, A. F., Castañeda, M. B., Nora, A., & Hengstler, D. (1992). The convergence between two theories of college persistence. *The Journal of Higher Education*, *63*(2), 143–164. https://doi.org/10.2307/1982157

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003a). Survival analysis part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*, 232–238. https://doi.org/10.1038/sj.bjc.6601118

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003b). Survival analysis part IV: Further concepts and methods in survival analysis. *British Journal of Cancer*, *89*, 781–786. https://doi.org/10.1038/sj.bjc.6601117

Cobre, J., Tortorelli, F. A. C., & de Oliveira, S. C. (2019). Modelling two types of heterogeneity in the analysis of student success. *Journal of Applied Statistics*, *46*(14), 2527–2539. https://doi.org/10.1080/02664763.2019.1601164

Cochran, J. D., Campbell, S. M., Baker, H. M., & Leeds, E. M. (2014). The role of student characteristics in predicting retention in online courses. *Research in Higher Education*, *55*(1), 27–48. https://doi.org/10.1007/s11162-013-9305-8

Emmert-Streib, F., & Dehmer, M. (2019). Introduction to survival analysis in practice. *Machine Learning and Knowledge Extraction*, *1*(3), 1013–1038. https://doi.org/10.3390/make1030058

Grau-Valldosera, J., Minguillón, J., & Blasco-Moreno, A. (2019). Returning after taking a break in online distance higher education: From intention to effective re-enrollment. *Interactive Learning Environments*, *27*(3), 307–323. https://doi.org/10.1080/10494820.2018.1470986

Hachey, A. C., Conway, K. M., Wladis, C., & Karim, S. (2022). Post-secondary online learning in the U.S.: An integrative review of the literature on undergraduate student characteristics. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-022-09319-0

Hachey, A. C., Wladis, C. W., & Conway, K. M. (2014). Do prior online course outcomes provide more information than G.P.A. alone in predicting subsequent online course grades and retention? An observational study at an urban community college. *Computers and Education*, *72*, 59–67. https://doi.org/10.1016/j.compedu.2013.10.012

James, S., Swan, K., & Daston, C. (2016). Retention, progression and the taking of online courses. *Journal of Asynchronous Learning Network*, *20*(2), 75–96. https://doi.org/10.24059/olj.v20i2.780

Kuzilek, J., Hlosta, M., & Zdrahal,Z. (2017). Open University Learning Analytics dataset. *Scientific Data*, *4*(1), 170171. https://doi.org/10.1038/sdata.2017.171

Larsen, M. S., Kornbeck, K. P., Kristensen, R. M., Larsen, M. R., & Sommersel, H. B. (2013). *Dropout phenomena at universities: What is dropout? Why does dropout occur? What can be done by the universities to prevent or reduce it?: A systematic review*. Danish Clearinghouse for Educational Research.

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, *59*(5), 593–618. https://doi.org/10.1007/s11423-010-9177-y

Lee, Y., & Choi, J. (2013). Discriminating factors between completers of and dropouts from online learning courses. *British Journal of Educational Technology*, *44*(2), 328–337. https://doi.org/10.1111/j.1467-8535.2012.01306.x

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2019). Prediction in MOOCs: A review and future research directions. *IEEE Transactions on Learning Technologies*, *12*(3), 384–401. https://doi.org/10.1109/TLT.2018.2856808

Muljana, P. S., & Luo, T. (2019). Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review. *Journal of Information Technology Education: Research*, *18*, 19–57. https://doi.org/10.28945/4182

Open University, The. (2022). *Changing your study plans policy 2022/23*. https://help.open.ac.uk/documents/policies/changing-your-study-plans/files/214/changing-your-study-plans-policy-2022-23.pdf

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, *135*(2), 185–207. https://doi.org/10.2307/2344317

Radovanović, S., Delibašić, B., & Suknović, M. (2021). Predicting dropout in online learning environments. *Computer Science and Information Systems*, *18*(3), 957–978. https://doi.org/10.2298/csis200920053r

Rienties, B., Boroowa, A., Cross, S., Farrington-Flint, L., Herodotou, C., Prescott, L., Mayles, K., Olney, T., Toetenel, L., & Woodthorpe, J. (2016). Reviewing three case-studies of learning analytics interventions at the Open University UK. In D. Gašević & G. Lynch (Chairs), *LAK '16: Proceedings of the sixth international conference on learning analytics and knowledge* (pp. 534–535). https://doi.org/10.1145/2883851.2883886

Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning: A decision tree based approach. *Computers and Education*, *137*(2), 32–47. https://doi.org/10.1016/j.compedu.2019.04.001

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2022). Beyond one-size-fits-all in MOOCs: Variation in learning design and persistence of learners in different cultural and socioeconomic contexts. *Computers in Human Behavior*, *126*(C), Article 106973. https://doi.org/10.1016/j.chb.2021.106973

Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *Internet and Higher Education*, *6*(1), 1–16. https://doi.org/10.1016/S1096-7516(02)00158-6

Shah, M., & Cheng, M. (2019). Exploring factors impacting student engagement in open access courses. *Open Learning*, *34*(2), 187–202. https://doi.org/10.1080/02680513.2018.1508337

Simpson, O. (2010). *"22%—can we do better?": The CWP retention literature review*. Centre for Widening Participation, Open University UK. https://doi.org/10.13140/RG.2.2.15450.16329

Strang, K. D. (2017). Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes? *Education and Information Technologies*, *22*(3), 917–937. https://doi.org/10.1007/s10639-016-9464-2

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89–125. https://doi.org/10.3102/00346543045001089

Villano, R., Harrison, S., Lynch, G., & Chen, G. (2018). Linking early alert systems and student retention: A survival analysis approach. *Higher Education*, *76*(4), 903–920. https://doi.org/10.1007/s10734-018-0249-y

Xavier, M., & Meneses, J. (2020). *Dropout in online higher education: A scoping review from 2014 to 2018*. Universitat Oberta de Catalunya. https://doi.org/10.7238/uoc.dropout.factors.2020

Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, *58*, 119–129. https://doi.org/10.1016/j.chb.2015.12.007

Xing, W., Tang, H., & Pei, B. (2019). Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. *Internet and Higher Education*, *43*, Article 100690. https://doi.org/10.1016/j.iheduc.2019.100690

Artículo 5: Impact of assessment characteristics on course dropourevit: evidence gathered through survival analysis

Martínez-Carrascal, J. A., Sancho-Vinuesa, T. Currently being published as part of the LASI Spain 2023 Proceedings.

Nota: Dado que el artículo se encuentra en fase de publicación definitiva, se presenta seguidamente una versión preliminar. Indicar también que las referencias numéricas a figuras y tablas contenidas en el artículo que se presenta a continuación deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

Impact of assessment characteristics on course dropout: evidence gathered through survival analysis

One of the measures used while analysing teaching quality is dropout rate. It has a significant effect on higher education institutions, in particular, those offering *online* studies. This fact justifies the pursuit of a deeper comprehension of the phenomenon in order to mitigate its detrimental effects.

This study focuses on the impact that assessment characteristics can have on course dropout. The analysis is carried out from a methodological perspective and translated into a real case-study. Survival analysis will be used as main tool to explore the impact of the characteristics under consideration. In addition, the impact of time between assessments and the type of tests carried out will be analysed based on an open dataset provided by the Open University.

The results show that increasing the maximum time between assessments impacts the withdrawal decision, leading to an increase in dropout ratios. As an additional result, we cannot draw decisive conclusions about the number and typology of assessments. In addition, the method permits a straightforward extension to analyse the impact of other assessment criteria that may influence course dropout.

Keywords: withdrawal; dropout risk; assessment characteristics; survival analysis.

1 Introduction

Withdrawal, understood as the decision to abandon an enrolled course, is one of the biggest problems of the educational system at all levels. Al the highest, it involves a waste of resources – time and money - (Aulck et al., 2016). From the student perspective, it causes frustration regarding her expectations (Lee & Choi, 2013). The issue is complex and unresolved. It is especially prevalent in higher education, prompting policymakers and universities to seek ways to decrease it.

Diverse factors influence the decision to withdraw (Aljohani, 2016). Some of these factors relate to the unique circumstances of the student. Other elements, however, are under the educational institution's control. Teaching-related factors are acknowledged as an influential factor. Disengagement-inducing factors would raise dropout rates. Given that assessment characteristics, specifically assessment frequency, difficulty, and diversity, have a moderating effect on student involvement, they would likewise influence dropout.

Our research focuses on determining the impact of assessments characteristics on dropout, uncovering which strategies can be more convenient in terms of its reduction. Further than providing specific findings for a given set of characteristics, we provide a method that can be used to explore other potentially relevant aspects. Assessments are a nuclear part of the learning process, and a proper design contributes to improve learning (Gibbs et al., 2003). Its link to academic achievement has also been reported in different studies (Elton & Laurillard, 1979; Gibbs, 1999; Gibbs et al., 2003). However, no precise correlation has been demonstrated with academic withdrawal. The literature suggests a connection between assessment design and student engagement, but neither a quantification nor a direct connection to academic dropout.

The methodological approach will be based on the use of survival analysis and translated into practice considering an open dataset, provided by the Open University (Kuzilek et al., 2017b). It includes data from over 30,000 students enrolled in seven distinct higher education *online* courses, totaling 22 course editions. In various versions of particular courses, several assessment strategies are utilised. Specifically, the use of different time between assessment and computer marked assessments (CMA) as a component of the learning process. We will analyse how variations in these assessment parameters impact withdrawal rates for a specific course. This dataset has been considered for use since it is an open dataset, permitting replication and even expansion of the current research. In addition, due to the Open University's active policy to reduce withdrawal, the time of student withdrawal is also recorded.

To the best of our knowledge, no previous studies have performed this kind of analysis to explore the influence of assessment characteristics in withdrawal at course level. The use of survival analysis is common in medicine (Clark et al., 2003a), to estimate life expectancy of a patient, or how different treatments can influence this expectancy. While different studies using this technique are also found on *learning analytics* scenarios (Ameri et al., 2016; Cobre et al., 2019; Spronken-Smith et al., 2018; Villano et al., 2018), its use to analyse withdrawal at course level is mainly limited to MOOC environments (Moreno-Marcos et al., 2019; Rizvi et al., 2022; Xing et al., 2019) or specific programs (Spronken-Smith et al., 2018) and do not consider the specific impact of assessment characteristics. Beyond the specific findings linked to the courses analysed, the method can be particularly effective for determining the influence of different evaluation strategies on withdrawal, so assisting in the improvement of teaching quality by identifying those assessment characteristics that perform the best in terms of reducing course dropout.

2 Theoretical framework

2.1 Assessment characteristics and academic performance

Assessment is defined in The Glossary of Education Reform as ‘the wide variety of methods or tools that educators use to evaluate, measure, and document the academic readiness, learning progress, skill acquisition, or educational needs of students’ (*Assessment Definition*, n.d.). Its relevant impact in the learning process justifies the broad research on the topic. A compilation of works together with their structuration into different research lines can be found in (Pereira et al., 2016).

The influence of assessment in academic performance was already present in early studies on the topic (Elton & Laurillard, 1979) and confirmed in more studies, either considering direct influence or moderating effects. (Day et al., 2018a; Wang & Zhang, 2020). Some authors consider the impact of assessment even greater than that of teaching (Miller &

Parlett, 1974). (Gibbs, 1999) outlined the impact of assessment in learning outcomes. Years later, (Gibbs et al., 2003) conclude that ‘there is more leverage to improve teaching through assessment than through anything else’. This later study proposed a practical framework detailing an assessment framework with specific tactics.

Maximizing the potential of assessment requires a thoughtful exercise (Bloxham & Boyd, 2007). Assessment strategy includes the definition of a clear purpose for each assessment item, the context of the assessment ant a detailed view of the tasks to develop and demonstrate learning. It has a major impact on academic performance and its implementation provides indicators of teaching and learning effectiveness and overall quality (*Standards and Guidelines for Quality Assurance in the European Higher Education Area*, n.d.)..

Focusing at course level, (Simpson, 2016) indicates that there is evidence in the literature showing differences in patterns of assessment and mark distribution among different disciplines. This difference does not necessarily translate to the degree level, where the discipline itself can be of greater influence. (Lynam & Cachia, 2018) focus on the specific impact of assessment characteristics that impact student learning. (Day et al., 2018a) compile previous research and conclude that assessments characteristics act combined to influence student’s grades. They focus on intermediate assessments and reveal the relevance of both mandatory and voluntary assessments, the potential benefits of peer assessments, and the effects of assessment rewards.

Differences between continuous and final exam-based evaluation has also been analysed (Richardson, 2015). Marks are reported to be higher in environments that combine coursework and examinations when compared to those that use examinations alone. Continuous evaluation provides also higher attainment (Ellison & Jones, 2019).

Focusing on continuous evaluation assessments, differences focused on assessment types are analysed in (Day et al., 2018b). As a relevant finding, the article notices that the type of assessment may influence in non-continuous scenarios, and even depending on personal characteristics such as gender. However, this influence disappears in continuous evaluation models. No difference in results is detected based on whether assessments are mandatory or not (Day et al., 2018a).

The perception of the student regarding the level of difficulty is also found relevant. (Preston et al., 2020) indicate that the performance of the student mirrors her perception of the level of assessment difficulty. Assessments which require low effort, or which are based on predictable questions derive in low effort from the student, and translates into lower learning outcomes (Lynam & Cachia, 2018). From the student perspective, the same study indicates that learners appreciate assessments which are built on their skill set. Those fomenting creativity were preferred. The perception of student workload is also considered, indicating that students prefer an amount of ‘balanced’ workload.

Further that determining the impact, recent studies focus on its quantification. A recent study (Wang & Zhang, 2020) reveals the mediating effect and moderate impact of the specific assessment characteristics in overall performance. This study points out the moderating effect that frequency, difficulty, and diversity of assessment have on student performance.

While difficulty and diversity have to do with the type of assessment, frequency is linked to time**.** Timeliness refers to the importance of the timing of the delivery of assessment guidance and the timing of the assessments themselves (Lynam & Cachia, 2018). Most higher institutions implement evaluation models which include different assessments. The use of continuous summative assessment methods is time-consuming for teachers, but at the same time to provides benefits in terms of learning outcomes (Trotter, 2006).

(Gibbs et al., 2003) indicate the negative effect of a badly conceived strategy in terms of assessment tasks and frequency. Two of the authors performed further research (Gibbs & Simpson, 2004) showing that infrequent examinations tend to concentrate learning time and a more frequent assignment schedule is liked to more study time and better academic performance.

More recently, (Vandist et al., 2009) find improvements in academic performance in environments with a higher volume of assessments. The authors indicate that this improvement is linked to the higher frequency of feedback the student receives. (Day et al., 2018a) compile previous results and extend this finding, concluding that the proper number and frequency of assessments may depend on the course itself and also should take into account the duration of the course. However, the improvement in overall academic performance does not necessarily mean a reduction of dropout ratios (Foster & Francis, 2020) .

2.2 Specific impact on course withdrawal

The impact in performance abovementioned is related to the influence on final course marks. However, the number of studies drops when considering the link between assessment and course withdrawal

Course withdrawal can be defined as ceasing studying a course without the intention to resume the study of that course later(Open University, 2022). The concept is typically included within the broader term of dropout, which in higher education refers to 'situations where students leave the university study in which they have enrolled before they have obtained a formal degree' (Larsen et al., 2013).

Specific compilation of works analysing dropout in higher education (Behr et al., 2020) and specifically in *online* environments (Orellana et al., 2016) can be found. The studies collected reflect the relevance of dropout both at degree and course level, and also the higher impact in *online* institutions. It is also noticeable that early works already indicated that course design and characteristics are present in 28% of withdrawal cases (Aragon & Johnson, 2008). Assessment characteristics would be among those aspects, but are not explicitly analysed.

The potential link between assessment characteristics and dropout would be based on the mediating effect of assessment characteristics in student engagement reflected in (Wang & Zhang, 2020), and the demonstrated effect of engagement in dropout (Landis & Reschly, 2013). However, the link between engagement in broad terms and assessment characteristics would still be unclear. (Wang & Zhang, 2020) indicate a higher engagement when the frequency, difficulty or diversity of assessments is high. However, (Naude et al., 2016) indicate that higher workload is not necessarily linked to higher engagement.

As it can be seen, 'the evidence for what factors within assessments actually contribute to student engagement is not fully understood and more research is required' (Lynam & Cachia, 2018). This is particularly true when considering the impact on course dropout.

2.3 Research question

Based on this context, we investigate a methodological strategy for determining how and to what extent specific assessment characteristics influence course withdrawal decisions in higher education. To the best of our knowledge, no previous studies have analysed the effect of a variation in certain assessment characteristics across various editions of the same course on its dropout rates. The approach will be based on the use of survival analysis, as it has proven relevant in previous research (Author; Omitted during review process)

Besides the methodogological foundation, we explore some specific parameters. In particular, we hypothesise that leaving the students for a long time without assessments forces disengagement, and this disengagement translates into course withdrawal. For this reason we raise the following research question:

RQ1: How does the maximum time between assessments impact withdrawal decisions of students at course level?

The approach will be translated into practice through an open dataset provided by The Open University (Kuzilek et al., 2017a). In order to explore other aspects that might impact, the dataset includes information regarding the use of computer marked assessments (CMAs) through the course. We will use this information to validate the method on another assessment parameter by posing this question:

RQ2: Does the inclusion of CMAs affect the withdrawal decision? To what extent?

While time between assessments and the use of CMAs will be analysed, the method provided is extensible to any other assessment characteristics whose potential influence would be under consideration.

3 Methodology

To be able to answer the RQs, two factors are needed. First, the establishment of a methodology for analysing possible withdrawal disparities based on different evaluation variables. Second, a specific dataset containing a group of courses that may be compared with respect to various assessment features. These topics are covered in two separate subsections. A final subsection ties these two notions together by demonstrating how data is effectively modelled for our needs and is ready for analysis.

3.1 Using survival analysis to analyse the impact of assessment differences on course dropout

Different techniques are used to identify students at risk of dropping out (Behr et al., 2020; Xing et al., 2019). Recent research, however, encourages the use of longitudinal data techniques. Despite its potential, only a small number of studies use longitudinal data approaches. Some authors indicate their ability to outperform traditional prediction algorithms (Xing et al., 2016).

Among these techniques, survival analysis demonstrates particularly beneficial when addressing dropout. While most studies at course level focus on MOOC scenarios (Moreno-Marcos et al., 2019; Rizvi et al., 2022; Xing et al., 2019), recent research has proved its ability to identify at-risk populations and quantify the impact of differences among student's populations in regular higher education *online* courses (Author;Omitted during review process). We extend this approach to analyse the impact of differences based on differences in assessment characteristics among course editions.

The principles of survival analysis are outlined in (Bradburn et al., 2003a, 2003b; Clark et al., 2003a, 2003b). The method is particularly suited for situations when the relevant component is the time remaining until a specified event occurs. The strategy is typically employed in the medical field when examining the time to relapse or death as well as the significance and impact of various causes or treatments during this time.

These ideas are directly transferable to *learning analytics* scenarios when three key factors are taken into account: a precise definition of the event being studied (in this case, dropout), the observation period (which in this case will last from the time of enrolment to the end of the course), and survival as a whole. In the study of dropout, survival as a whole refers to the period of time during which the student remains engaged in academic life without formally resigning. Once this mapping is complete, we can evaluate the effects of various assessment tactics.

Kaplan-Meier estimates will reflect the impact of this potential effect (Clark et al., 2003a). They provide a graphical representation of the likelihood of surviving after a specific period, enabling comparison of the effects of a particular change. They provide a clear view of whether life expectancy – time to dropout in our case – varies depending on the assessment method or implementation. In addition to the graphical representation, the analytical formulation of the curve can be used to ascertain the influence of a particular element on the survival probability or how changing this factor, which is typically referred to as a covariate, affects the total survival probability.

3.2 Dataset

The reseach conducted is based on an open dataset provided by the Open Univesity (OU) (Kuzilek et al., 2017b). This dataset contains information about 22 editions, or presentations in OU terminology, of 7 distinct courses. Every course includes at least two editions. 32,593 students are enrolled in these courses, which the Open University refers to as modules. The average presentation duration for these modules is around nine months. As a remarkable aspect, the dataset includes, for each study, the date of withdrawal from the course in the eventuality of withdrawal.

All of the courses in the dataset are delivered through a virtual learning environment (VLE) and have more than 500 students. Courses have also been chosen based on the proportion of students who do not graduate the course, either by failing or dropping out. Table 1 provides a summary of academic performance in these courses in terms of pass, failure, and withdrawal.

Table 1. Global view of academic achievement in the dataset

Concept	Enrolled	Fail	Pass	Withdraw

# Students	32.593	7,052	15,385	10,156
% of total	100%	21.64%	47.20%	31.16%

As Table 1 shows, 31,16% of the students enrolled withdraw before the course ends. This population constitutes our analysis target. All courses share a common framework for evaluation, which includes a set of tutor marked assessments (TMA) and optionally some computer marked assessments (CMA). Besides, there is normally a final exam at the end of the course. Relevant variables related to assessment characteristics are summarized in Table 2.

Table 2. Characteristics under consideration in the OU dataset linked to assessment characteristics

Variable	Meaning
assessment_type	TMA, CMA, or Final Exam.
date	information about the cut-off day of the assessment
weight	weight of the assessment as part of a summative assessment evaluation

In addition to the information in Tables 1 and 2 above, we have course parameters such as presentation length. We also have data regarding the enrolment date of the student and, as indicated, the date when a student withdraws is also recorded for those students who decide to abandon the course. This point is particularly relevant for mapping the problem as a survival analysis scenario.

3.3 Linking survival analysis to a specific learning dataset

Once the concepts of lifespan (time to dropout), event under consideration (dropout as such) and the period of observation (course duration) are clear, we focus on the impact on this event derived from specific assessment characteristics. Data pre-processing has uncovered that certain students dropout even before the course begins. These students have not been considered for the analysis. The primary reason is that it may be challenging to connect this decision to assessment-related issues. As stated, our primary objective is to identify variations in the presentations that influence the decision to withdraw. The withdrawals prior to the start date may be attributable to other factors, or at the very least, the reasons may not be related to components of course instruction.

The withdrawal decision is recorded in the *final_result* field in the student record with the designation 'Withdrawn.' Regarding the observation period, we considered the beginning of the course to be the initial time (t=0). Prior to that time, dates (especially enrolment dates) will be reported in days and deemed negative. In terms of lifespan, our analysis will be valid so long as the presentation is active. Students with life expectancies shorter than the duration of the lecture are more likely to drop out.

Despite our emphasis on two specific parameters – time and the potential inclusion of CMAs – we also searched for assessment-related parameters that could contribute to statistically significant variations in withdrawal rates. Taking into account the dataset, and specifically the data presented in Table 2, we consider the following parameters as being of potential interest:

- Maximum time between assessments
- Type of assessments –TMA or CMA- performed
- Number of assessments performed
- Date of the first assessment
- Length of the course presentation

Keeping in mind the medical approach, the idea behind is to segregate populations that have been treated with different methods – i.e. different assessment characteristics in our case – and compare dropout expectancy. Due to the absence of A/B testing over the same presentation of a course in the provided dataset, we will search for courses with different parameters in separate presentations. It may not be poasible to establish the relevance of a given factor, as – for example – the number of tests conducted in all presentations of a given course could be the same for all courses, making it impossible to segregate populations based on this criterion. When feasible, we will establish distinct populations based on the analysed courses and examine the evolution of withdrawal over time.

As it may be noticed, the survival approach could be done to analyze the impact of other parameters that are present in the database but not linked to assessments in the withdrawal decision. However, we focus on this subset to address our RQs.

4 Results

4.1 Population segregation based on differences in assessment characteristics

In accordance with the explanation in the preceding section, we begin by segregating populations that have been enrolled in presentations with varied assessment parameters in order to measure the impact on dropout rates. This is done as we do not have an A/B testing scenario where different assessment strategies have been explicitly conducted on different groups. The information summarised in Table 3 pertains to the various presentations contained in the dataset under consideration.

Table 3. Presentation parameters linked to assessment characteristics for the different courses in the dataset

Course	Presentation	Num, TMA	Num. CMA	Max Between TMA (days)	FirstTMA (dat)	Duration	Kind
AAA	2013J	5	0	63	19.0	268	Social
	2014J	5	0	63	19.0	269	
BBB	2013B	6	5	42	19.0	240	Social
	2013J	6	5	49	19.0	268	
	2014B	6	5	42	12.0	234	
	2014J	5	0	56	19.0	262	
CCC	2014B	4	4	70	32.0	241	STEM
	2014J	4	4	77	32.0	269	
DDD	2013B	6	7	49	25.0	240	
	2013J	6	0	42	25.0	261	
	2014B	6	0	42	25.0	241	
	2014J	6	0	49	20.0	262	
EEE	2013J	4	0	56	33.0	268	

	2014B	4	0	49	33.0	241	
	2014J	4	0	63	33.0	269	
FFF	2013B	5	7	42	19.0	240	
	2013J	5	7	49	19.0	268	
	2014B	5	7	42	24.0	241	
	2014J	5	7	63	24.0	269	
GGG	2013J	3	6	63	61.0	261	Social
	2014B	3	6	56	61.0	241	
	2014J	3	6	63	61.0	269	

Data in Table 3 shows that some courses will not be suitable for the comparison of strategies. For instance, no segmentation can be performed for course AAA, as both presentations share common parameters. For course BBB, segmentation will be possible, as the last presentation does not contain CMAs while the others do. We focus on generating populations that help to answer our RQs. This requires considering the time between assessments and the potential use of CMAs.

Segregation based on differences in maximum time between assessments

Although there are minor differences in some cases, we have considered only differences around two weeks. Populations will be generated according to Table 5:

Table 5. Courses containing presentations with relevant differences in maximum time between assessments

Course	Presentations with a lower time between assessments (#days)	Presentations with a higher time between assessments (#days)
EEE	2014B (49)	2014J (63)
FFF	2013B,2014B (42)	2014J (63)

No further analysis can be conducted on course duration, as all courses are annual (duration between 234 and 269 days), and also the duration of the different editions is similar for every given course. The same applies to the time of the first test, as there are only minor differences (in the largest case it does not reach a week).

Use of computer marked assessments (CMAs)

Courses BBB and DDD show differences based on their use of CMAs. In particular, Table 4 summarizes the differences detected.

Table 4. Courses containing presentations with differences in CMA

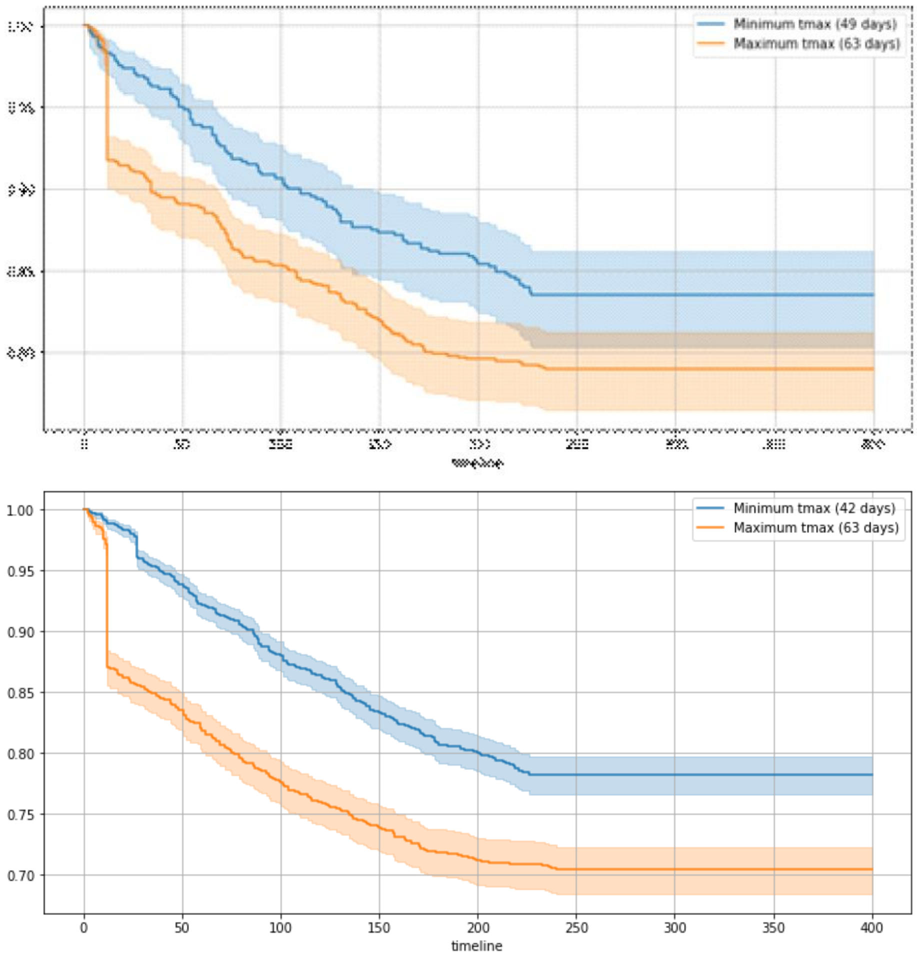
Course	Presentations without CMA	Presentations with CMA (# CMA)
BBB	2014J	2013B, 2013J,2014B (5)
DDD	2013J, 2014B, 2014J	2013B (7)

We use the information in Table 3 and Table 4 to segregate groups of students that have been enrolled in the same course, but with different assessment parameters. The following two subsections will reflect the impact of this difference on dropout.

4.2 Maximum time between assessments: increasing leads to higher dropout

For this test, analysis is performed in courses EEE and FFF. Results are shown in Figure 1:

Fig. 1. Survival curve comparison for groups with a relevant difference in maximum time between assessments. EEE (above) and FFF (below)

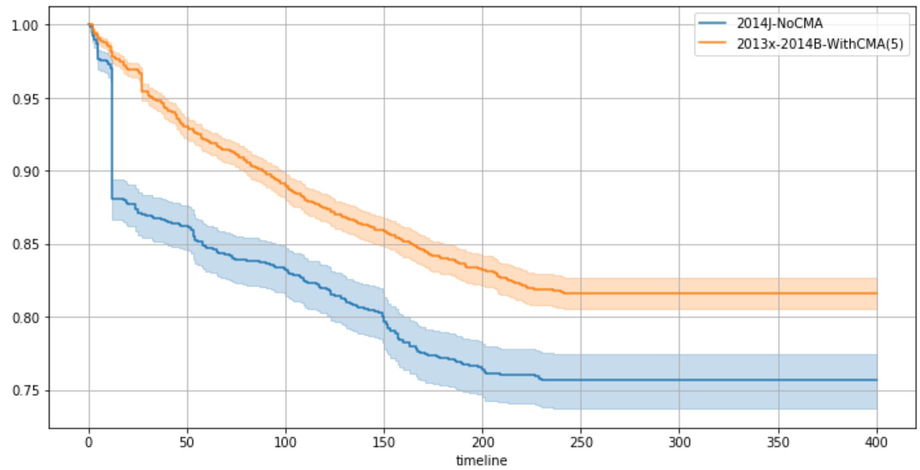


p-Values are 0.01 for the first case and p<0.001 for the second, providing statistical significance in both cases. As the curves reflect, longer periods without assessment derive in higher dropout rates.

4.3 Impact of CMAs: no concluding results can be extracted

Population comparison has been carried out with data in Table 4. Separate analyses have been done for courses BBB and DDD as they correspond to different subjects with different global abandon characteristics. Figure 2 shows curves for both courses.

Fig. 2. Survival curve comparison for groups with and without CMA for courses BBB (above) and DDD (below)



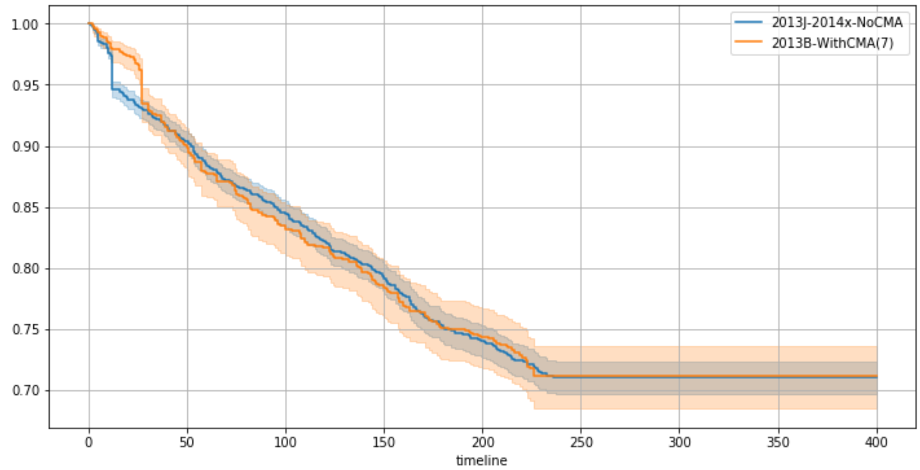


Figure 2 depicts a difference for course BBB, but not for course DDD. Statistical tests to support the null hypothesis (both groups having the same risk hazard curve) indicate $p<0.005$ for course BBB, making the difference statistically significant. As expected from the graph, differences are not significant for course DDD ($p=0.95$)

5 Discussion

Looking for a concise answer to the RQs addressed in this paper, certain assessment characteristics impact the withdrawal decisions of students enrolled in higher education *online* courses. Specifically, increasing time between consecutive assessments increases withdrawal probability. Regarding the inclusion of CMA as part of the evaluative activities, definitive conclusions cannot be drawn based on the data analysed.

From a broader perspective, two main conclusions derive from the research conducted and explained in this paper. First, some assessment strategies work better than others when addressing potential reductions of dropout ratios. Second, survival analysis provides an interesting method for analyzing dropout in general and withdrawal at the course level in particular.

We can compare the impact of assessment characteristics that our results have shown with previous works. Our findings extend the impact on academic performance revealed in (Wang & Zhang, 2020), showing that certain aspects explicitly impact course dropout. In particular, frequency of assessments has shown a clear impact. We cannot draw conclusions – based on our case study – regarding the potential impact of combining different type of assessments. While the use of CMAs has shown impact on one of the courses, it has not been so on another. This might suggest that the impact might depend on the course under analysis.

Regarding these analyzed aspects, time between assessments, and for a given course, a long time without assessments increases also dropout probability. Our finding is consistent with the results in (Holder, 2007) and also aligned with (Kim et al., 2017). The first of these studies indicates that successful students in *online* programs show higher time and study management capabilities. Reducing the maximum time between tests would encourage students to maintain more control, leaving less room for personal judgment. This would help those students with poorer time management capabilities who may be adversely affected by long periods without activity control. The overall effect would be a reduction in the number of withdrawal decisions. The second study, from a different perspective, indicates that setting up constraints in the learning process – and so, reducing student freedom – would provide better academic results in on-line courses. Regarding the use of CMAs, we can raise no conclusion based on the data in our study.

The specific impact of frequency of assessments is also aligned with the causes suggested in (Ellison & Jones, 2019). This frequency would increase attainment, and considering the effect of low attainment in dropout, increasing attainment would also mean a reduction in dropout that has been shown in Figure 2.

Further than the impact of specific factos, we would also like to focus on the method itself. The method exposed has shown of great help to uncover the impact of factors influencing dropout and opens an interesting approach to understand the impact of assessment parameters on it. Kaplan-Meier curves provide a graphical view of differences between groups and allow for clear interpretation. At the same time, tests can be performed to determine whether the results are statistically significant.

The method could be used to extend some results such as those suggested in (Day et al., 2018a). This study outlines the relevance of the appropriate number of assessments rather than just an increase in frequency. While we had no additional data to compare more strategies based on differences in the frequency of assessments, the method provided could help to validate these particular results or other hypothesis linked to the evaluation of comparative impact. Specific data should be needed for this purpose.

The results provided should also be contextualized considering the theories behind dropout. Internal factors are considered one of the four components explaining dropout in the CPM outlined in (Rovai, 2003). Pedagogical factors (learning and teaching styles) are included inside this group. For *online* studies, in particular, we believe that fitting course designs to student's expectations increases academic integration. This was already reflected in (Tinto, 1975) for traditional studies and is also part of the institution teaching quality indicated as a concept influencing dropout in (Behr et al., 2020).

We would want to conclude with the statement from (Gibbs et al., 2003) that "there is more leverage to improve teaching through assessment than through anything else." We feel that this observation holds true when particularly considering potential dropout reduction, and we advocate the analysis of ideal assessment parameters in terms of dropout reduction using the approach presented.

6 Final remarks and future lines

Our study has revealed the impact that certain assessment characteristics may have on dropout at course level. Considering the importance of dropout reduction in higher education, we encourage the use of the methodology exposed to extend this work to analyse other aspects further than those covered in this research.

The fact of using an open database has been useful to perform an unbiased study, but we lack deeper pedagogical knowledge of the internals of the course analysed which might be relevant. We have not been able to compare or contextualize aspects such as difficulty of assessments, neither have been able to deepen into other parameters further than the impact of frequency or the use of CMA. We suggest to record potential changes in assessments policies that may lead to differences, and even to perform A/B testing scenarios based on differences on assessments.

We believe that our study opens up interesting research lines. To name a few, the execution of a comparable analysis focusing on other factors and the extension of the study to undertake a comparative analysis between various types of learning disciplines and institutions are potential future directions. In addition, the investigation of specific applications oriented to design tailored assessment procedures, which could lead to personalised assessment policies oriented to lower dropout rates and ultimately boost academic achievement. The authors are open to collaborate in these research lines.

7 References

Aljohani, O. (2016). A comprehensive review of the major studies and theoretical models of student retention in higher education. *Higher Education Studies*, *6*(2), 1–18. https://doi.org/10.5539/hes.v6n2p1

Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016). Survival analysis based framework for early prediction of student dropouts. *International Conference on Information and Knowledge Management, Proceedings*, 903–912. https://doi.org/10.1145/2983323.2983351

Aragon, S. R., & Johnson, E. S. (2008). Factors influencing completion and noncompletion of community college *online* courses. *International Journal of Phytoremediation*, *22*(3), 146–158. https://doi.org/10.1080/08923640802239962

Assessment Definition. (n.d.). Retrieved August 30, 2021, from https://www.edglossary.org/assessment/

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting Student Dropout in Higher Education. *Undefined*.

Behr, A., Giese, M., Teguim Kamdjou, H. D., & Theune, K. (2020). Dropping out of university: A literature review. *Review of Education*, *8*(2), 614–652. https://doi.org/10.1002/rev3.3202

Bloxham, S., & Boyd, P. (2007). Developing Effective Assessment in Higher Education. In *Developing Effective Assessment in Higher Education: A Practical Guide*.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003a). Survival analysis part II: Multivariate data analysis- An introduction to concepts and methods. In *British Journal of Cancer* (Vol. 89, pp. 431–436). https://doi.org/10.1038/sj.bjc.6601119

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003b). Survival analysis part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit. In *British Journal of Cancer* (Vol. 89, pp. 605–611). https://doi.org/10.1038/sj.bjc.6601120

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003a). Survival analysis part I: Basic concepts and first analyses. In *British Journal of Cancer* (Vol. 89, pp. 232–238). https://doi.org/10.1038/sj.bjc.6601118

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003b). Survival analysis part IV: Further concepts and methods in survival analysis. In *British Journal of Cancer* (Vol. 89, pp. 781–786). https://doi.org/10.1038/sj.bjc.6601117

Cobre, J., Tortorelli, F. A. C., & de Oliveira, S. C. (2019). Modelling two types of heterogeneity in the analysis of student success. *Journal of Applied Statistics*, *46*(14), 2527–2539. https://doi.org/10.1080/02664763.2019.1601164

Day, I. N. Z., van Blankenstein, F. M., Westenberg, M., & Admiraal, W. (2018a). A review of the characteristics of intermediate assessment and their relationship with student grades. *Assessment & Evaluation in Higher Education*, *43*(6), 908–929. https://doi.org/10.1080/02602938.2017.1417974

Day, I. N. Z., van Blankenstein, F. M., Westenberg, P. M., & Admiraal, W. F. (2018b). Explaining individual student success using continuous assessment types and student characteristics. *Higher Education Research and Development*. https://doi.org/10.1080/07294360.2018.1466868

Ellison, L., & Jones, D. (2019). First year law students: the impact of assessment type on attainment. *Assessment and Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2018.1496398

Elton, L. R. B., & Laurillard, D. M. (1979). Trends in Research on Student Learning. *Studies in Higher Education*. https://doi.org/10.1080/03075077912331377131

Foster, C., & Francis, P. (2020). A systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes. *Assessment and Evaluation in Higher Education*, *45*(6). https://doi.org/10.1080/02602938.2019.1696945

Gibbs, G. (1999). Using assessment strategically to change the way students learn. *Assessment Matters in Higher Education*.

Gibbs, G., & Simpson, C. (2004). Does your assessment support your students' learning? *Learning and Teaching in Higher Education*.

Gibbs, G., Simpson, C., & Macdonald, R. (2003). Improving student learning through changing assessment – a conceptual and practical framework Assessment tactics that emphasise the ' Eleven conditions .' *EARLI Conference Padova*.

Holder, B. (2007). An investigation of hope, academics, environment, and motivation as predictors of persistence in higher education *online* programs. *Internet and Higher Education*. https://doi.org/10.1016/j.iheduc.2007.08.002

Kim, T. dong, Yang, M. young, Bae, J., Min, B. a., Lee, I., & Kim, J. (2017). Escape from infinite freedom: Effects of constraining user freedom on the prevention of dropout in an *online* learning context. *Computers in Human Behavior*, *66*. https://doi.org/10.1016/j.chb.2016.09.019

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017a). Open University Learning Analytics dataset. *Scientific Data*, *4*(1), 170171. https://doi.org/10.1038/sdata.2017.171

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017b). Data Descriptor: Open University Learning Analytics dataset. *Scientific Data*, *4*(1), 1–8. https://doi.org/10.1038/sdata.2017.171

Landis, R. N., & Reschly, A. L. (2013). Reexamining gifted underachievement and dropout through the lens of student engagement. *Journal for the Education of the Gifted*, *36*(2). https://doi.org/10.1177/0162353213480864

Larsen, M. S., Kornbeck, K. P., Kristensen, R. M., Larsen, M. R. &, & Sommersel, H. B. (2013). *Dropout phenomena at universities: What is dropout? why does dropout occur? what can be done by the universities to prevent or reduce it? : A systematic review*. Danish Clearinghouse for Educational Research.

Lee, Y., & Choi, J. (2013). Discriminating factors between completers of and dropouts from *online* learning courses. *British Journal of Educational Technology*, *44*(2), 328–337. https://doi.org/10.1111/j.1467-8535.2012.01306.x

Lynam, S., & Cachia, M. (2018). Students' perceptions of the role of assessments at higher education. *Assessment & Evaluation in Higher Education*, *43*(2), 223–234. https://doi.org/10.1080/02602938.2017.1329928

Miller, C., & Parlett, M. (1974). *Up to the Mark: A Study of the Examination Game.*

Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., & Kloos, C. D. (2019). Prediction in MOOCs: A review and future research directions. *IEEE Transactions on Learning Technologies*, *12*(3), 384–401. https://doi.org/10.1109/TLT.2018.2856808

Naude, L., Nel, L., van der Watt, R., & Tadi, F. (2016). If it's going to be, it's up to me: first-year psychology students' experiences regarding academic success. *Teaching in Higher Education*, *21*(1), 37–48. https://doi.org/10.1080/13562517.2015.1110788

Open University, T. (2022). *Changing Your Study Plans Policy 2022/23*. https://help.open.ac.uk/documents/policies/changing-your-study-plans/files/214/changing-your-study-plans-policy-2022-23.pdf

Orellana, D., Segovia, N., & Cánovas, B. R. (2016). El abandono estudiantil en programas de educación superior virtual: revisión de literatura. *Revista de La Educación Superior*, *49*(194), 45–62. https://doi.org/10.36857/resu.2020.194.1124

Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited: a review of research in Assessment and Evaluation in Higher Education. *Assessment and Evaluation in Higher Education*, *41*(7). https://doi.org/10.1080/02602938.2015.1055233

Preston, R., Gratani, M., Owens, K., Roche, P., Zimanyi, M., & Malau-Aduli, B. (2020). Exploring the Impact of Assessment on Medical Students' Learning. *Assessment & Evaluation in Higher Education*, *45*(1), 109–124. https://doi.org/10.1080/02602938.2019.1614145

Richardson, J. T. E. (2015). Coursework versus examinations in end-of-module assessment: a literature review. *Assessment & Evaluation in Higher Education*, *40*(3), 439–455. https://doi.org/10.1080/02602938.2014.919628

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2022). Beyond one-size-fits-all in MOOCs: Variation in learning design and persistence of learners in different cultural and socioeconomic contexts. *Computers in Human Behavior*, *126*(C), 106973. https://doi.org/https://doi.org/10.1016/j.chb.2021.106973

Rovai, A. P. (2003). In search of higher persistence rates in distance education *online* programs. *Internet and Higher Education*, *6*(1), 1–16. https://doi.org/10.1016/S1096-7516(02)00158-6

Simpson, A. (2016). Assessment and its outcomes: the influence of disciplines and institutions. *Assessment & Evaluation in Higher Education*, *41*(6), 917–937. https://doi.org/10.1080/02602938.2015.1052369

Spronken-Smith, R., Cameron, C., & Quigg, R. (2018). Factors contributing to high PhD completion rates: a case study in a research-intensive university in New Zealand. *Assessment and Evaluation in Higher Education*, *43*(1). https://doi.org/10.1080/02602938.2017.1298717

Standards and Guidelines for Quality Assurance in the European Higher Education Area. (n.d.).

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89–125. https://doi.org/10.3102/00346543045001089

Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment and Evaluation in Higher Education*, 505–521. https://doi.org/10.1080/02602930600679506

Vandist, K., de Schryver, M., & Rosseel, Y. (2009). Semisupervised category learning: The impact of feedback in learning the information-integration task. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/APP.71.2.328

Villano, R., Harrison, S., Lynch, G., & Chen, G. (2018). Linking early alert systems and student retention: A survival analysis approach. *Higher Education*, *76*(4), 903–920. https://doi.org/10.1007/s10734-018-0249-y

Wang, S., & Zhang, D. (2020). Perceived teacher feedback and academic performance: the mediating effect of learning engagement and moderating effect of assessment characteristics. *Assessment & Evaluation in Higher Education*, *45*(7), 973–987. https://doi.org/10.1080/02602938.2020.1718599

Xing, W., Tang, H., & Pei, B. (2019). Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. *Internet and Higher Education*, *43*, 100690. https://doi.org/10.1016/j.iheduc.2019.100690

Capítulo 4 Minería de procesos educativos

“La educación es un ornamento en la prosperidad y un refugio en la adversidad.”
Aristóteles (384 - 322 a.C.)

Artículos que se presentan:

Martínez-Carrascal, J. A., Valderrama, E., & Sancho-Vinuesa, T. (2020).

Combining clustering and sequential pattern mining to detect behavioral differences in *log* data: conceptualization and case study. Proc. of Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?, *online*, June 15-16, 2020, http://ceur-ws.org/Vol-2671/paper04.pdf

Martínez-Carrascal, & Sancho-Vinuesa, T. (2022).

Using *Process mining* to determine the relevance and impact of performing non-evaluative quizzes before evaluative assessments. Proc. of Learning Analytics Summer Institute Spain 2022. Learning Analytics: Here to stay, Salamanca, Spain, June 20-21, 2022, http://ceur-ws.org/Vol-3238/paper7.pdf

Martínez-Carrascal, J. A., Munoz-Gama, J., & Sancho-Vinuesa, T. (2023).

J. A. Martínez-Carrascal, J. Munoz-Gama and T. Sancho-Vinuesa, "Evaluation of Recommended Learning Paths using Process Mining and Log Skeletons: Conceptualization and Insight into an Online Mathematics Course", in IEEE Transactions on Learning Technologies, doi: 10.1109/TLT.2023.3298035.

4.1. Contextualización de los artículos

El capítulo previo mostró ya la conveniencia de las aproximaciones temporales en el análisis del rendimiento académico, y muy especialmente en el del abandono. Sin embargo, el análisis de supervivencia – y específicamente tal y como lo hemos planteado– está claramente orientado al análisis y predicción temprana del abandono, y pese a su enfoque dinámico, no proporciona una visión integral del proceso de aprendizaje. Esta visión de proceso requiere de otras técnicas, que serán las que permitan abordar visiones más completas del aprendizaje.

En este capítulo continuaremos analizando técnicas de uso poco común en entornos educativos. Dos son las razones que impulsan este análisis. Por un lado, la búsqueda de una visión más completa del proceso de aprendizaje. Por otro, explorar la mejora académica más allá del abandono haciendo uso igualmente de aproximaciones temporales. Como veremos, estas técnicas buscan una visión más completa, y a la vez, presentan mayor novedad. Aun cuando no es el foco del presente trabajo, son también computacionalmente más complejas por tratarse de modelos más sofisticados.

Con el mismo espíritu que el mostrado en el capítulo anterior, hemos explorado como se analizan procesos en diferentes ámbitos más allá del educativo. Esto nos llevó a considerar inicialmente la búsqueda de secuencias. Comprobamos que en entornos empresariales ciertas secuencias podían ser indicadoras de un mal funcionamiento del proceso subyacente. Esta aproximación tenía la ventaja de simplificar el análisis de procesos más complejos, donde un modelado completo puede resultar realmente complejo.

Trasladando estas ideas al ámbito educativo, nuestra exploración comenzó por la búsqueda de secuencias específicas en las tareas que pudieran indicar un potencial riesgo desde el punto de vista académico. Se trataba de una búsqueda de patrones (y en particular, de técnicas de *sequence mining*) que ahondaran en la búsqueda de patrones de *'disengagement'* que se habían mostrado particularmente relevantes cuando se observaba el abandono.

El primero de los artículos de este bloque versa sobre esta cuestión (Martínez-Carrascal, Valderrama Vallés, et al., 2020). Dicho estudio extiende la visión temporal con que se cerró el capítulo previo, realizando una búsqueda de patrones de comportamiento. Son patrones preliminares, que cuantifican el volumen de actividad a lo largo del tiempo. No muestran aún la visión de qué tareas específicas realiza el estudiante en cada intervalo, pero sí detectan períodos de inactividad que sostenidos en el tiempo pueden anticipar el fracaso.

Esta línea de investigación mantiene la visión temporal, profundiza más allá del abandono, y se acerca a una visión de proceso. Aun así, no constituye formalmente un modelado del proceso, entendido como la secuenciación temporal de un conjunto de actividades específicas. Realizar un modelado del proceso en este sentido comportará utilizar técnicas de minería de procesos (*process mining*). Aunque su uso en trabajos realizados desde la perspectiva de *learning analytics* no es particularmente extendida, creemos que presenta oportunidades significativas en el análisis del proceso de aprendizaje, y en particular, abre vías a la mejora académica.

Con esta idea en mente, el segundo de los artículos de esta sección (Martínez-Carrascal & Sancho-Vinuesa, 2022) inicia la exploración metodológica del uso de *process mining* orientado a la mejora académica a nivel de curso. Una visión clara de qué hace el estudiante y cuándo permite explorar por qué puede estar potencialmente en riesgo. Esta exploración se completará con una propuesta metodológica de evaluación de los caminos de aprendizaje con especial foco en el impacto de dicho camino en el rendimiento (Martínez-Carrascal, Munoz-Gama, et al., 2023).

Adicionalmente, y para aquellas aplicaciones más versadas en la analítica predictiva, este último artículo abre la que puede ser considerada como una de las vías más interesantes: el análisis no del proceso en su totalidad, sino de sus características principales, para determinar hasta qué punto el estudiante sigue o no el proceso sugerido por el docente. Como se verá en dicho artículo, será un aspecto de suma importancia a la hora de evaluar caminos de aprendizaje, y contribuye tanto a evaluar la conveniencia de un camino de aprendizaje como al diseño de soluciones de aprendizaje personalizado (‘*tailored learning*’). La validación de la conveniencia de seguir un camino determinado así como de detectar discrepancias cuantificables contribuirán a la mejora del rendimiento académico, objetivo último de la presente tesis.

Con el propósito de dar sentido y coherencia a estos tres artículos, las siguientes subsecciones exponen los conceptos clave implicados. La sección 4.2. expone los fundamentos de las técnicas de búsqueda de secuencia, englobadas bajo el concepto genérico de '*sequence pattern matching*' (SPM). Seguidamente, la sección 4.3. expone las bases de la minería de procesos. Finalmente, incluimos una sección específica sobre los '*skeletons*' asociados a un proceso, que se demostrarán como una herramienta de particular utilidad y que abren muchas posibilidades en la investigación futura sobre aprendizaje personalizado (Sección 4.4). Un breve apartado resume las conclusiones de este bloque (Sección 4.5).

4.2. Sequence Pattern matching (SPM)

La aproximación hacia la modelización del aprendizaje como proceso empezó por el análisis de secuencias. Haciendo uso de nuevo del símil médico – como sucedió con el análisis de supervivencia –, la idea inicial fue buscar patrones que pudieran anticipar un bajo rendimiento. De la misma forma que se buscan cadenas en el ADN indicadoras de una determinada patología, pueden llegar a establecerse patrones que anticipen un determinado problema de rendimiento.

Las técnicas que exploraremos en este capítulo constituyen técnicas propias de *Educational Data Mining* (EDM). Estarían así alineadas con las exploradas en el capítulo 2. El hecho de no haberlas tratado con anterioridad se debe a que son menos comunes. Remarcar que, pese a ello, buscaremos siempre su sentido educativo más allá del algoritmo (es decir, buscaremos siempre la interpretación en términos de *learning analytics*). Dentro de las técnicas de *process mining*, se habla de minería de procesos educativos (*Educational Process mining*, EPM) (Bogarín et al., 2018; Reimann et al., 2014). En esa exploración nuestro análisis empieza por la búsqueda de patrones secuenciales (*Sequential Pattern Mining* o SPM por sus siglas en inglés).

Denominaremos secuencia a la sucesión ordenada de eventos, que ocurren en un espacio de tiempo objeto de análisis. Para su análisis se incluyen diferentes técnicas que pueden clasificarse según su orientación:

- Técnicas orientadas a la *clusterización* o agrupación, cuando el objetivo es identificar grupos que presentan secuencias similares. Estas técnicas están basadas generalmente en modelos de Markov.
- Técnicas orientadas a la clasificación, en las que se aportan un conjunto de secuencias, identificadas como pertenecientes a una determinada categoría. Este conjunto constituye un grupo de test, que servirá como base para clasificar otras muestras. Está basado en técnicas como SVM ya expuesta en el capítulo 2 del presente trabajo.
- Técnicas orientadas a la predicción, donde el objetivo es identificar el siguiente elemento de una secuencia.
- Técnicas orientadas a la segmentación, donde el objetivo es encontrar subsecuencias dentro de secuencias más amplias.

Una visión detallada de los diferentes algoritmos asociados a estas categorías puede encontrarse en (Fournier-Viger et al., 2016). El propio autor proporciona una lista actualizada de algoritmos propuestos que incluye a fecha de hoy 254 implementaciones, hecho que da idea de la variabilidad y complejidad de las técnicas subyacentes.

En este marco, el primero de los artículos presentados en este capítulo (Martínez-Carrascal, Valderrama Vallés, et al., 2020) explora cómo la combinación de secuenciación y *clustering* puede contribuir a determinar el fracaso académico a nivel de curso. En lo que refiere específicamente a la secuenciación, se hace uso del algoritmo GSP (*generalized sequential pattern*) cuyos fundamentos pueden encontrarse en (Agrawal & Srikant, 1995). Este algoritmo es uno de los sugeridos y comúnmente utilizados (Fournier-Viger et al., 2017). Para la clusterización, y basándonos en nuestra experiencia del Capítulo 2, utilizaremos k-Means, combinando así técnicas clásicas y técnicas estrictamente basadas en secuencias.

Es destacable también que más allá de su evolución de técnicas – que pasa del análisis de supervivencia al análisis de secuencias – el foco del estudio también evoluciona. Se pasa del análisis del abandono a la detección de diferencias en el comportamiento aprobado-suspenso. Este avance es necesario para tener una visión global de la mejora académica, que incluya tanto la reducción del abandono, como la mejora de las calificaciones de los estudiantes que continúan en el curso.

El artículo ahonda en la relevancia de la actividad – ya vista en los trabajos preliminares mostrados en el capítulo 2–. De todas formas, esta parte de la investigación ya no se limita a constatar una visión estática de la actividad, sino que pone de manifiesto la evolución temporal. Específicamente, el estudio refleja que combinar períodos de baja actividad supone un aumento notable en la posibilidad de fracaso académico. Se avanza así en la visión temporal, pero aún no se llega a una visión clara del proceso de aprendizaje. Para ello será necesario un paso más: la minería de procesos en sentido estricto, que se aborda seguidamente.

4.3. Minería de procesos *(process mining)*

La minería de procesos (*process mining*) completa la exploración de técnicas, y en particular, la trasposición al mundo educativo de métodos más comunes en otras disciplinas. Estas técnicas surgieron inicialmente en los entornos empresariales, para analizar procesos de negocio. En el ámbito empresarial, ciertas tareas están claramente constituidas por etapas que se deben completar en un orden específico. Aun así, la realidad a menudo difiere del modelo teórico, provocando ineficiencias en los procesos. Es este escenario es donde la minería de procesos actúa con el fin último de mejorar la ejecución de los mismos. El siguiente objetivo de nuestra investigación es analizar cómo estas técnicas pueden aportar información relevante en un proceso radicalmente diferente: el proceso de aprendizaje.

Las bases de esta metodología se describen en (van der Aalst, 2012). A efectos de nuestra investigación, resulta fundamental establecer cómo se realiza el modelado de eventos, para a partir de ahí entrar en los aspectos nucleares del *process mining*: el descubrimiento de procesos (*process discovery*), la búsqueda de alineación entre un modelo y la ejecución real del mismo (*conformance checking*), y finalmente, la mejora del proceso en base al análisis de resultados (*process enhancement*). Gráficamente la figura 4.1 relaciona los conceptos fundamentales.

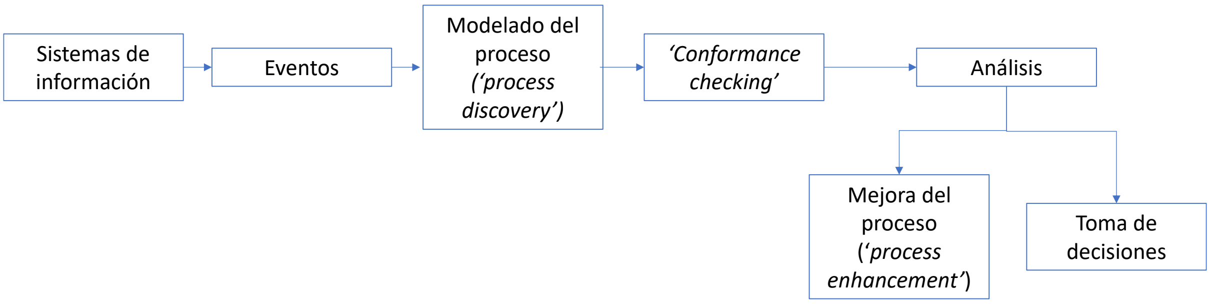


Figura 4.1– Conceptos fundamentales asociados a la minería de procesos.

La portabilidad de estas técnicas a entornos académicos ha sido explorada en los últimos años dando lugar a la conocida como *educational process mining* (Bogarín et al., 2018; Ghazal et al., 2018). Analizando la potencialidad de la herramienta que reflejan los estudios referenciados, se propone un marco de investigación que se estructura en las siguientes etapas. En primer lugar, el modelado del curso en forma de proceso. Esta fase debe proporcionar un modelo automático mediante algoritmos de descubrimiento del proceso de aprendizaje. Para ello se requiere la traducción del propio fichero de *log* a secuencia de eventos con significado a nivel de aprendizaje. Se evoluciona desde la visión estática de variables individuales (el número de *clicks* o número de *logins*) hacia un modelo que contempla la visualización de una unidad didáctica o la realización de un cuestionario específico. Serán estos los eventos reales que conformarán nuestro proceso.

Realizado el modelado teórico y conseguido el *log* de eventos con significado pedagógico, nuestro objetivo era validar hasta qué punto el estudiante sigue el camino pautado. Esto lo planteamos con técnicas de *conformance checking* que permitieran visualizar hasta qué punto el camino seguido por el estudiante es coherente con el planteado por el docente y, adicionalmente, evaluar si seguir dicho camino es relevante a nivel de desempeño académico.

Estos conceptos se materializan en (Martínez-Carrascal & Sancho-Vinuesa, 2022). De cara a validar el proceso de modelado descrito, se partió del resultado de una investigación previa (Figueroa-Cañas & Sancho-Vinuesa, 2021). El estudio analiza un curso de Matemáticas asociado a un grado universitario *online*, y muestra cómo realizar cuestionarios de práctica antes de las pruebas evaluativas tiene un impacto positivo en el rendimiento académico.

Nuestra investigación busca validar dicho resultado, pero a partir del modelado del proceso de aprendizaje mediante técnicas de minería de procesos.

Este estudio inicial sirvió para la exploración de técnicas, pero puso de manifiesto dos problemáticas relevantes. Por un lado, los estudiantes – pese a tener un modelo trazado – realizan trayectorias diferentes a las planteadas por el profesorado. Esto complica el descubrimiento del modelo asociado al comportamiento real, y hace particularmente enrevesado determinar hasta qué punto se está siguiendo un camino sugerido. Esto es así por dos razones. En primer lugar, el alto número de actividades y caminos que puede seguir el estudiante hace difícil establecer hasta qué punto alguien sigue un determinado camino. Los procesos empresariales suelen tener pocas actividades, y en todo caso, un orden claramente establecido. En ámbitos educativos, es altamente improbable que dos estudiantes sigan una misma secuencia. Por otra parte, los algoritmos de *conformance* habituales no disponen de indicadores cuyo valor indique, sin necesidad de contextualización adicional, la similitud entre un modelo descubierto y un *log* específico. Existen técnicas de *'match'* entre un *log* y un modelo que miden dónde el *log* se desvía del modelo, o bien el modelo del propio *log*, pero el significado e interpretación de la desviación resultante depende del modelo y del mismo *log*.

4.4. *Skeletons*: concepto y potencialidad

La búsqueda de modelos más simples que a su vez permitieran una mejor evaluación de la alineación llevó al uso de los conocidos como *skeletons* (Kudo et al., 2013). Éstos constituyen una visión declarativa del proceso, basada en las relaciones habituales que se dan en un conjunto de eventos. Es decir, en lugar de definir o buscar una visión gráfica del proceso completo, se buscan relaciones que se dan o no se dan habitualmente, así como frecuencias habituales de las distintas tareas.

Considerando un hipotético proceso académico, la idea es – a modo de ejemplo – visualizar si un estudiante que presenta un buen rendimiento académico siempre realiza una determinada tarea antes que otra, o bien, nunca realiza según qué tipo de tareas o secuencias. No se trata por tanto de obtener una visión gráfica completa (que como muestra la literatura, y como confirma nuestro estudio (Martínez-Carrascal, Munoz-Gama, et al., 2023), da lugar a visiones de difícil interpretación) sino de una visión de alto nivel de las relaciones que se dan en un proceso. Son elementos habituales de este *skeleton* los mostrados en la Tabla 4.1.:

Tabla 4.1 – Relaciones fundamentales que incluye el *skeleton* de un proceso

Concepto	Descripción
Equivalencia	Actividades que muetran la misma frecuencia en todas las variantes del proceso
Siempre después (A,B)	Detección de actividades (A) que son siempre seguias de otras dadas (B) en el futuro.
Siempre antes (B,A)	Detección de actividades (B) a las que siempre precede la ejecución de otras (A)
Nunca juntas	Actividades que nunca se encuentran en una misma traza
Sigue directamente (A,B)	Actividades (A) que siempre siguen directamente a otras (B) en todas las trazas
Frecuencia de la actividad	Máximo número de veces que una traza está presente en las trazas analizadas

Estos *skeletons* presentan una ventaja adicional y es que disponen de métodos específicos para contrastar hasta qué punto un determinado *log* se ajusta a las características contenidas en el *skeleton*, proporcionando un indicador acotado de interpretación simple en el rango [0,1], donde 0 indica una discrepancia total entre el *log* y el *skeleton* y 1 un ajuste perfecto. Su potencialidad se ha reflejado en aplicaciones orientadas a la clasificación de comportamiento fuera del ámbito educativo (Verbeek & de Carvalho, 2018).

Partiendo de esta idea, planteamos utilizar el *skeleton* del proceso, en lugar de su visión completa como modelo del proceso. Basados en este skeleton, seremos capaces de determinar el ajuste (*conformance*) de una ejecución específica en relación al modelo. Es decir: determinaremos un índice de similitud entre lo que hace realmente el alumno, y el camino sugerido por el profesorado. Adicionalmente, podremos determinar cuáles son las divergencias de comportamiento más habituales.

El último artículo que se presenta en este compendio (Martínez-Carrascal, Munoz-Gama, et al., 2023) materializa metodológicamente esta idea y muestra cómo aplicarla para determinar el seguimiento que hace el estudiante del camino propuesto por el docente (Martínez-Carrascal, Munoz-Gama, et al., 2023). Determina también la correlación existente entre este seguimiento y el rendimiento académico, medido como la nota final de la asignatura.

La aplicación práctica muestra el potencial de los *skeletons* basándose en una aproximación novedosa, que utiliza para generar el modelo no las recomendaciones teóricas del profesorado, sino la ejecución real de actividades que realizan los estudiantes que siguen dichas indicaciones. La Figura 4.2. refleja el proceso sugerido.

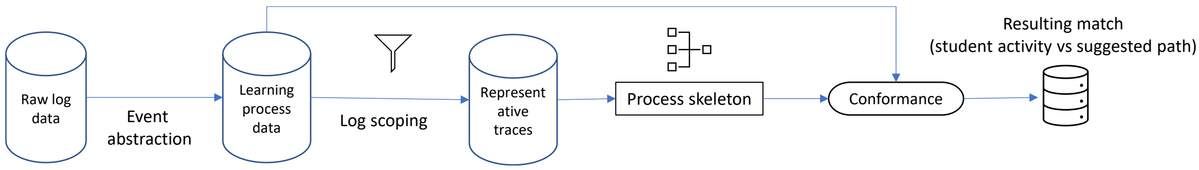


Figura 4.2–Modelo conceptual para el análisis de trayectorias académicas y su impacto en el rendimiento académico

El proceso planteado contempla 4 etapas fundamentales:

1. Obtención de los eventos semánticamente relevantes a partir de los datos recogidos (*event abstraction*)
2. Aplicación de reglas específicas para extraer aquellas trayectorias que se muestran conformes con el proceso deseado (*log scoping*)
3. Cálculo del *skeleton* asociado a las trazas resultantes del proceso de *scoping*.
4. Cuantificación y evaluación de las diferencias de trazas específicas respecto al modelo resultante (*conformance*).

Remitimos al lector a la sección *'3.1. Modelling learning paths through a combination of theoretical constraints and real student activity'* del último artículo de esta sección para una explicación detallada de los diferentes pasos y los aspectos metodológicos. La potencia del método radica en la posibilidad de modelado de cualquier curso, permitiendo establecer la relevancia y conveniencia de seguir uno u otro camino. Esta contribución específica permite establecer caminos convenientes de aprendizaje, y adicionalmente, detectar las divergencias que pueden suponer un riesgo. En lo que respecta a nuestros objetivos, estos aspectos contribuyen a la mejora del rendimiento académico que se persigue en este trabajo.

4.5. Conclusiones

La visión dinámica que se planteó en el capítulo 3 con el análisis de supervivencia focalizado en el abandono, se ha completado con la extensión al modelado del aprendizaje como proceso. Tras la exploración inicial de técnicas de análisis de secuencia, se ha llegado a un modelado basado en las técnicas de minería de procesos. Hemos sugerido un método – basado en el uso de *skeletons* – que permite modelizar el camino de aprendizaje combinando la propuesta teórica del profesorado con la ejecución real que realiza el estudiante.

De esta forma, se completa el modelado del proceso del aprendizaje a nivel de curso, aportando un método aplicable a cualquier curso. Las potencialidades que abre esta aproximación son múltiples. Indicamos como las más relevantes la contrastación de diferentes propuestas de caminos de aprendizaje, la evaluación de divergencias del estudiante que pueden anticipar un problema académico, y la validación en sí misma de los caminos de aprendizaje sugeridos a nivel de curso. Estas aportaciones contribuyen de forma clara a la mejora académica, ayudando al diseño de potenciales intervenciones, tanto mediante la definición de caminos que se demuestran adecuados para la superación del curso como mediante la actuación en sí cuando se detecte una divergencia de dicho camino.

Presentamos seguidamente los tres artículos vinculados a este capítulo. Estaremos así en condiciones de abordar la respuesta a las preguntas de investigación que se plantearon en el capítulo introductorio y que constituirán el capítulo final de esta tesis.

Artículo 6: Combining clustering and sequential pattern mining to detect behavioral differences in *log* data: conceptualization and case study.

In Proc. of Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?, *online*, June 15-16, 2020, http://ceur-ws.org/Vol-2671/paper04.pdf

Scopus: SJR 2020: 0.177, Category: Computer Science (miscellaneous)

Nota: Las referencias numéricas a figuras y tablas contenidas en el artículo que se presenta a continuación deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

Combining clustering and sequential pattern mining to detect behavioral differences in log data: conceptualization and case study

Juan Antonio Martínez-Carrascal [1,2*] [0000-0002-7696-6050], Elena Valderrama [2] [0000-0001-7673-2310] and Teresa Sancho-Vinuesa[1] [0000-0002-0642-2912]

[1] Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain
[2] Universitat Autonoma de Barcelona, Engineering School, Campus UAB, 08193 Bellaterra, Spain
jmartinezcarra@uoc.edu

Abstract. Many on-campus universities are shifting their methodologies towards blended learning models. In these models, students cover some content online – normally associated with traditional lectures and quiz practice–, and attend also on-campus activities. Online activity is recorded in Learning Management Systems (LMS) logs, where interactions of the students with content are recorded.

In this article, we propose a method to detect differences in behavior considering only data recorded in the LMS log. We begin by clustering students based on activity log information. This process is carried out on a periodical basis. Clustering results are translated into meaningful states and then sequenced. The generated sequence is mined through sequential pattern mining (SPM).

Besides method description, we apply our method to a specific case-study to prove its validity. In particular, we analyze differences between passing and failing students in a blended-learning course. We prove that the method can generate meaningful sequences which – once analyzed – show relevant behavioral differences between students who pass and those who fail. In particular, failing students show more disengagement patterns than those who pass, while working attitudes - in particular, continuous working - are more common among the passing group.

Keywords: behavioral analysis, sequential pattern mining, process mining, student modelling, learning analytics.

1 Introduction

Learning management systems (LMSs) are nowadays a common piece in any university. Some of them were in fact born thanks to these systems. On the other side, some traditional on-campus universities were initially reluctant to develop them on their campuses. The perceived added value of the campus life – including campus classes – was considered a core value which did not admit changes.

This initial view is nowadays no more than a reminder of the days where the Internet was stilly maturing. Today, it would be difficult to find any university which has not adopted some kind of LMS [1]. LMSs have contributed to nuclear changes in the way instructors teach and students learn [2]. Beyond its relevant participation in the learning process, it has provided researchers with a valuable item: data.

Data coming from LMSs can be analyzed alone, combined with pre-existing student data, or with data coming from other systems. A whole set of data-mining techniques has been developed to solve different problems that range from performance prediction to analysis of social engagement [3]–[5]. These techniques can be analyzed from a computational perspective – which in the educational field leads to Educational Data Mining (EDM) [6] – or looking for its application in the learning process – giving place to Learning Analytics (LA) [7], [8]-.

Among these techniques, educational process mining (EPM) has gained popularity in recent years [9], [10]. EPM derives from generic process mining (PM) [11] but applies specifically to educational data. Inside this subset, sequential pattern mining (SPM) [12] tries to find interesting subsequences in a dataset. In the learning scenario, potential applications of this technique covers a broad range of topics. While our interest focuses on the detection of behavioral differences, other applications include better curriculum design or performance prediction [13]–[15].

In this paper we describe a method based on SPM to detect behavioral differences between passing and failing students in a blended-learning course based on log data. The method is tested on a first-year Engineering course offered at a public university. The subject under analysis was designed with blended-learning approach and lasts for 12 weeks, including three main milestones which correspond to intermediate tests performed in weeks 4 and 8 and a final test in week 12. The edition we analyze has a total of 337 students enrolled – 199 of them passing the subject -. A detailed description of the course structure can be found in [16], where results show that both groups effectively behave differently.

Our goal is to evaluate if we can detect these differences through a new method. The method we describe begins by clustering log activity on a per-week basis. Output of this clustering process is then interpreted to determine the weekly state of the student. We iterate this process in consecutive periods to create a sequence of states. Data is then split in two groups – associated to passing and failing students – and SPM is used to detect behavioral differences. In particular, we raise the following research question (RQ):

- RQ: Can the method exposed collect and show behavioral differences among both group of students?

To answer the question, we describe the method in detail and apply it to the aforementioned case study. Once done, we analyze if different behavioral patterns can be found in the sequences present in the failing and passing group. Results will reveal that behavior is different for the passing and failing group, and specific patterns which anticipate potential failure arise. This fact, combined with the simplicity to gather and process data, opens interesting applications such as predictor or alarm indicator.

2 Theoretical framework

Research on learning analytics [7], [8] focuses on different aspects of learning, being one of them the analysis of the learning process itself [9], [17]. In this context, studies that apply process analysis tool to learning environments have emerged with rising interest in recent years, as different compilations show [9], [10].

The process approach was common in other disciplines [18], in particular in business and industry. In the learning scenario, specific areas of interest include curriculum mining, computer-based assessment or LMS log analysis. Expected results include the detection of learning difficulties, learning flows or sequential patterns[9].

While processes can be designed based on theoretical behavior, complex or unstructured processes need a process discovery stage [18]. The learning process totally fits into this categorization. Process discovery constitutes a discipline on its own, where one of the approaches is to model processes based on log records, that are translated into real models, including the detection of variants in the same process[19]–[21]. The field is promising, but initial models usually conform *spaghetti-like* graphs, which require simplification [22]. In addition, direct interpretation is not straightforward.

For these reasons, some studies focus on analysis of sub-processes, or look for meaningful sequence of actions. The basics for these techniques were introduced in [23] and constitute the core of SPM. In particular, SPM deals with "sequences of events, items, or tokens occurring in an ordered metric space appear often in data and the requirement to detect and analyze frequent subsequences is a common problem.“ [24].

SPM applications are common in different knowledge areas ranging from medicine to business processes [12]. Applications are broad, depending on the topic under analysis. For instance in the medical field, they range from prescription [25] to detection of specific patterns, such as gene mutation [26].

In the learning field, we can find different studies and applications. [27] looks for better understanding of the learning process through the use of a specific algorithm. [13], [28] show applications in recommender systems. Other uses include impact of behavior during practical sessions on final performance [29] or discovering of navigational patterns [14]. Applications to behavioral analysis can also performed through SPM [30].

Practical implementation of SPM can be done in two different ways: either a-priori or using pattern growth. A-priori methods are based on [23], and rely on the hypothesis that if a sequence is not frequent, super sequences based on it can neither be frequent. Pattern-growth methods are based on [31]. Any of these methods can provide, for each of the detected sequences, its support, which indicates the percentage of items where the sequence is present.

In order to feed these SPM algorithms with data coming from log files, preprocessing is needed, as a sequence of states is required as input. [32] remarks that one of the challenges when translating logs into process is the granularity of log data. In order to translate this log data into states, an aggregation is needed, as working with low-granularity data can be useless [14].

An interesting approach to this aggregation is the use of clustering based on log data, which can be found in studies such as [33], [34]. In particular, in [34] clustering is used to detect groups, while SPM is used to detect sequences in that group. Following this research line, we plan to create the sequences based on the clustering process itself. In other words, we do not plan to cluster students, but to cluster states on a periodical basis that will finally constitute a sequence.

To feed clustering algorithms, commonly interesting variables include the number of lectures viewed, quizzes assessed or time between consecutive sessions. These parameters show relevant for student success, with relevance depending on the case under study [35]. Some of these parameters require a deep knowledge of the course structure. While it is relatively simple to know how many times a user logins per week, knowing how many lectures or quizzes a user performs requires prior classification of LMS items – associating a category to each individual item–.

Regarding the algorithm itself, different techniques can be used. [36] performs a compilation of potentially interesting algorithms which are common in e-Learning problems.

3 Methodology

Linking these ideas, we build a sequence of states for each user and time interval. The states are created through clustering based on data extracted from log files. These states are sequenced and a search is performed. Analysis will look for relevant patterns that can show behavioral differences.

We decided to analyze activity on a per-week basis for two reasons. First, due to the dynamics of the blended course. Instructors anticipate contents the students should cover before attending class, and this is normally done on a weekly basis. Second reason influencing this decision takes into account student habits. While some students are full-time students and can possibly cover contents during the week, some others may only be able to cover them during the weekend. The weekly analysis accommodates this situation, showing whether students are effectively engaged during the different weeks of the course.

Clustering results for the different weeks is then be analyzed and labelled according to meaningful patterns. Once done, a sequence is created for each student. This sequence will be meaningful, as labels will have been assigned to each of the states.

Finally, and in order to validate the method with a practical case study, we will apply it to a real case. We separate students into two groups. This segmentation will be done according to academic result – pass or fail – For each of these groups, we will perform

SPM. After getting the results, comparison will be made between groups, in order to validate if the method properly detects differences in support for specific patterns between groups.

3.1 Clustering: input data and algorithms

As outlined in the introduction, we will focus in online activity. The data we analyze corresponds to that gathered in the LMS. We do not include data from other sources, neither take into account any evaluative marks obtained by the students. We consider the global amount of online activity performed, but also classify interactions according to the kind of content covered.

Online activity will be classified based on the categorization of the tasks students are instructed to follow. On each class, students are suggested to cover specific contents before to prepare forthcoming sessions. These contents fit into one of these categories:

- Lectures: which correspond to encapsulated videos provided prior to face-to-face sessions.
- Problem sets: where the student can test to what extent she has acquired knowledge properly
- Evaluative quizzes: that correspond to quizzes that have impact on final grade
- Specific non-assessed contents: which correspond to contents related to the subject, and which are covered in classes, but that are not assessed in any evaluation, and do not have impact on the final grade
- Suggested readings.

Instructors were asked to detect and inform relevant content for each week of the course. Students are specifically instructed to cover these contents before attending specific face-to-face sessions. A total of 150 items were considered. Table 1 shows the type and amount of activities considered.

Table 1. Number of items in the course for each content type

Kind of activity	Number of items
Lectures	54
Problem sets	31
Evaluative quizzes	7
Specific non-assessed contents	12
Suggested readings	46

Data for each of these five kind of content are kept along with the number of login sessions the user performs in the period under consideration. For each week and student, we summarize the number of items in each category – for instance, a student can watch 5 lectures, review 3 problem sets and do not perform any other kind of activity –, performing 4 login sessions. This information will constitute the input to the clustering algorithm. To be able to extract this information, instructors must classify contents in advance in order to properly account each access. It is interesting to note that once this is done, the approach is computationally simple, and data to feed the clustering algorithm is readily available.

Our idea when running this experiment was to obtain a clear view of different group behaviors for each week of the course. For instance, we expected to detect a cluster containing students who show high activity, or a cluster clearly focused on evaluative assessments. Clustering in the different stages should be consistent to allow comparison among weeks – same label should indicate same behavior –. In this way, we could check temporal evolution. For instance, a sequence could indicate that a student begins in the 'high activity' group and then changes to the 'assessment oriented' during the following week.

Regarding the clustering technique, we opted to select k-Means as clustering algorithm for being commonly used [36]. In this technique, clusters are created based on the distance to a centroid, which constitutes the center of this cluster. Interpretation of centroid data will allow to assign meaningful labels to the clusters obtained.

To implement this approach, we needed to consider the number of clusters in advance. Literature indicates 4-5 clusters is a common number for this kind of environments[33], [37]. Our tests will be done considering k=4 as a potentially interesting number of clusters, as our scenario can be considered similar to [33] in terms of course methodology – flipped –, course duration and LMS data as main data source.

The translation form cluster labels to meaningful naming will be done based on centroid information analysis. Centroids will be kept to allow proper interpretation and appropriate labelling of the cluster. We must keep in mind that the sequence we look for would be useless having non-meaningful states such as 'cluster *n*'. Analyzing centroids will allow us to interpret the meaning of the cluster, and label a particular group with meaningful attributes such as 'high activity' or 'assessment oriented'.

Besides individual week cluster labeling, centroids will also be used to provide cluster coherence among different weeks. In other words, the same detected behavior should map to the same label, even among different weeks. For clarity purposes, we will try to keep the same number of clusters – and interpretation if possible - for all weeks. This analysis will be done manually, as human intervention is needed to properly interpret cluster results, and to provide coherence among weeks.

3.2 Establishing and mining sequence of actions

In order to properly model student behavior through the course, we map the information obtained through the clustering process into a sequence of situations. For instance, assuming the clustering process leaves three groups, labeled as 'Low activity' (1), 'Quiz oriented'(2) and 'Low login' (3), a sequence such as (1,2,3,3,3,3) would mean the user begins by performing low (1), she then has a quiz-oriented week (2), and after that four weeks with low login activity (associated to the 3333 in the sequence.

This sequence will be treated as a sequence of states that will be mined with sequence mining tools. Data will be split between passing and failing students, in order to detect differences in support for the most relevant patterns. As noted in the theoretical framework, different techniques exist [12]. We selected generalized sequential pattern

(GSP) algorithm for being commonly used [12], due to the existence of proven implementations, and considering performance is not a constraint (n=337 students). Our focus will be set on the interpretation of results, and not on the algorithm itself. We will keep sequence and associated support for each of the groups under analysis. An open-source implementation will be used [38].

4 Results

4.1 Clustering

The clustering process was carried out with k-means. x-means was previously used to explore the potential number of interesting clusters and confirmed k=4 was a proper number, which could accommodate clustering results for the different weeks.

The centroid analysis provided also interesting results. While almost all studies suggest there is a low activity group and a high activity group, some other behaviors exists. For instance, gamers who try to game the system and perform high number of quizzes but do not follow lessons in such a way. Table 2 shows centroid data for the first two weeks:

Table 2. Sample of centroid data (first two weeks)

Week	Cluster ID	Lectures	Quizzes	Evaluative	Non assessed	Suggested	Login sessions
1	cl0	18,93	9,64	8,93	24,86	0,86	5,75
1	cl1	2,51	1,27	1,50	1,03	0,07	1,78
1	cl2	15,12	6,24	6,79	12,30	0,54	5,06
1	cl3	9,31	4,06	5,07	4,50	0,27	4,31
2	cl0	3,32	1,53	2,06	1,30	0,17	2,52
2	cl1	18,21	6,94	9,14	11,30	0,87	6,54
2	cl2	35,82	9,82	10,45	22,82	2,55	7,18
2	cl3	11,28	4,24	5,53	4,43	0,42	5,29

As Table 2 shows, for our case study the group with higher values for login sessions per week shows also higher activity in the different categories. This finding suggests that the clustering is really showing the amount of work performed by the student. For instance, the low on-line activity group – which for Table 2 would be cl1 for week 1 or cl0 for week 2 - is always present, showing low performance in all items (lectures, quizzes, …). The other three clusters are graded according to their amount of work. For this reason, we identified the clusters as low (L), medium (M), high (H) and extreme (E) activity.

4.2 Segmentation

The results of the clustering process were compiled into sequences for each of the students. As stated, we segmented the dataset into two groups according to final academic result. This segmentation is performed in order to detect differences in patterns between the passing and the failing group.

Once segmentation is performed, GSP algorithm is run on each of the resulting sequence dataset.

4.3 Relevant patterns for the failing group

GSP algorithm run on the failing group dataset looking for sequences with a minimum support of 0.6. Among the resulting sequences, Table 3 shows those with greater support. As it could be expected, long periods of low-activity are present among those students who finally fail the subject. It is also noticeable that most sequences include one or more low activity periods (L).

Table 3. Sequences that are present in higher percentage among the failing group (Top 10)

Sequence	% failing students showing sequence
L	0.95
L - L	0.88
M	0.84
L - L - L	0.80
M - L	0.74
M - M	0.73
L - L - L - L	0.72
L - L - L - L - L	0.69
M - L - L	0.68
H	0.68

4.4 Relevant patterns for the passing group

We carried out the same process for the passing group. Again, we used 0.6 as minimum support. Most common sequences and their support are shown in Table 4:

Table 4. Sequences that are present in higher percentage among the passing group (Top 10)

Sequence	% passing students showing sequence
M	0.97
M - M	0.92
H	0.87

M - M - M	0.84
L	0.84
M - H	0.83
M - L	0.81
H - H	0.77
H - M	0.76
M - M - M - M	0.74

4.5 Comparison of patterns

Despite Tables 3 and 4 already show noticeable differences, we perform a specific search to determine the support for sequences in the failing group inside the passing group. In this case, support for a specific sequence can be below 0.6, as it can be common only in the failing group. We also sort the table according to this difference in support. Results are shown in Table 5:

Table 5. Support for Top-10 failing sequences among the passing group

Sequence	% failing students showing sequence	% passing students showing sequence	Difference
L - L - L - L - L	0.69	0.3	0.38
L - L - L - L	0.72	0.42	0.29
L - L - L	0.8	0.56	0.24
L - L	0.88	0.68	0.2
L	0.95	0.84	0.11
M - L - L	0.68	0.62	0.05
M - L	0.74	0.81	-0.06
M	0.84	0.97	-0.13
M - M	0.73	0.92	-0.18
H	0.68	0.87	-0.19

5 Discussion

Results in Tables 3,4 and 5 allow us to answer the RQ raised in the introduction. The application of the described method to our case study has proven valid to detect differences in behavior between the two groups under study: students who pass and those who fail behave differently. That means that behavior is kept in the state sequence. We deepen into the process itself and its results for this particular case.

The process described uses SPM to mine sequences generated through clustering. Clustering is the initial stage, and in our case, produced pure activity groups. Students who show higher volume of activity show it on all kind of items. In particular, and for instance, students with higher number of login sessions show also higher lecture activity and higher quiz completions.

This fact can be compared with other clustering analysis present in the literature. A similar scenario – university course, first year engineering, computer science topic and flipped design – can be found in [33]. In this case, a clustering process is also performed aimed to detect student strategies. As a key different to our study, assessment data is included into the clustering process. The study detects four initial clusters, two oriented to assessments – formative or summative – and two related to content – one more oriented to video lecture and one to reading materials-.

[37] provides also a study of two courses focused on activity. Four clusters are also identified, being two of them clearly identified as highly active and low active. This study includes only activity, gathered also from a LMS platform. Clusters showing activity show also higher activity for the values considered – in this case, resource view, forum view and forum participation – with the exception of one single group showing least forum activity.

In the MOOC environment, this kind of studies is also present to analyze engagement in courses [39], [40]. We believe this scenario shows relevant differences in behavior to our case. This reason explains different pattern detection, such as samplers or returners. We believe this behavior is common in MOOC courses, but not so much in regular university courses.

Regarding behavioral sequences, Table 3 shows more common behavior for failing students. 95% of them show low activity in at least one of the weeks, and almost 90% in two consecutive weeks. A week with low activity is present in top-5 sequences, and in 7 out of 8 of those sequences with a minimum of two items. For the passing group, the most common situation is to follow medium or high engagement combinations.

Differences become more evident if we have a look at Table 5. While almost 70% of failing students show 5 consecutive weeks of low engagement with content, only 30% of passing students show this behavior. At the same time, it is also noticeable that sequences showing two or more low access weeks show the higher differences with passing students.

In fact, and according to Table 4, sequences which include at least one week of medium or high activity are more common in passing students. That indicates that passing students perform higher volume of online activity. From a pedagogical point of view, the interpretation of results in Table 5 shows that while one disengagement week makes no major difference, failure probability increases as the number of disengaged weeks does. In other words, the continuous detection of low online activity can indicate the student is more likely to fail.

Besides specific interpretation of this case study, we believe the approach provides an interesting insight to log analysis and its transformation into a process model. Log processing is simple and no specific restrictions have been imposed to algorithms for clustering or SPM. We have also indicated a potential selection of specific tools, such as k-means clustering and GSP.

We consider two parameters make the approach particularly attractive. First, the system is easy to implement. Only data obtained from log is needed. No sociological, preexisting or data coming from other record systems is needed. Classification does not require prior categorization as the work is done according to individual behavior in relation to the group.

Second, the process takes into account not only static values, but a dynamic picture of the student. A low engagement week may not be relevant, but it can become a problem if two consecutive weeks – or more – are accumulated. In short, the model is capturing a relevant part of the learning process. And this learning process is not a static picture. Analysis can only be done when the process is seen in perspective. In this sense, we consider the method depicted can provide a new and interesting insight to many problems related to research in learning analytics.

6 Conclusions

The process described generates a sequence of states based on behavioral clustering. This sequence is then analyzed in order to detect differences between two groups of students (passing and failing). Results show that behavior is effectively different and that this difference is contained in the sequences analyzed.

While we have focused on the method itself, we envision two groups of potential applications of this process. First, the use as a potential failure indicator. Second, as a detector of points of disengagement during the course, which could lead to curriculum redesign.

In order to implement potential applications in any of these groups, a previous extension of the study would be advisable. This extension can be done to successive editions of the same course, or to other courses, opening interesting research lines.

In the first case, results could be potentially extended to analyze forthcoming editions of the same course. While the issue of portability has not been addressed, we believe the study could open a different approach in prediction processes. The method could be carried out on a per-week basis as described and raise alarms when sequences indicating failure are detected.

Regarding portability to other courses, it would be interesting to compare results among courses, and even deep into the pedagogical implications of course type and methodology in results. For instance, results could help to detect differences not only in terms of passing and failing groups, but can detect differences among on-campus or on-line courses, methodologies - i.e. blended, flipped, MOOC- or even topic – STEM vs social -. This extensions could allow deeper comparison of results with some references analyzed in this study (for instance [37], [39], [41] for the MOOC case).

Finally, and while our interest has remained on pure non-grading activity data, the method could also be extended to other scenarios, and include other aspects in the clustering process, such as sociological data or even impact of specific learning activities (i.e. quizzes or evaluative assessments). Authors are open to collaborate in these open scenarios, in particular looking for practical applications and contributions to better learning designs.

36

References

[1] W. Lasanthika and W. Tennakoon, “Assessing the Adoption of Learning Management Systems in Higher Education,” *GATR Glob. J. Bus. Soc. Sci. Rev.*, vol. 7, no. 3, pp. 204–209, Sep. 2019, doi: 10.35609/gjbssr.2019.7.3(5).

[2] H. Coates, R. James, and G. Baldwin, “A critical examination of the effects of learning management systems on university teaching and learning,” *Tert. Educ. Manag.*, vol. 11, no. 1, pp. 19–36, Mar. 2005, doi: 10.1007/s11233-004-3567-9.

[3] F. Castro, A. Vellido, À. Nebot, and F. Mugica, “Applying data mining techniques to e-learning problems,” *Stud. Comput. Intell.*, 2007, doi: 10.1007/978-3-540-71974-8_8.

[4] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, “Data mining algorithms to classify students,” in *Educational Data Mining 2008 - 1st International Conference on Educational Data Mining, Proceedings*, 2008.

[5] C. Vieira, P. Parsons, and V. Byrd, “Visual learning analytics of educational data: A systematic literature review and research agenda,” *Comput. Educ.*, vol. 122, 2018, doi: 10.1016/j.compedu.2018.03.018.

[6] R. S. J. D. Baker and K. Yacef, “The State of Educational Data Mining in 2009 : A Review and Future Visions,” *J. Educ. Data Min.*, 2009, doi: http://doi.ieeecomputersociety.org/10.1109/ASE.2003.1240314.

[7] G. Siemens, “Learning analytics: Envisioning a research discipline and a domain of practice,” in *ACM International Conference Proceeding Series*, 2012, doi: 10.1145/2330601.2330605.

[8] R. Ferguson, “Learning analytics: drivers, developments and challenges,” *Int. J. Technol. Enhanc. Learn.*, vol. 4, no. 5, pp. 304–317, 2012, doi: 10.1504/IJTEL.2012.051816.

[9] A. Bogarín, R. Cerezo, and C. Romero, “A survey on educational process mining,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 1, p. e1230, Jan. 2018, doi: 10.1002/widm.1230.

[10] M. A. Ghazal, O. Ibrahim, and M. A. Salama, “Educational process mining: A systematic literature review,” in *Proceedings - 2017 European Conference on Electrical Engineering and Computer Science, EECS 2017*, Jul. 2018, pp. 198–203, doi: 10.1109/EECS.2017.45.

[11] W. Van Der Aalst, “Process mining: Overview and opportunities,” *ACM Transactions on Management Information Systems*, vol. 3, no. 2. pp. 1–17, Jul. 2012, doi: 10.1145/2229156.2229157.

[12] P. Fournier-Viger, J. Chun, W. Lin, R. U. Kiran, Y. S. Koh, and R. Thomas, “A Survey of Sequential Pattern Mining,” 2017.

[13] M. Salehi *et al.*, “Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering,” doi: 10.1007/s10639-012-9245-5.

[14] L. K. M. Poon, S. C. Kong, M. Y. W. Wong, and T. S. H. Yau, “Mining sequential patterns of students’ access on learning management system,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10387 LNCS, pp. 191–198, doi: 10.1007/978-3-319-61845-6_20.

[15] P. Fournier-Viger, J. Li, J. C. W. Lin, T. T. Chi, and R. Uday Kiran, "Mining cost-effective patterns in event logs," *Knowledge-Based Syst.*, vol. 191, Mar. 2020, doi: 10.1016/j.knosys.2019.105241.

[16] J. A. Martínez-Carrascal, D. Márquez Cebrián, T. Sancho-Vinuesa, and E. Valderrama, "Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course," *Comput. Appl. Eng. Educ.*, Mar. 2020, doi: 10.1002/cae.22229.

[17] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," 2015. Accessed: Apr. 11, 2019. [Online]. Available: http://www.educationaldatamining.org/EDM2015/proceedings/short392-395.pdf.

[18] W. Van Der Aalst, "Process mining: Overview and opportunities," *ACM Transactions on Management Information Systems*, vol. 3, no. 2. Jul. 2012, doi: 10.1145/2229156.2229157.

[19] N. Tax, N. Sidorova, R. Haakma, and W. M. P. van der Aalst, "Mining local process models," *J. Innov. Digit. Ecosyst.*, 2016, doi: 10.1016/j.jides.2016.11.001.

[20] A. Bolt, M. de Leoni, and W. M. P. van der Aalst, "Process variant comparison: Using event logs to detect differences in behavior and business rules," *Inf. Syst.*, vol. 74, pp. 53–66, May 2018, doi: 10.1016/j.is.2017.12.006.

[21] S. Rizvi, B. Rienties, J. Rogaten, and R. F. Kizilcec, "Investigating variation in learning processes in a FutureLearn MOOC," *J. Comput. High. Educ.*, vol. 32, no. 1, pp. 162–181, Apr. 2020, doi: 10.1007/s12528-019-09231-0.

[22] D. Fahland and W. M. P. van der Aalst, "Simplifying Mined Process Models: An Approach Based on Unfoldings," Springer, Berlin, Heidelberg, 2011, pp. 362–378.

[23] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings - International Conference on Data Engineering*, 1995, pp. 3–14, doi: 10.1109/icde.1995.380415.

[24] C. H. Mooney and J. F. Roddick, "Sequential pattern mining - Approaches and algorithms," *ACM Computing Surveys*, vol. 45, no. 2. Feb. 2013, doi: 10.1145/2431211.2431218.

[25] A. P. Wright, A. T. Wright, A. B. McCoy, and D. F. Sittig, "The use of sequential pattern mining to predict next prescribed medications," *J. Biomed. Inform.*, vol. 53, pp. 73–80, Feb. 2015, doi: 10.1016/j.jbi.2014.09.003.

[26] A. S. Johnsha Ali, "Frequent Sequential Patterns (FSP) Algorithm for Finding Mutations in BRCA2 Gene," *Int. J. Recent Technol. Eng.*, no. 3, pp. 2277–3878, 2019, doi: 10.35940/ijrte.C6507.098319.

[27] A. Bogarín, R. Cerezo, and C. Romero, "Discovering learning processes using inductive miner: A case study with learning management systems (LMSs)," *Psicothema*, vol. 30, no. 3, pp. 322–329, 2018, doi: 10.7334/psicothema2018.116.

[28] J. K. Tarus, Z. Niu, and A. Yousif, "A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining," *Futur. Gener. Comput. Syst.*, vol. 72, pp. 37–48, Jul. 2017, doi: 10.1016/j.future.2017.02.049.

[29] R. Venant, K. Sharma, P. Vidal, P. Dillenbourg, and J. Broisin, "Using sequential pattern mining to explore learners' behaviors and evaluate their correlation with performance in inquiry-based learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol.

10474 LNCS, pp. 286–299, doi: 10.1007/978-3-319-66610-5_21.

[30] J. S. Kinnebrew, K. M. Loretz, and G. Biswas, "A Contextualized, Differential Sequence Mining Method to Derive Students' Learning Behavior Patterns," *JEDM | J. Educ. Data Min.*, vol. 5, no. 1, pp. 190–219, May 2013, doi: 10.5281/ZENODO.3554617.

[31] J. Han and J. Pei, "Mining frequent patterns by pattern-growth," *ACM SIGKDD Explor. Newsl.*, vol. 2, no. 2, pp. 14–20, Dec. 2000, doi: 10.1145/380995.381002.

[32] M. Zhou, Y. Xu, J. C. Nesbit, and P. H. Winne, "Sequential pattern analysis of learning logs: Methodology and applications," in *Handbook of Educational Data Mining*, 2010.

[33] J. Jovanović, D. Gašević, S. Dawson, A. Pardo, and N. Mirriahi, "Learning analytics to unveil learning strategies in a flipped classroom," *Internet High. Educ.*, vol. 33, pp. 74–85, Apr. 2017, doi: 10.1016/j.iheduc.2017.02.001.

[34] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O. R. Zaane, "Clustering and sequential pattern mining of online collaborative learning data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 6, pp. 759–772, Jun. 2009, doi: 10.1109/TKDE.2008.138.

[35] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS," *IEEE Trans. Learn. Technol.*, 2017, doi: 10.1109/TLT.2016.2616312.

[36] A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *IEEE Access*, 2017, doi: 10.1109/ACCESS.2017.2654247.

[37] I. P. Ratnapala, R. G. Ragel, and S. Deegalla, "Students behavioural analysis in an online learning environment using data mining," in *2014 7th International Conference on Information and Automation for Sustainability: "Sharpening the Future with Sustainable Technology", ICIAfS 2014*, Mar. 2014, doi: 10.1109/ICIAFS.2014.7069609.

[38] P. Fournier-Viger *et al.*, "The SPMF open-source data mining library version 2," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, doi: 10.1007/978-3-319-46131-1_8.

[39] R. Ferguson and D. Clow, "Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs)," in *ACM International Conference Proceeding Series*, Mar. 2015, vol. 16-20-March-2015, pp. 51–58, doi: 10.1145/2723576.2723606.

[40] L. Shi and A. I. Cristea, "In-depth exploration of engagement patterns in MOOCs," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Nov. 2018, vol. 11234 LNCS, pp. 395–409, doi: 10.1007/978-3-030-02925-8_28.

[41] M. Khalil and M. Ebner, "Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories," *J. Comput. High. Educ.*, vol. 29, no. 1, pp. 114–132, Apr. 2017, doi: 10.1007/s12528-016-9126-9.

Artículo 7: Using *Process mining* to determine the relevance and impact of performing non-evaluative quizzes before evaluative assessments.

In Proc. of Learning Analytics Summer Institute Spain 2022. Learning Analytics: Here to stay, Salamanca, Spain, June 20-21, 2022, http://ceur-ws.org/Vol-3238/paper7.pdf

Scopus: SJR 2021: 0.228, Category: Computer Science (miscellaneous)

Nota: Las referencias numéricas a figuras y tablas contenidas en el artículo que se presenta a continuación deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

Using Process Mining to determine the relevance and impact of performing non-evaluative quizzes before evaluative assessments

Juan Antonio Martínez-Carrascal[1] and Teresa Sancho-Vinuesa[1]

[1] Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain

Abstract

Different data mining techniques are commonly used to extract behavioural patterns from activity logs. However, they often offer static views and lack interpretability. In this paper, we describe the procedure for obtaining meaningful learning processes for a Maths course based on Moodle logs. Log data is transformed into process models in order to be analysed. We use the method described to specifically analyse the potential relevance and detailed impact of performing non-evaluative assessments before evaluation tests on a mathematics online university course. Our preliminary results outline statistical-ly significant differences between those students who practice before submitting evaluative tests and those who decide to proceed to evaluation directly. Besides this particular result, the process can be expanded to detect behavioural patterns and differences in learning processes among groups of students, in particular, in online learning scenarios.

Keywords

Process Mining, Behavioural patterns, Formative assessment, Academic performance, Learning Process.

1. Introduction

Process mining constitutes a discipline between business process management and data mining [1]. It constitutes a discipline on its own, that was initially focused on industrial processes and business management analysis. However, its potentialities and interpretability of results have interested other related fields, such as learning analytics.

In this paper, we use process mining to get meaningful information from Moodle logs, aimed to obtain relevant outcomes that can help to improve the learning process. This introductory work aims to set the basics, envision the potentiality and shown one potential application.

As specific application, we analyse the impact of the student behaviour before taking evaluative tests. In particular, we focus on the impact of taking online non-evaluative quizzes before carrying out evaluative tests. The focus is not set on the simple execution of the formative quiz, but in whether it is executed before taking the evaluative activity. We choose this specific aspect due to the existence of previous research work on a similar course we will work on whose results indicate that prac-tising has an impact[2]. The course uses classical statistical techniques. However, we aim to use this experience to develop a framework allowing deeper analysis of learn-ing processes and in particular, of behavioural learning patterns.

The rest of this paper is structured as follows. Section 2 sets the theoretical frame-work, reviewing the basics behind process mining and its link to learning analytics. The section ends by stating the research question covered in this paper. Once done, Section 3, provides details on the course under analysis, and also on the steps carried out to conform a meaningful process based on Moodle log data. Section 4 shows the results obtained. Section 5 discusses these results, and finally, Section 6 concludes with interesting open lines for future research.

Learning Analytics Summer Institute Spain (LASI Spain) 2022, June 20–21, 2022, Salamanca, Spain
EMAIL: jmartinezcarra@uoc.edu (A. 1); tsancho@uoc.edu (A. 2)

CEUR Workshop Proceedings (CEUR-WS.org)

2. Theoretical framework

Process mining is a discipline oriented to 'discover, monitor, and improve real processes by extracting knowledge from event logs'[1]. Although it was originally focused on business-oriented models, the range of applications is broad. In particular, and in recent years, the learning analytics community has shown growing interest on this topic.

Three main categories can be found into process mining: process discovery, conformance checking and process enhancement. Process discovery consists of identifying underlying processes hidden in log data. Given a log file, with information regarding real process executions, three main characteristics are extracted: a group of cases, a set of actions, and a timestamp for each execution associated. Each case is finally an execution of the process, which consists of a set of actions that take place in a specific order at given timestamps.

While a single execution can be represented through a graph, integrating a huge number of executions is far more complex and makes interpretation difficult. Directly follows graphs (DFGs) constitute a first level of representation, which include a set of states (based on the actions in the log) linked by arcs, but with intrinsic limitations [3]. When the number of cases is high, the DFG turns into a spaghetti-like flow, which makes interpretation more complex. The analysis of these DFGs normally requires a filtering process, oriented to get either commonly performed activities, or common transitions. Different tools both commercial [4, 5]and non-commercial [6, 7] are available for processing these flows.

From a formal perspective, Petri nets [8]are behind the theoretical foundation of these models. This kind of nets can both represent processes and simulate iterations on it, by placing a token on the start point and simulating process executions. Petri nets can be designed from a theoretical perspective, by using BPMN language, common to generic process analysis tools [9]. However, the potential behind process mining is to automatically obtain these nets, by analysing logs. This process is called process discovery [10] and constitutes a discipline on its own. While theoretical models represent what should happen, the discovered models show what is actually happening. For a given log, there is neither a single technique to infer models, nor a single resulting model. In fact, if the number of cases and associated actions increase, the problem can be computationally challenging. Specific algorithms try to reduce computational complexity at the cost of providing models which lack some characteristics of more complex models.

Characteristics of discovered models include fitness, precision, generalisation and simplicity[11]. Fitness indicates to what extent the model reflects all the activity seen in the event log. Precision refers to the fact that the model does not allow for behaviors that are completely unrelated to the log from which it derives. As logs are finite, generalisation indicates the ability of the model to generalise similar behaviours. Finally, simplicity refers to the fact of the model being as simple as possible. There is always a balance between these characteristics. For instance, a model can potentially be made simpler at the cost of not reflecting all behaviour, or at introducing potentially unrelated behaviour.

Specific algorithms oriented to practical applications are included into two main categories: heuristic and fuzzy [11]. Heuristic miner is focused on reflecting the main behaviour reflected in the log. Fuzzy miner is used when the number of activities and cases is particularly high, and the behaviour reflected in the log is unstructured. However, other algorithms are also used. Alpha-miner algorithm is also commonly used, and some articles on Educational data mining make use of inductive miner[12].

The generated models can be compared with a log associated to the same process. This comparison is called conformance checking [10, 11] and indicates to what extent the information contained in the log is consistent with a given model. From a practical perspective, conformance aims to detect where the logged behaviour diverges from the model. While different possibilities exist to establish conformance indexes, the most common approach is to check alignments. It is considered an accurate technique, which overcomes the limitations of previous algorithms. This technique tries to map a trace that does not necessarily adjust to the model and find either where the log differs from the model or where the model differs from the log.

Besides conformance indicators, the disagreements between the log and the model set the basics for process enhancement. The idea is to improve the model based on the information gathered from real executions of the process. With this idea in mind, and focusing on learning scenarios, we approach the improvement of learning processes,

A compilation of works classified by education domains can be found in [13] . This compilation reveals works in different unrelated areas, such as curriculum min-ing, analysis of student registration processes or professional training. To cite some relevant works on aspects linked to student's behaviour in courses, [14] remarks the potentialities of process mining to detect behavioural patterns in learning environments focusing on conformance analysis in a blended learning course. [15] uses process mining oriented to improve the learning experience of students enrolled in MOOC courses. [16] analyses online assessment data and recently [17] focuses on assessments in self-regulated learning environments.

However, and to the best of our knowledge, there are no references in the literature that focus on the impact of non-evaluative assessments through process mining on academic performance. Existing literature focusing on this impact are based on classical techniques. Results indicate that practising improves overall outcome [2, 18]. This last work, focused on the same course we focus on, also indicates the potentiality of using formative quizzes, concluding that there is a link between questionnaire scores and overall performance. However, the link with the order of the activities is not considered in this article.

Given this background, and linking these ideas together, we aim to analyse the learning process of a mathematical course, based on the activity logged, and the analysis of the underlying process. Specifically, we raise the following research question:

RQ: Does completing non-evaluative online quizzes prior to performing evaluative assessments help students pass the course? What is the quantitative impact of an af-firmative response?

The next section provides additional details on the course and precises the methods used.

3. Methodology

3.1. Course description

The course under analysis is an introductory course to Mathematics. This course is offered online and covers mathematical requirements for students entering Computer Science and Multimedia related degrees. The course is structured into 11 learning units, covering from Calculus basics to differential calculus. The course lasts for one academic semester.

The course is offered completely online. Students have a course landing page, where they can access materials in different formats, post doubts and contact the instructor. However, the core of the course is contained into a Moodle environment, where non-evaluative quizzes and evaluative activities are offered. Each unit has a similar structure. Students are supposed to begin practising non-evaluative questionnaires. Upon submission, they can check the answers. Additional retries are possible. Successive submissions are provided with slight modifications, allowing for extra practice. There is neither penalty for failing these tests, nor for not submitting them. For each of the units, students should carry out an evaluative assignment for the unit under consideration.

Both kinds of tests are opened on a periodical basis (weekly or bi-weekly, depending on the unit). Non-evaluative quizzes are available two days before evaluative activities, with the aim of increasing student practice. Once a non-evaluative quiz is open, it remains open for the rest of the course, and students can check it later - and even resubmit -. When it comes to evaluative tests, they usually have a deadline, which is usually a week after they are offered. However, when a student opens an evaluative test, she must submit it within 24 hours. Figure 1 shows the schematic of the expected process for a given unit of the course.

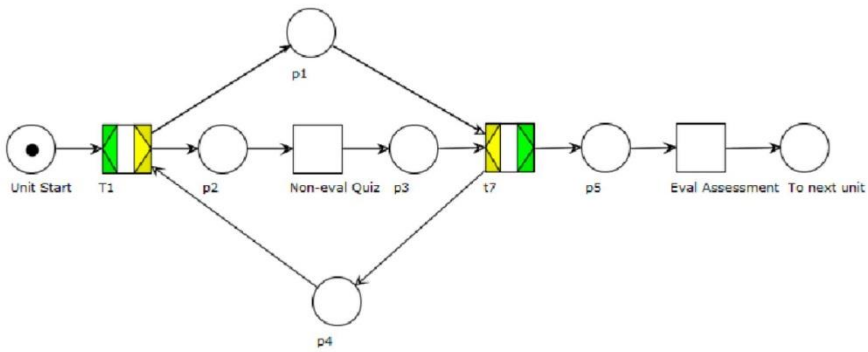


Figure 1: Expected student behaviour within a course unit.

Figure 1 reflects the Petri Net associated to the expected behaviour for a given unit. As the unit starts, the student can choose to perform the non-evaluative quiz (going through the place, P2), or simply skip (through P1). In any case, he can still decide to take non-evaluative quizzes again, as many times as needed. When the student is confident take the evaluative assessment, she should take it and proceed. This is the last step covered before entering a new unit.

It is also noticeable that satisfaction surveys carried out on different editions of the course show a high level of satisfaction. The passing rations for the course under analysis is normally over 80%. In particular, the edition analysed in this paper has 144 students. 121 of them pass the course, while 12 fail and 11 withdraw from the course.

3.2. Fitting data for process discovery

Besides the theoretical perspective of the process described in Figure 1, we aim to obtain a process model based on real executions from the students. To do so, we need to map a case identifier, an activity and a timestamp. Before performing the mapping, we have a look at the Moodle data we are working with. Table 1 reflects this data, where users have been anonymised.

Table 1
Sample of Moodle data gathered

Date	User	Activity	Scope	Description	Detail
18/02/2021 17:05	User1	Cuestionario: Nombres PRÀCTICA	Cuestionario	Modulo de curso visto	The user with id '125433' viewed the 'quiz' activity with course module id '25433'.
18/02/2021 17:06	User2	Cuestionario: Qüestionari inicial sobre el funcionament del curs	Cuestionario	Intento de cuestionario visualizado	The user with id '152301' has viewed the attempt with id '1347943' belonging to the user with id '152301' for the quiz with course module id '87859'.

18/02/2021 17:07	User2	Cuestionario: Qüestionari inicial sobre el funcionament del curs	Cuestionario	Intento del cuestionario revisado	The user with id '152301' has had their attempt with id '1347943' re-viewed by the user with id '152301' for the quiz with course module id '87859'.
18/02/2021 17:07	User2	Cuestionario: Qüestionari inicial sobre el funcionament del curs	Cuestionario	Intento enviado	The user with id '152301' has sub-mitted the attempt with id '1347943' for the quiz with course module id '87859'.

A preliminary view indicates that we can recover time information from the Date column on Table 1. We will also consider the User column as a case identifier. In other words, we consider that each student performs an execution of the learning process all through the course. Finally, and regarding the activities, we decided to adjust the level of granularity to adjust potential findings to reflect the process depicted in Figure 1.

In order to do so, and to provide shorter mnemonics we ordered the different learning units and assigned them a short name. The first learning unit will be identified as 1. Non evaluative assessments will be marked as PR, indicating that they are intended to practice. They correspond to non-evaluative quizzes and – as indicated – are expected to be covered before evaluative assessments (shortened as EVAL).

In addition, and to allow for extra granularity, we also distinguish between the different possibilities Moodle adopts for a given activity. In particular, and thinking about assessments, we can have INIT (when the user initiates a submission), SUBMIT, when the student effectively submits, and REVIEW, when the user checks a form already sent. All these transformations are common in case studies that use Moodle[12, 17], and in general should be performed when dealing with LMS data. In our case, they have been performed through Python scripts, providing an output which is suitable for process analysis. Table 2 provides a sample of the resulting output.

Table 2
Sample of Moodle log adapted for process analysis.

Timestamp	Case ID	Activity
25/02/2021 21:45	User2	M1-PR-INIT
25/02/2021 22:33	User2	M1-PR-SUBMIT
26/02/2021 22:34	User2	M1-PR-INIT
26/02/2021 23:09	User2	M1-PR-SUBMIT
28/02/2021 11:12	User2	M1-EVAL-INIT
28/02/2021 21:27	User2	M1-EVAL-SUBMIT

According to this sample, User2 has tried non-evaluative tests twice, before attempting the evaluative assessment. With this information, we can proceed to graph process data linked to our course under analysis.

4. Results

Once the data for the course has been pre-processed according to the procedure de-scribed in Section 3, it was load into Celonis [4] for detailed analysis. Figure 2 shows an initial view of the process.

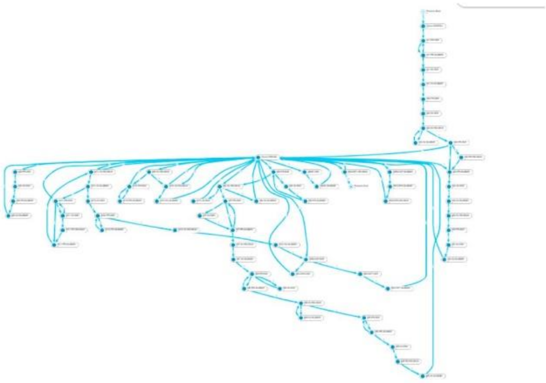


Figure 2: Overview of the discovered learning process model.

While the image is only provided as a sample, we can clearly see that the process begins in a quite linear way, but after the initial week, it turns into a spaghetti-like process. This is in fact typical of complex processes with a high number of cases and activities. It can also be filtered based on specific activities, transitions or groups of students. In our case, we have filtered the view in Figure 2. This filtering confirms the results of previous analysis based on the same course indicating the relevance of practicing before evaluative activities [2]. For this reason, we have filtered the log in order to determine to what extent students cover the practice tests before the evaluative ones. Table 3 indicates the number of students performing non-evaluative quizzes before evaluation for each unit, classified according to their final course result.

Table 3

Distribution of students performing non evaluative quizzes before evaluative assessments for the different modules based on final course mark (pass, fail, withdrawn).

Actions	Pass (121)	Fail (12)	WD (11)
M1-PR before M1-EVAL	101	10	3
M2-PRM2-EVAL	104	9	3
M3-PR before M3-EVAL	98	6	1
M4-PR before M4-EVAL	97	5	1
M5-PR before M5-EVAL	95	2	1
M6-PR before M6-EVAL	108	3	1
M7-PR before M7-EVAL	90	1	1
M8-PR before M8-EVAL	88	0	1
M9-PR before M9-EVAL	83	0	1
M10-PR before M10-EVAL	88	0	1

The analysis of data in Table 3 can be analysed in terms of probability. We compute the conditional probability of passing the course, considering whether the students perform non evaluative quizzes before evaluative assessments or not. In addition, we also determine statistical significance through χ^2 test, which is shown in Table 4.

Table 4
Probability to pass the subject based on whether the student carries out formative quizzes before evaluative assessements in each module. Legend: ns: non-significant - *: p<0.05- **: p<0.01- ***: p <0.001 - ****: p<0.0001

Actions	Pass (121)	Fail (12)	WD (11)
M1	0.85	0.8	ns
M2	0.90	0.61	**
M3	0.92	0.62	***
M4	0.94	0.59	****
M5	0.97	0.57	****
M6	0.96	0.41	****
M7	0.98	0.60	****
M8	0.99	0.60	****
M9	0.99	0.63	****
M10	0.99	0.60	****

From a graphical perspective, Figure 3 reflects the probability to pass considering whether non-evaluative activities are performed in advance or not.

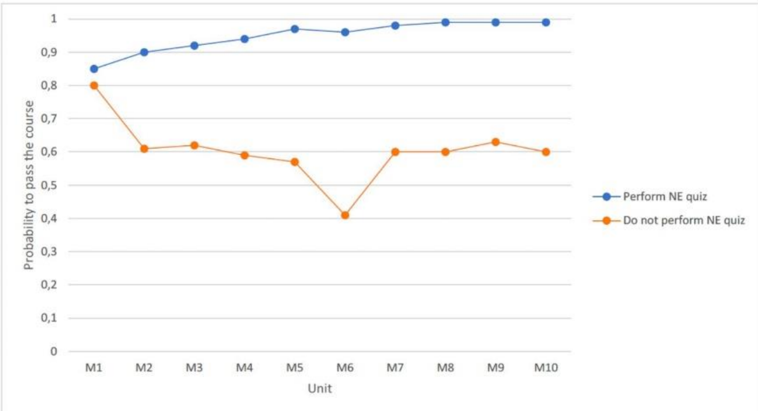


Figure 3: Differences in probabilities to pass the course.

5. Discussion

Results in the previous section allows us to answer the question stated in Section 2 of this paper. Performing non-evaluative quizzes before the evaluate assessment linked to the unit under study positively impacts the chances to pass the subject. Differences in passing ratios have been quantified and shown as statistically significant. While it shows not significant for the first unit, the differences get noticeable after the second unit. Difference in passing ratios can be as high as 55% in Unit 6.

The irrelevance of the first unit can even be considered normal from a conceptual perspective, even more considering the propaedeutic nature of the course. The first unit can be considered too introductory for some students, and they simply approach the evaluative activity directly. However, and from the second week, and consistently with the increasing complexity, the relevance of testing in advance also increases.

These findings are also consistent with previous findings linked to this course [2]. The referred work indicates that those students that submit practice tests show higher performance that those who choose not to submit them. While this work is based on classical techniques, the results agree with the process mining approach.

Regarding process mining, and while this article is part of a preliminary research stage, results are promising. Specific findings have emerged from the analysis performed. We aim to further explore its application to deepen into the analysis of student behaviour and learning paths. We believe it can outperform other tools, due to its focus on the process and not on static views, and also due to the ability to get meaningful results it provides.

Finally, we would like to remark the pedagogical implications. In particular, and based on these findings, we would encourage Math teachers in general to design non-evaluative quiz activities, and to remark the relevance of training before approaching evaluative tests. At the same time, students should be encouraged to always train before evaluation. Finally, and from a learning analytics perspective, we believe that those systems oriented to raise alarm should not only focus on the activities performed but on the process carried out.

6. Limitations and future work

This article shows an initial stage of our research regarding the potential use of process mining to analyse online courses. In this sense, results should be considered as those of a preliminary work. Nevertheless, the work carried out outlines the potential of process mining in learning analytics scenarios. Among its potential applications, we currently focus on the analysis of student behaviour along the course, and in par-ticular, on the learning paths they follow. This analysis can provide relevant insight into when and why students decide to abandon the expected course path and how this impacts academic outcomes. This is relevant both to detect behaviours that can potentially lead to unsuccessful course outcomes and to improve course design. Authors are open to collaborate in initiatives aimed to achieve these goals.

7. References

[1] van der Aalst W (2012) Process mining: Overview and opportunities. ACM Transactions on Management Information Systems 3:1–17

[2] Figueroa-Cañas J, Sancho-Vinuesa T (2021) Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous online calculus module. Interactive Learning Environments 29:. https://doi.org/10.1080/10494820.2018.1559864

[3] van der Aalst WMP (2019) A practitioner's guide to process mining: Limita-tions of the directly-follows graph. Procedia Computer Science 164:321–328. https://doi.org/10.1016/J.PROCS.2019.12.189

[4] Process Mining and Execution Management Software | Celonis. https://www.celonis.com/. Accessed 22 Apr 2022

[5] Process Mining and Automated Process Discovery Software for Professionals - Fluxicon Disco. https://fluxicon.com/disco/. Accessed 22 Apr 2022

[6] van Hee K, Oanea O, Post R, et al (2006) Yasper: A tool for workflow model-ing and analysis. Proceedings - International Conference on Application of Concurrency to System Design, ACSD 279–281. https://doi.org/10.1109/ACSD.2006.37

[7] start | ProM Tools. https://www.promtools.org/doku.php. Accessed 22 Apr 2022

[8] Peterson JL (1977) Petri Nets. ACM Computing Surveys (CSUR) 9:223–252. https://doi.org/10.1145/356698.356702

[9] Dijkman RM, Dumas M, Ouyang C (2008) Semantics and analysis of busi-ness process models in BPMN. Information and Software Technology 50:. https://doi.org/10.1016/j.infsof.2008.02.006

[10] Leemans SJJ, Fahland D, van der Aalst WMP (2018) Scalable process dis-covery and conformance checking. Software and Systems Modeling 17:. https://doi.org/10.1007/s10270-016-0545-x

[11] van der Aalst WMP (2014) Process mining in the large: A tutorial. In: Lecture Notes in Business Information Processing

[12] Bogarín A, Cerezo R, Romero C (2018) Discovering learning processes using inductive miner: A case study with learning management systems (LMSs). Psicothema 30:322–329. https://doi.org/10.7334/psicothema2018.116
[13] Bogarín A, Cerezo R, Romero C (2018) A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Dis-covery 8:e1230. https://doi.org/10.1002/widm.1230
[14] Mukala P, Buijs J, Leemans M, van der Aalst W (2015) Exploring Students' Learning Behaviour in MOOCs using Process Mining Techniques. Compu-ting Conference
[15] Umer R, Susnjak T, Mathrani A, Suriadi S (2017) On predicting academic performance with process mining in learning analytics. Journal of Research in Innovative Teaching & Learning 10:. https://doi.org/10.1108/jrit-09-2017-0022
[16] Pechenizkiy M, Trčka N, Vasilyeva E, et al (2009) Process mining online assessment data. In: EDM'09 - Educational Data Mining 2009: 2nd Interna-tional Conference on Educational Data Mining
[17] Cerezo R, Bogarín A, Esteban M, Romero C (2020) Process mining for self-regulated learning assessment in e-learning. Journal of Computing in Higher Education. https://doi.org/10.1007/s12528-019-09225-y
[18] Fitkov-Norris E, Lees B (2012) Online formative assessment: Does it add up to better performance in Quantitative modules? In: Proceedings of the 11th ECRM

Artículo 8: Evaluation of Recommended Learning Paths using *Process mining* and Log *Skeletons*: Conceptualization and Insight into a Mathematics Course

Martínez-Carrascal, J. A., Munoz-Gama, J., & Sancho-Vinuesa, T. (2023). Evaluation of Recommended Learning Paths using Process Mining and Log Skeletons: Conceptualization and Insight into an Online Mathematics Course. IEEE Transactions on Learning Technologies, doi: 10.1109/TLT.2023.3298035.

Scopus: SJR 2022: 1.14 Q1 Education, Q1 eLearning, Q1 Computer Science applications, Q1 Engineering (Miscellaneous)

WoS: JCR 2022: 3.7 Q1 Education & Educational Research Q2 Computer Science, Interdisciplinary Applications

Nota: El artículo siguiente es una versión previa de trabajo, por lo que se recomienda consultar la versión publicada del artículo (doi: 10.1109/TLT.2023.3298035). Las referencias numéricas a figuras y tablas contenidas en el artículo que se presenta a continuación deben ser consideradas en el contexto del propio artículo. Igualmente sucede con las referencias bibliográficas.

Evaluation of Recommended Learning Paths using *Process mining* and Log *Skeletons*: Conceptualization and Insight into a Mathematics Course

Juan Antonio Martínez-Carrascal[1] [0000-0002-7696-6050], Jorge Muñoz-Gama[2] [0000-0002-6908-3911] and Teresa Sancho-Vinuesa[1] [0000-0002-0642-2912]

[1] Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain

[2] Department of Computer Science, Pontificia Universidad Católica de Chile, Chile.

jmartinezcarra@uoc.edu

Abstract. Learning paths are deemed to be more effective at reaching learning outcomes that just working autonomously on a collection of tasks that the student completes at her own pace. However, there are limited references to methodologies that assess whether students effectively follow the path, quantify to what extent, and establish whether following a certain learning path is effective for a particular course.

This study presents a method to address these issues. The methodology is founded on process mining*process mining*, and more specifically on the potential of process *skeletons*. In addition to describing the method, it is applied to a real-world case study. On the basis of a mathematics course offered in an *online* university, we model the learning path indicated by the teachers, quantify the amount to which students' *online* behaviour fits the suggested learning path, and then determine if following the path is associated with improved academic performance.

Results indicate that few students strictly follow all the indications. However, the majority subscribes to them in broad terms. Following then correlates with final grades in our particular scenario. The technique described can be of significant assistance in validating the suitability of a given learning path, comparing other paths in terms of their impact on outcomes, and solving a multitude of *learning analytics* issues.

Keywords: *Process mining*, learning paths, academic performance, student activity modelling, learning process, intervention, process *skeleton*.

1 Introduction

‘‘Everyone knows a great deal about what he has to do to learn” (Mowrer, 1960). However, sixty years after this fact was publicly recognised, instructional and learning designers continue to work toward developing better learning materials and constructing effective learning paths. Despite these efforts, a significant number of students continue to fail academic courses. Generic or personalized learning paths emerge as just a tool to facilitate the learning process and enhance overall outcomes.

At the course level, learning paths are prescribed sequences of learning activities to be completed in a specific order. These paths should aid in passing the course under consideration if followed. Nonetheless, students deviate from the theoretical path. While learning designers can establish a baseline, forcing students to precisely adhere to a predetermined path is nearly impossible and impractical. In addition to following the primary path, students can also undertake ramifications based on the diversity of learner characteristics and the inherent complexity of the learning process.

This background highlights the primary focus of this study. We determine how closely a student is adhering to a prescribed learning path. This is particularly relevant from a *learning analytics* perspective. First, because this sets the basics to redirect student behaviour when significant deviations from the course are identified. Second, because it allows to evaluate, from a learning design standpoint, the relevance of following the path to achieving learning objectives. In other words, the applicability and effectiveness of a specific learning path can be formally validated.

To achieve this goal, it is necessary to establish a mathematical model of the learning path. Instead of simply considering the theoretical constraints the instructors might provide, our approach is to build a model based on real student activity that matches the theoretical learning path. Such model will respect the learning constraints of the theoretical path, while reflecting the activities that the students effectively perform. This model will be created and analysed by means of *process mining* artifacts (W. M. P. van der Aalst et al., 2003; W. M. P. van der Aalst & Weijters, 2004). The selection of these tools is done considering that *process mining* aims to "improve and support business processes" based on gathered log data (W. van der Aalst, Adriansyah, de Medeiros, et al., 2012).

Its focus on providing meaningful and clear information that would otherwise remain hidden in event logs has motivated its application to fields that range from industrial (W. M. P. van der Aalst et al., 2007) to medical (Rojas et al., 2016) environments. These techniques have also been used in education (Bogarín et al., 2018a). To cite specific applications, it has been used in MOOC courses to discover learning strategies (Maldonado-Mahauad et al., 2018), or to analyse learning trajectories uncovering behavioural patterns (Mukala et al., 2015b). References to process discovery on regular university courses can also be found (Bogarín et al., 2018b), including the analysis of curricular trajectories (Salazar-Fernandez et al., 2021a).

However, its potential for analysing learning paths at the course level has not yet been explored. Our research demonstrates the applicability of *process mining* for this analysis, particularly the usage of *skeletons*. *Skeletons* record the essential aspects of real-world learning path executions that are linked to theoretical paths. Once done, it will be possible to determine how closely additional executions correspond to this *skeleton*. We estimate the impact on academic outcomes by considering both the results attained by students in the course and how close they are to the intended path. In addition, students exhibiting relevant divergences may be considered intervention targets.

In addition to conceptualising, we put these principles into practise by analysing the activities of students enrolled in an *online* university's mathematics course. The course instructors provide a clear learning route consisting of a sequence of training and evaluative activities. However, students exercise their own autonomy and choose whether or not to take specific activities, review particular topics, or avoid others. We place particular emphasis on training and evaluative activities, since prior research associated with this course suggests their importance (Figueroa-Cañas & Sancho-Vinuesa, 2021a).

The suggested method can be used to analyse learning paths regardless of the course itself, giving to a better understanding of students' learning paths throughout the course and eventually aiding them in "doing what is necessary to learn."

The rest of this paper is structured as follows. Section 2 sets the theoretical framework and states formally the research questions. Section 3 exposes the approach in detail. It also provides details on the course that will be used for practical implementation. Section 4 presents the results, which will be discussed in Section 5. Finally, section 6 concludes and outlines future research lines that derive from this research.

2 Theoretical framework

In order to gain insight into the modelling and evaluation of learning paths we begin by considering aspects directly linked to learning paths (Section 2.1). Once done, and deriving from the consideration of *learning as a process* we analyse *process mining* and its use for our purpose (Section 2.2). At this point, we will be able to raise the research questions addressed in this paper (Section 2.3).

2.1 Learning paths

At course level, a learning path is defined as "a sequence of learning tasks or activities which are designated to assist the student in improving their knowledge or skill in the particular subject" (Yang et al., 2010). Since then, a large number of works has analysed them and their personalization (Nabizadeh et al., 2020). Its implementation requires the organization of learning activities and a clear definition of how to evaluate student learning and should be designed to match the learning outcomes requirements.

According to its definition, two tasks should be accomplished. First, establishing a clear sequence of activities. Second, determining that the path effectively helps to achieve knowledge. These tasks demonstrate complex (Yang et al., 2010). The first aspect would be linked to the design of the path itself. Different woks show different approaches for the design, and its potential personalization (to cite some works (Davies et al., 2021; Karampiperis & Sampson, 2005; Limongelli et al., 2009; Liu & Yang, 2005)). However, as the compilation referred (Nabizadeh et al., 2020) clearly indicates “evaluation is always one of the main challenging phases in the learning path personalization methods”. While this is particularly true when considering path personalization, its evaluation is also needed in environments where a single path is suggested for the students taking the course.

The evaluation of learning paths involves the need to determine to what extent students are effectively following the path. Otherwise, it would be impossible to determine whether the offered learning path is assisting students to improve their knowledge. Different works have analysed how this “sequence of tasks” is effectively materialized (Ramos et al., 2021). Initial works on path visualization performed a Gannt chart as a simple tool to represent the path (Teutsch & Bourdet, 2010). Adesina and Molloy (Adesina & Molloy, 2011) suggest the use of tools linked to business process, and in particular, the use of BPMN models. Recent works include the use of and graphs (Cerezo et al., 2014; Ramos et al., 2021) and sequences (Saito & Watanobe, 2020).

While these works can help to visualize a theoretical learning path, none of them provide formal methods oriented to compute to what extent a student is following the path. References to mathematical formulations addressing the evaluation of the learning paths are difficult to find. (Yang et al., 2010) suggest a learning path model for its construction. However, this formulation does not include the evaluation stage. The evaluation is performed through a specific survey. This survey showed that 90% of the participants find useful to link learning tasks and learning outcomes. The same survey indicated that the model itself provides a clear understanding of how the students progress through the course.

Working with personalized paths, (Govindarajan et al., 2017) propose an evaluation based on the quantification of precision and sensibility. These parameters are linked to the predicted student learning style used for classification, rather than on the path itself. (Klašnja-Milićević et al., 2011) compare results of a control group with regard to a group where the personalized learning path was introduced. (Essalmi et al., 2010) indicate the relevance to analyse student satisfaction as a measure to evaluate the learning path.

As it can be seen, most works are focused on designing and providing personalized learning paths rather than on providing formal methods to evaluate those paths. Questions such as: *To what extent is the student following the path?* or *Which is the impact to diverge from the suggested track?* are not formally addressed. These questions emerge as particularly relevant in environments where a path is suggested, but where the student has freedom to select learning objects.

2.2 Modelling paths: the role of *process mining*

Modelling student activity is a recurring task in *learning analytics*. As different compilations show, quantitative parameters linked to the amount of activity have been commonly considered (Conijn et al., 2017). Early works dealing with this topic (Macfadyen & Dawson, 2010) consider parameters directly derived from LMS data – such as total number of discussion messages posted, or total number of mail messages sent-. Recent works aim to detect patterns indicating specific behaviors in log data (Araka et al., 2020), as they can constitute better indicators of potential problems. In addition, as a recent work (Lerche & Kiel, 2018) remarks, quantitative measures of activities tightened to LMS parameters, often lack meaningful information linked to didactics. Values such as the number of clicks, or the number of days a student logins, may reveal information on how much or how often the student performs activities, but not on what she is really doing. It also lacks the temporal perspective. Nothing can also be inferred regarding potential previous contents not covered, or a lack of previous activities.

The focus on meaningful information, the detection of patterns, and the consideration of the potential match to a suggested learning path makes us consider the suitability of *process mining* to analyse student behaviour, and in particular, learning paths. It addresses “the analysis of processes using event data” (W. van der Aalst, 2012). While its initial emphasis was on business-oriented scenarios, its current range of applications is extensive (Rojas et al., 2016; W. van der Aalst, 2012).

In the educational field, it has given birth to educational *process mining* (EPM) (Bogarín et al., 2018c; Ghazal et al., 2018) with a broad range of applications including the analysis of assessments (Cerezo et al., 2020; Pechenizkiy et al., 2009), performance prediction (Umer et al., 2017), detection of patterns and strategies – in particular in SRL environments - (Bannert et al., 2014; de Leoni et al., 2016) or overall improvement of the learning process (Bogarín et al., 2014). It has also used in MOOCs (Mukala et al., 2015a; Zhang et al., 2020), in particular to detect behavioral patterns.

Process mining includes three relevant stages: process discovery, conformance checking, and process enhancement (W. M. P. van der Aalst, 2011). Process discovery deals with the identification of hidden processes inside log data. Conformance refers to the ability to determine the matching between a process model and a real log trace. Based on the results of discovery and conformance, the potential enhancement of processes emerges.

Regarding process discovery, there is neither a single technique to infer models nor a single resulting model for a given log. Theoretical models represent what should occur, whereas found models depict what actually occurs, giving place to a potential computationally difficult problem. Specific algorithms attempt to reduce computing complexity by delivering models with fewer features than those of more complicated models (Leemans et al., 2018).

Conformance can help to determine either to what extent a model includes the information in a given log, or to establish whether a log effectively fits to the model. The common approach to determine conformance is by means of alignments (W. M. P. van der Aalst, 2011), which aim to determine where the log diverges from the model and vice versa. The analysis of those discrepancies between the log and the model provides the foundation for process improvement.

While alignments constitute the classical approach to conformance, recent studies perform an alternative approach. Instead of using the full model, they compute its *skeleton*. This *skeleton* includes the relevant characteristics of the discovered process. *Skeletons* have shown particularly useful to determine whether a given behaviour matches a given process (Verbeek & de Carvalho, 2018).

However, none of the works referred or – to the best of our knowledge – any other has explored the potentialities of *process mining*, and in particular, that of *skeletons*, to analyze to what extent the student activity is consistent with a given learning path, quantifying the similarity. The closest application found explores curricular trajectories at the program level, but does not quantify to what extent it corresponds to an expected trajectory neither considers the use of *skeletons* (Salazar-Fernandez et al., 2021b). The same study suggests that *process mining* can constitute a novel method to unveil student trajectories, which we believe that can be applied at course level. The next subsection formally defines the research questions that we address.

2.3 Research questions

Linking these ideas together, we explore the potentialities of *process mining*, and in particular the use of *skeletons*, to model and evaluate learning paths. This translates formally into the first of the research questions (RQs) covered in this paper, which addresses the theoretical foundations behind the process:

RQ1: How can we measure to what extent students' activity matches the expected behaviour suggested by the teachers for a given course?

Further than this, we cover the practical implementation, evaluating whether following the path is relevant in terms of course grades in a higher education *online* mathematics course. We address the following question:

RQ2: For the course under consideration, and considering path execution and course grades, to what extent is following the path linked to course grades?

This second question will serve as a validation for the method exposed, that should be portable to any course where the evaluation of a learning path is under consideration.

3 Methodology

The methodology used to address the research questions in Section 2.3. is covered in this section. First, Section 3.1. formally exposes the method we propose to analyse learning paths through *process mining*. This method consists of four stages. Before introducing them in detail, we provide details on the course where practical implementation will be carried out in Section 3.2. Once done, Sections 3.3. to 3.6 will cover each of the method's four stages.

3.1 Modelling learning paths through a combination of theoretical constraints and real student activity

Theoretical models of a learning path can be constructed using standard notations such as BPMN, DFGs, or Petri Nets (Dijkman et al., 2008; W. M. P. van der Aalst, 2014). However, real processes tend to differ from theoretical models, and in particular, when the process is complex. This makes the modelling of learning processes particularly challenging, as no single student strictly follows the same path. Despite a main path can be suggested, real behavioural patterns include skipping certain activities, reviewing some others, or performing some out of time, deriving into a high number of potentially different implementations.

Our approach to materialize a model of a learning path is to **generate a process model based on real activity extracted from those logs associated to the students that follow the learning path suggested by the teachers**. To accomplish this task, four stages are needed:

1. The activity log is pre-processed and translated into a meaningful log from a learning process perspective.
2. This processed log is filtered to extract only the trajectories associated to students that follow the indications of the learning path suggested by the teacher. The resulting filtered log contains real activity that conforms to the learning path.
3. The *skeleton* of the process associated to the filtered log is computed.
4. A conformance process is carried out to quantify the differences between a given student trace and the *skeleton* computed above.

This process is depicted in Figure 1.

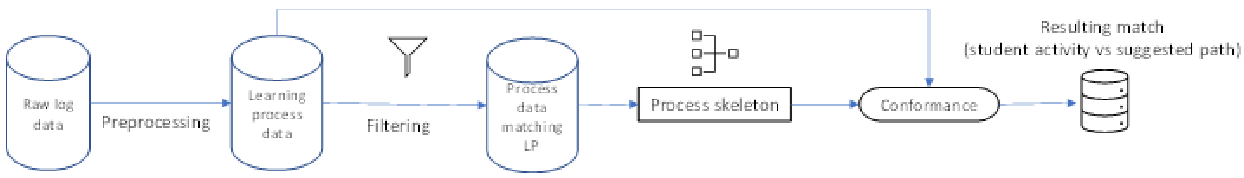


Fig. 1. Conceptual framework for model generation

Figure 1 shows the process we carry out. First, raw data in LMS logs is translated into activities linked to a process, meaningful from a learning process perspective (Stage 1). This conversion is dependent on the course itself and includes a detailed view of the activities that conform the course. This processed log will be filtered, based on the constraints determined by the theoretical learning path (Stage 2). This learning path is expressed in terms of sequences of activities, paths not to be covered, or any indications based on pedagogical considerations.

The filtering process will provide a subset of the original log containing traces associated to students whose activity matches that suggested in the learning path. This information will be considered as real trajectories associated to the theoretical learning path. Using *process mining* techniques, the associated *skeleton* will be computed (Stage 3). This *skeleton* will allow to compare the activity performed by any given student (which is present in the processed log in Stage 1) to the learning path suggested by the instruction, whose core parameters are contained into the *skeleton* (Stage 4).

We provide a detailed explanation of these four stages in the following subsections, considering also its practical implementation in a real course. To do so, we begin by explaining the relevant aspects linked to this course.

3.2 Course description

The course under analysis is an introductory mathematics course offered in a European *online* university. It is offered to all students entering engineering degrees and lasts for one semester. The enrolled students perform their *online* activity asynchronously through a Moodle platform. Successive editions of the course have focused on increasing student retention. Student surveys reflect a great satisfaction with the structure and development of the course. The course design includes a set of modules (11 in total) containing training and evaluative activities. An initial quiz – without any impact on grades – is used both to explain how the course works and to validate previous knowledge. The suggested learning path is clearly exposed.

Each modules includes training and evaluative activities. Training activities have shown nuclear to pass this specific course (Figueroa-Cañas & Sancho-Vinuesa, 2021b). These activities do not impact on grades, but students are encouraged to perform them before taking evaluative assessments. The modules should be taken in sequential order. To accomplish these goals, the modules are not open till the suggested initial date. In addition, non-evaluative tests are opened before evaluative assessments to encourage practice. Regarding evaluative activities, they have a deadline, which is always set before the start date of the next unit. No submissions are allowed after the deadline. Once opened, both kind of activities remain open for the rest of the course and can be used by students to review or perform extra practice. Two tests recapitulating the concepts are also performed by the students, with impact on grades. One after 5^{th} module and one at the end of the course. All this activity is performed in a Moodle environment, where data regarding the date, the user, the specific activity and a detailed description of the action is logged as low-level log.

Getting into deeper detail, for each module, a student can perform practice or evaluation activities. When practice is performed at its due date, we consider it real practice. However, the student can decide to skip it, and perform it later, generating a late activity. She can also perform it on time, and review it again at a later time. In this case we talk about a review process. These concepts are shown in Table 1 .

Table 1. Activity description

Activity mnemonic	Meaning
Mn-PR	Practice (non-evaluative) activity for module n
Mn-EV	Evaluative activity for module n
Mn-REVIEW	Review of contents associated to Module n once the module is completed. The contents have been however worked on the due date
Mn-LATE	Work on contents associated to Module n after module is over. These LATE contents have not been covered when the student was instructed to cover them

The learning path suggested for a specific module consists in performing practice before approaching evaluation. As indicated, practice cannot be carried out before the suggested date as the module contents are not opened in advance. Potential review of previous modules is optional. Evaluative assessments can not be performed late, as there is a due date for evaluative assessment. Figure 2 shows a graphic visualization of the theoretical suggested learning path for the i^{th} module following BPMN notation. M_j-REV indicates potential review of previous modules (i.e. $j<i$). Students can practice and review as much as they consider before evaluation but should at least practice once.

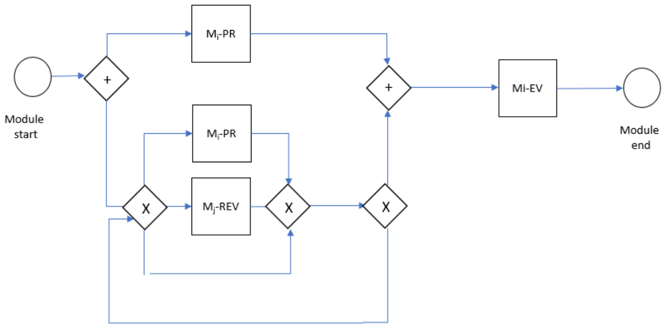


Fig. 2. Expected student behaviour within a course unit.

The course learning path consists of a sequence of 11 modules following the model in Figure 2, and the completion of an intermediate test after module 5 and a final test after Module 11. Neither failing nor skipping non-evaluative assessments impact grades directly.

Regarding the number of students enrolled, we analyse data linked to 6 different course groups totalizing 330 students, 280 of which pass the subject.

Once the main characteristics linked to our case study are known, we get back to the details of the process, as indicated in Section 3.1.

3.3 From raw logs to process-oriented logs (Stage 1)

As indicated in the previous subsection, data for our case study is gathered from Moodle logs. It includes the required fields for *process mining*: a user identification, a description of the low-level action performed and a timestamp. However, the log is not suitable from a *process mining* perspective. A filtering process is needed to accommodate the log to a meaningful process (Bogarín et al., 2018b). In our particular case-study, the information needs to be consistent with the activities reflected in Table 1 and depicted in Figure 2.

To better understand this concept let us consider a student that performs a non-evaluative quiz associated to first module. According to the learning path, it is not the same to perform this activity when this module is running (the student is practising) or doing it later – in this case, the student is either late or reviewing depending on whether she has covered it previously -. While these situations will be equally logged in the LMS, they correspond to different activities from a learning process perspective.

Once this preprocessing has been carried out, we get meaningful learning processes. As an example, Figure 3 shows the initial stages associated to a student learning process:

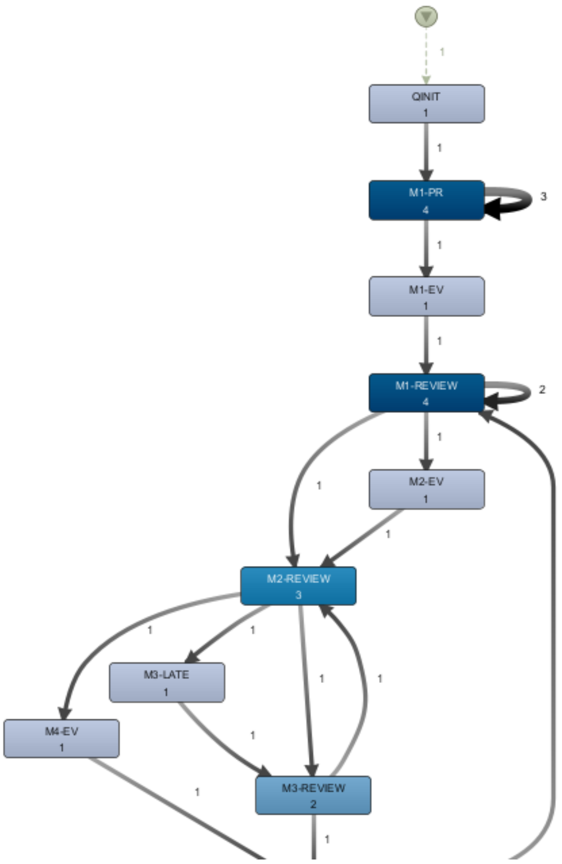


Fig. 3. Initial stages of the path followed by a sample student (DFG format)

In the figure above, the student begins with the introductory quiz for the course, before attempting non-evaluative quizzes. After different work on these quizzes, she decides to take the evaluative assessment. After that, he reviews the contents covered in module 1, before directly attempting the evaluative assessment for the second module. This content is reviewed (M2-REVIEW), before covering module 3 contents. These contents are covered late, and reviewed after. The graph also shows the evaluative activity associated to M4. No reference is found to covering M3 contents. As we can see, the track that the student is following can be analysed, and in this particular case, would not be following the path suggested by the teachers, as she has skipped contents associated to the third module.

3.4 Filtering processed log according to learning constraints (Stage 2)

As the previous subsection shows, the processed log can be interpreted to determine whether a student adheres to the learning path constraints or not. In our case study, the suggested learning path includes the following constraints:

- Practice should be covered for all modules before performing the evaluative assessments associated to that module.
- Evaluative assessments must be performed for all modules.
- Students should also perform the mid-term and final assessments.

These constraints can be translated into rules to filter the processed log obtained in Section 1. The results of this filtering process correspond to activity of students who follow the path. However, we want to remark that this does not correspond to a single path. As an example, there are no restrictions associated to review of contents. Some students can decide to review contents of Module 2 (for instance) while some others not. Some will do it when reviewing the third module, and some others later. Finally, we are considering real activity which matches theoretical constraints of a learning path. This implies that it includes the theoretical constraints but at the same time includes the real activity that – consistently with the complexity of a learning process – is dependent on the individual student's needs.

3.5 Generation of the *skeleton* associated to the learning path (Stage 3)

Once the data is filtered, modelling activity as a process requires three main concepts. The case identifier (*case_id*), which refers to the each running instance of a process. The activity taking place, which refers to each step carried out during the process and finally, the timestamp of each activity, which refers to the specific moment when the activity is carried out.

In our case, and consistently with studies modelling courses as a process (Mukala et al., 2015a), an anonymised student id will be used as *case_id*. For each student, we will generate a process containing the key activities she performs. As timestamp we use the moment when the student carries out the activity itself. Each student will generate a trajectory. These trajectories will be filtered, and only those that adhere to the learning path constraints (Stage 2) will be kept.

Due to the potentially high number of tasks and students, the associated process can be highly complex. In addition, classical conformance techniques – in particular, alignments (W. van der Aalst, Adriansyah, & van Dongen, 2012)– do not provide a fitter matching index providing straight interpretation. For both reasons, instead of computing the whole process, we will keep the *skeleton* associated to the process that derives from the filtered trajectories that adhere to the learning path. *Skeletons* aim to simplify the view of a process, analysing a log, and summarizing common activities, relations, and order between them. Table 2 indicates the concepts normally included in process *skeletons*:

Table 2. Contents of a process *skeleton*

Concept	Description
Equivalence	Activities that show the same frecuency in all running instances of a process
Always after (A,B)	Detection of activities (A) that are always followed by some others in the future (B)
Always before (B,A)	Detection of activities (B) that were always preceeded by the execution of others (A)
Never together	Activities which are never found together in a single trace
Directly follows (A,B)	Detection of activities (A) which are always followed by another activity (B) in all traces considered
Activity frecuency	Maximum number of times each activity is performed in the traces analysed

As it can be seen, *skeletons* can summarize and simplify the analysis of complex processes. While they are not commonly used, recent studies indicate its potentiality when considering classification scenarios (Verbeek & de Carvalho, 2018).

3.6 Computing deviations from the modelled learning path (Stage 4)

Inside *process mining*, conformance techniques are used to determine to what extent a trace matches a process(W. van der Aalst, Adriansyah, & van Dongen, 2012). However, interpretation of conformance values depends on the process under analysis. The range associated to the matching index depends on the process itself and on the length of the trace. As an additional limitation, and derived from this fact, conformance techniques do not perform particularly well for classification purposes. While our goal is not strictly to classify traces, we finally want to know if a trace matches a model or not – quantifying in addition to what extent -.

Besides simplifying the process view, *skeletons* provide conformance algorithms (Kudo et al., 2013; Przybylek, 2011). They compute the extent to which a trace matches the *skeleton*, providing a matching value ranged [0,1]. This allows for easy interpretation, which is nuclear for our analysis. Also, and as indicated in the previous subsection, they have proved powerful as classifiers. Our approach is to use the *skeleton* of the process rather than the process itself to compute the matching index.

Keeping in mind that we have generated the *skeleton* based on traces including real behaviour that strictly adhere to the learning path constraints, the result of matching index will indicate to what extent a trace conforms to the main characteristics of the modelled learning path. For the purposes of our research, this information is particularly relevant, as it provides the following information:

- From a teaching perspective, information on whether students are effectively following the path, providing a clear quantitative view both at group and student level.
- In terms of learning design, a method to validate learning paths, determining the relevance of following a suggested path.

Before presenting the results for our case study, we want to remark that the process described in this section can be applied to any course with logged activity – as far as this log includes activity, student identification and moment in time –.

4 Results

This section provides the results of the process explained in the previous subsection, focusing on providing answers to the research questions raised in Section 2.3.

4.1 Modelling the suggested learning path: real path is not a linear process

The process described in the previous section has been performed on our data log. To do so, the activities have been defined as indicated in Table 1. Learning path constraints indicated in Section 3.4 have been considered. Following the nomenclature in Table 1, we have filtered those students where M_i-PR is always performed before M_i-EV for all activities along the course (i.e. for i values from 1 to 11, corresponding to all the modules along the course). The mid-term assessment must also be performed.

Filtering has been carried out using the filtering functionalities of a *Process mining* library for Python(*Pm4py · PyPI*, n.d.) . Results indicate that only 8 of the 330 students strictly follow all the required steps. However, and despite this relative low number, the sequence of activities is not a straight path, but presents different trajectories. We visualize the associated DFG in Figure 4.

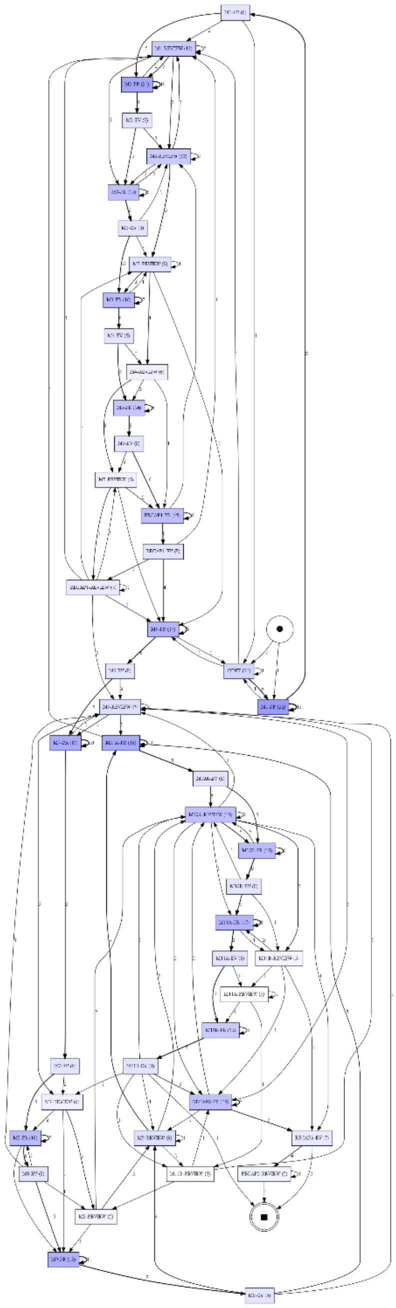


Fig. 4. DFG of students performing all activities suggested timely and in proper order

As it can be seen, the real implementation of the process is not a simple linear flow. While the process is meant to follow all the steps suggested by the theoretical learning path, the students take other activities – i.e. they review content at different stages of the course before approaching certain others -. As a last stage, we compute the *skeleton* for this process that will allow to compare a student's trajectory against this model.

4.2 Matching user activity and process *skeleton* provides a view of whether students follow the path, but also of how relevant is following it

Once we have computed the *skeleton*, we can compare log traces for the rest of the students with it. Index match is a [0,1] range, with 0 indicating no match at all and 1 indicating perfect match. Figure 4 plots the histogram of the values obtained:

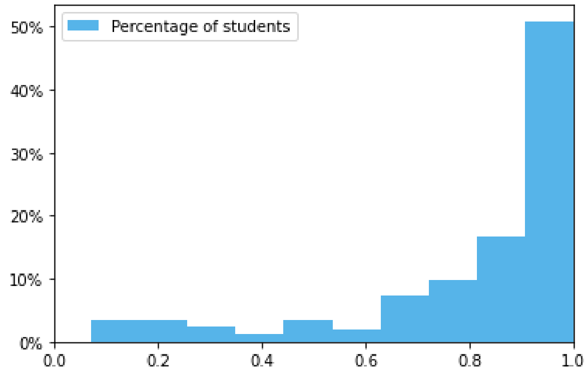


Fig. 5. Trace fitness histogram (%) of students not conforming the model

Despite only 8 students strictly follow the expected path, Figure 5 reflects that most students follow paths that resemble considerably to the *skeleton* of the model generated. In other words, the students are following the indications of their instructions. However, the analysis in Figure 5 does not reflect whether following the path is convenient to achieve better academic results. To verify this specific point, we compare the matching index with the *skeleton* against the marks obtained by students in the course. Results are shown in Figure 6.

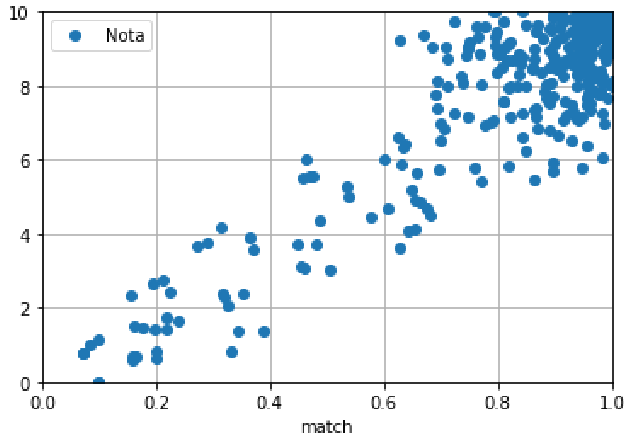


Fig. 6. Match index vs Final grade plot

Figure 6 clearly indicates that those students with higher matching to the model also obtain higher marks. Correlation between the matching index and final grade results in R^2=0.7996. In other words, and for our course under analysis, a great deal of the variance in final mark is contained in the index indicating whether the students follow the path. In addition, following the path that the instructors suggest, increases the probability to pass the course and obtain higher scores.

5 Discussion

Two research questions were addressed in this paper. Considering the method exposed in Section 3. and the specific results in Section 4. we are in a position to answer them. We summarize findings before getting into deeper detail.

Regarding the first question, the use of *process mining* techniques, and in particular, the consideration of the *skeleton* of the process, are key components to measure in a quantitative and interpretable way to what extent a student is following the path. The steps indicated in Section 3.1. have led to a path model. As a noticeable fact, this model allows for a matching fitted index, which provides clear interpretation.

The second question addressed the translation of this method into practice. In our particular case, it has been demonstrated that following the path is relevant in order to pass the course, and specifically a correlation R^2=0.7996 has been determined. This constitutes empirical evidence of the relevance of a learning path. It must be pointed out that we have found no previous references in the literature approaching this problem in a methodological way that provides this kind of indicator. While the results are for our specific case study, the method can be applied to any other course under consideration. This specific point can help to improve the quality of teaching, by designing learning paths that prove effective in terms of passing the subject.

5.1 Comparison with the literature

The transversality of the process carried out allows for a comparison both in terms of the process followed and of its specific result. First of all, and from a generic perspective, our work confirms the relevance of activity, and in particular that carried out *online*, in student achievement. The correlation found between the process-mapped activity and the final grade is consistent with the research works so far (to cite a couple of compilations (Agudo-Peregrina et al., 2014; Conijn et al., 2017)).

The process carried out puts the learning path – and in fact, the learning process – as the center of the analysis. This is completely aligned with works suggesting meaningful analysis rather than prediction ratios (Lerche & Kiel, 2018; Romero et al., 2008). While our goal was not to establish predictions, it is noticeable that for our particular course, the early activity has not shown so different between passing and failing students in terms of process execution. This is reflected in Figure 7(a) and would indicate that other parameters are needed if the goal is to establish predictions.

Regarding the specific analysis of learning paths, and its potentialities to deepen into the analysis of the global learning process, we consider that the method described extends visualization techniques – as those compiled in (Ramos et al., 2021) - , allowing to quantify the difference. In addition, and from a *process mining* perspective, the use of *skeleton* has the ability to provide an indicator that allows for clear interpretation – for being clearly defined in a [0,1] interval – in comparison to other conformance checking techniques based on alignments.

Regarding the comparison of the path followed by the students – either considered as a group, or at individual level – and that suggested, our work adds meaningful measures to existing works. Considering those suggesting surveys (Yang et al., 2010), these can be helpful, they are also difficult to implement and tightened to a specific course. A similar problem derives when there is a need to implement control and reference groups (Klašnja-Milićević et al., 2011). While our approach avoids this need focusing on determining the relevance of the path in terms of its relation to the learning outcome, it could also accommodate potential comparison between different paths.

As an additional consideration, the approach we propose can be considered regardless of the course. Regarding the works considering parameters such as precission or sensibility(Govindarajan et al., 2017), we believe our approach adds a process perspective to the classification approach linked to these measures.

5.2 Implications

The method carried out translates the activity performed by the student into a process that allows for analysis, determining how close is the activity performed to the one suggested by the instructor.

The method exposed, and the quantitative approach, can help in different problems and questions linked to *learning analytics*. First, it allows for path validation. It has proven useful to effectively show that following the path leads to a better learning outcome. This is particularly relevant when designing learning paths. We have focused on the evaluation of one path – as the course under analysis had no personalized paths -. However, the extension of this method to compare different paths would be straightforward.

The process is also useful from an intervention perspective. The process exposed can be carried out along time, determining when the student is diverging from the path. The goal of our study was not specifically determining a given moment in time, or a specific divergence threshold suitable for intervention. We believe this approach is based on meaningful information, adding pedagogical information which is not evident in other methods.

From a data-mining perspective, we encourage the use of *skeletons* to match the analysis of modelled and executed processes. Our results would confirm the potentialities of this technique suggested by (Verbeek & de Carvalho, 2018). In that study, it is used in terms of classification, but we believe it also allows for other relevant aspects, such as the quantification of divergences – as shown in our study – and the analysis of those divergences – while it is out of the scope of this research, we encourage a future research line on this topic -. We extend these potential research lines opened by this fact in the next section.

6 Conclusion

Process mining has proven as a relevant method to analyse student activity at course level, and in particular, to evaluate the relevance and conveniency of following a given learning path. The use of *skeletons* has helped to synthesise the differences between the suggested and the executed path in an indicator concentrating the match.

We encourage the use of this technique to deepen into the pedagogical implications that the method opens. First, and as noticed in the previous subsection, the analysis can be extended to determine the specific differences between the suggested path and real executions. This would determine what is the difference rather than how relevant it is in quantitative terms.

Further than this, we would recommend its use in environments that apply personalized learning paths. While our course is based on a traditional approach, with a clear suggested learning path, the extension to compare different paths could provide relevant insight. This comparison could be performed either in systems that use personalized learning paths, or those implemented through A/B testing strategies. Authors are open to collaborate in works related to these research lines.

7 References

Adesina, A., & Molloy, D. (2011). A business process management based virtual learning environment: Customised learning paths. *CSEDU 2011 - Proceedings of the 3rd International Conference on Computer Supported Education*, *1*. https://doi.org/10.5220/0003273203650368

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for *learning analytics* and their relation with performance in VLE-supported F2F and *online* learning. *Computers in Human Behavior*, *31*(1), 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—systematic review (2008–2018). *Research and Practice in Technology Enhanced Learning*, *15*(1), 1–21. https://doi.org/10.1186/S41039-020-00129-5/TABLES/3

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). *Process mining* techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*(2), 161–185. https://doi.org/10.1007/s11409-013-9107-6

Bogarín, A., Cerezo, R., & Romero, C. (2018a). A survey on educational *process mining*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1). https://doi.org/10.1002/widm.1230

Bogarín, A., Cerezo, R., & Romero, C. (2018b). Discovering learning processes using inductive miner: A case study with learning management systems (LMSs) | Descubriendo procesos de aprendizaje aplicando inductive miner: Un estudio de caso en learning management systems (LMSs). *Psicothema*, *30*(3), 322–329. https://doi.org/10.7334/psicothema2018.116

Bogarín, A., Cerezo, R., & Romero, C. (2018c). A survey on educational *process mining*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1), e1230. https://doi.org/10.1002/widm.1230

Bogarín, A., Romero, C., Cerezo, R., & Sánchez-Santillán, M. (2014). *Clustering for improving Educational Process mining*. https://doi.org/10.1145/2567574.2567604

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). *Process mining* for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education*, *32*(1), 74–88. https://doi.org/10.1007/s12528-019-09225-y

Cerezo, R., Suarez, N., Carlos Núñez, J., & Sánchez-Santillán, M. (2014). EGraph tool: Graphing the learning process in LMSs. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/2567574.2567596

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2016.2616312

Davies, J. N., Verovko, M., Verovko, O., & Solomakha, I. (2021). Personalization of e-learning process using ai-powered chatbot integration. *Advances in Intelligent Systems and Computing*, *1265 AISC*. https://doi.org/10.1007/978-3-030-58124-4_20

de Leoni, M., van der Aalst, W. M. P., & Dees, M. (2016). A general *process mining* framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems*, *56*, 235–257. https://doi.org/10.1016/j.is.2015.07.003

Dijkman, R. M., Dumas, M., & Ouyang, C. (2008). Semantics and analysis of business process models in BPMN. *Information and Software Technology*, *50*(12). https://doi.org/10.1016/j.infsof.2008.02.006

Essalmi, F., Ayed, L. J. ben, Jemni, M., Kinshuk, & Graf, S. (2010). A fully personalization strategy of E-learning scenarios. *Computers in Human Behavior*, *26*(4). https://doi.org/10.1016/j.chb.2009.12.010

Figueroa-Cañas, J., & Sancho-Vinuesa, T. (2021a). Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous *online* calculus module. *Interactive Learning Environments*, *29*(1). https://doi.org/10.1080/10494820.2018.1559864

Figueroa-Cañas, J., & Sancho-Vinuesa, T. (2021b). Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous *online* calculus module. *Interactive Learning Environments*, *29*(1). https://doi.org/10.1080/10494820.2018.1559864

Ghazal, M. A., Ibrahim, O., & Salama, M. A. (2018). Educational *process mining*: A systematic literature review. *Proceedings - 2017 European Conference on Electrical Engineering and Computer Science, EECS 2017*, 198–203. https://doi.org/10.1109/EECS.2017.45

Govindarajan, K., Kumar, V. S., & Kinshuk. (2017). Dynamic Learning Path Prediction - A Learning Analytics Solution. *Proceedings - IEEE 8th International Conference on Technology for Education, T4E 2016*. https://doi.org/10.1109/T4E.2016.047

Karampiperis, P., & Sampson, D. (2005). Adaptive learning resources sequencing in educational hypermedia systems. *Educational Technology and Society*, *8*(4).

Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers and Education*, *56*(3). https://doi.org/10.1016/j.compedu.2010.11.001

Kudo, M., Ishida, A., & Sato, N. (2013). Business process discovery by using process skeletonization. *Proceedings - 2013 International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2013*. https://doi.org/10.1109/SITIS.2013.158

Leemans, S. J. J., Fahland, D., & van der Aalst, W. M. P. (2018). Scalable process discovery and conformance checking. *Software and Systems Modeling*, *17*(2). https://doi.org/10.1007/s10270-016-0545-x

Lerche, T., & Kiel, E. (2018). Predicting student achievement in learning management systems by log data analysis. *Computers in Human Behavior*, *89*, 367–372. https://doi.org/10.1016/J.CHB.2018.06.015

Limongelli, C., Sciarrone, F., Temperini, M., & Vaste, G. (2009). Adaptive learning with the LS-Plan system: A field evaluation. *IEEE Transactions on Learning Technologies*, *2*(3). https://doi.org/10.1109/TLT.2009.25

Liu, H. I., & Yang, M. N. (2005). QoL guaranteed adaptation and personalization in e-learning systems. *IEEE Transactions on Education*, *48*(4). https://doi.org/10.1109/TE.2005.858398

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and Education*. https://doi.org/10.1016/j.compedu.2009.09.008

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open *Online* Courses. *Computers in Human Behavior*, *80*, 179–196. https://doi.org/10.1016/j.chb.2017.11.011

Mowrer, O. H. (1960). Learning theory and behavior. In *Learning theory and behavior.* John Wiley & Sons Inc. https://doi.org/10.1037/10802-000

Mukala, P., Buijs, J., Leemans, M., & van der Aalst, W. (2015a). Exploring Students' Learning Behaviour in MOOCs using *Process mining* Techniques. *Computing Conference*.

Mukala, P., Buijs, J., Leemans, M., & van der Aalst, W. (2015b). Learning analytics on coursera event data: A proceb mining approach. *CEUR Workshop Proceedings*, *1527*.

Nabizadeh, A. H., Leal, J. P., Rafsanjani, H. N., & Shah, R. R. (2020). Learning path personalization and recommendation methods: A survey of the state-of-the-art. In *Expert Systems with Applications* (Vol. 159). https://doi.org/10.1016/j.eswa.2020.113596

Pechenizkiy, M., Trčka, N., Vasilyeva, E., van der Aalst, W., & de Bra, P. (2009). *Process mining online* assessment data. *EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining*.

pm4py · PyPI. (n.d.). Retrieved September 5, 2022, from https://pypi.org/project/pm4py/

Przybylek, M. R. (2011). Skeletal algorithms. *ECTA 2011 FCTA 2011 - Proceedings of the International Conference on Evolutionary Computation Theory and Applications and International Conference on Fuzzy Computation Theory and Applications*. https://doi.org/10.5220/0003674700800089

Ramos, D. B., Ramos, I. M. M., Gasparini, I., & de Oliveira, E. H. T. (2021). A new learning path model for e-learning systems. *International Journal of Distance Education Technologies*, *19*(2). https://doi.org/10.4018/IJDET.20210401.oa2

Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). *Process mining* in healthcare: A literature review. In *Journal of Biomedical Informatics* (Vol. 61). https://doi.org/10.1016/j.jbi.2016.04.007

Romero, C., Ventura, S., Espejo, P. G., & Hervás, C. (2008). Data mining algorithms to classify students. *Educational Data Mining 2008 - 1st International Conference on Educational Data Mining, Proceedings*.

Saito, T., & Watanobe, Y. (2020). Learning path recommendation system for programming education based on neural networks. *International Journal of Distance Education Technologies*, *18*(1). https://doi.org/10.4018/IJDET.2020010103

Salazar-Fernandez, J. P., Sepúlveda, M., Munoz-Gama, J., & Nussbaum, M. (2021a). Curricular analytics to characterize educational trajectories in high-failure rate courses that lead to late dropout. *Applied Sciences (Switzerland)*, *11*(4), 1–15. https://doi.org/10.3390/app11041436

Salazar-Fernandez, J. P., Sepúlveda, M., Munoz-Gama, J., & Nussbaum, M. (2021b). Curricular analytics to characterize educational trajectories in high-failure rate courses that lead to late dropout. *Applied Sciences (Switzerland)*, *11*(4). https://doi.org/10.3390/app11041436

Teutsch, P., & Bourdet, J. F. (2010). How to see training paths in learning management systems? *Proceedings - 10th IEEE International Conference on Advanced Learning Technologies, ICALT 2010*. https://doi.org/10.1109/ICALT.2010.98

Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2017). On predicting academic performance with *process mining* in *learning analytics*. *Journal of Research in Innovative Teaching & Learning*, *10*(2). https://doi.org/10.1108/jrit-09-2017-0022

van der Aalst, W. (2012). *Process mining*: Overview and opportunities. In *ACM Transactions on Management Information Systems* (Vol. 3, Issue 2, pp. 1–17). https://doi.org/10.1145/2229156.2229157

van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., Bose, J. C., van den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., de Leoni, M., … Wynn, M. (2012). *Process mining* manifesto. *Lecture Notes in Business Information Processing, 99 LNBIP*(PART 1). https://doi.org/10.1007/978-3-642-28108-2_19

van der Aalst, W., Adriansyah, A., & van Dongen, B. (2012). Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(2). https://doi.org/10.1002/widm.1045

van der Aalst, W. M. P. (2011). *Process mining*: Discovery, Conformance and Enhancement of Business Processes. In *Media* (Vol. 136, Issue 2).

van der Aalst, W. M. P. (2014). *Process mining* in the large: A tutorial. *Lecture Notes in Business Information Processing*, *172 LNBIP*. https://doi.org/10.1007/978-3-319-05461-2_2

van der Aalst, W. M. P., Reijers, H. A., Weijters, A. J. M. M., van Dongen, B. F., Alves de Medeiros, A. K., Song, M., & Verbeek, H. M. W. (2007). Business *process mining*: An industrial application. *Information Systems*, *32*(5). https://doi.org/10.1016/j.is.2006.05.003

van der Aalst, W. M. P., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A. J. M. M. (2003). Workflow mining: A survey of issues and approaches. *Data and Knowledge Engineering*, *47*(2). https://doi.org/10.1016/S0169-023X(03)00066-1

van der Aalst, W. M. P., & Weijters, A. J. M. M. (2004). *Process mining*: A research agenda. In *Computers in Industry* (Vol. 53, Issue 3). https://doi.org/10.1016/j.compind.2003.10.001

Verbeek, H. M. W., & de Carvalho, R. M. (2018). *Log Skeletons: A Classification Approach to Process Discovery*. https://doi.org/10.48550/arxiv.1806.08247

Yang, F., Li, F. W. B., & Lau, R. W. H. (2010). An open model for learning path construction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6483 LNCS*. https://doi.org/10.1007/978-3-642-17407-0_33

Zhang, F., Liu, D., & Liu, C. (2020). Mooc video personalized classification based on cluster analysis and *process mining*. *Sustainability (Switzerland)*. https://doi.org/10.3390/su12073066

Capítulo 5 Conclusiones y líneas futuras

"Conducir con orden mis pensamientos, empezando por los objetos más simples y más fáciles de conocer, para ascender poco a poco, gradualmente, hasta el conocimiento de los más complejos, y suponiendo incluso un orden entre ellos que no se parecen naturalmente unos a otros"

René Descartes (1596-1650)

5.1. Contextualización

El capítulo 1 estableció cuál sería el problema abordado por el presente trabajo – la mejora del rendimiento académico –, el ámbito específico – los cursos de estudios de educación superior, muy específicamente los *online* – y la aproximación metodológica – el uso de técnicas que orientadas a la modelización temporal –. Se concretó así un objetivo: contribuir a la mejora del rendimiento académico en los estudios de educación superior, y muy específicamente en aquellos sustentados en entornos *online.*

Este objetivo genérico se materializó en un conjunto de preguntas de investigación que han guiado la realización de los artículos presentados. Hecha esta presentación, estamos en situación de dar respuestas a dichas preguntas. Las subsecciones siguientes abordan de forma secuencial las cuestiones de investigación planteadas en el Capítulo 1. Daremos respuesta a cada una de ellas, incluyendo adicionalmente las implicaciones en relación a potenciales acciones vinculadas a la mejora del rendimiento. Finalmente, y a modo de cierre, una sección final servirá de reflexión última del trabajo y planteará líneas de investigación que basadas en el presente estudio permitan la continuidad y evolución de la investigación llevada a cabo.

5.2. ¿Qué factores preexistentes influyen en el abandono académico? (RQ1)

Tal y como se mostró en el capítulo introductorio, muchos trabajos de investigación aportan evidencias relativas a que no solo los factores estrictamente académicos condicionan el abandono. Ciertos factores preexistentes pueden suponer diferencias significativas. Este hecho se ha abordado explícitamente en uno de nuestros artículos, que profundiza en el abandono desde el punto de vista de las características personales en un contexto de educación *online* (Martínez-Carrascal, Hlosta, et al., 2023).

En dicho artículo se exploran una serie de factores preexistentes. El artículo muestra que la influencia de ciertos factores es transversal, mientras que otros pueden tener impacto específico en algunos cursos (puede consultarse la Tabla 5 del citado artículo a modo de resumen). Entre los **factores transversales** destacan particularmente la **discapacidad, la situación económica y un bajo nivel académico previo**. Entre aquellos factores que pueden tener un impacto **en**

algunos de los cursos, destacamos factores como la **edad o el género.** Estos factores no han demostrado un impacto común o estadísticamente significativo en todos los cursos donde su efecto ha sido analizado. Lo mismo sucede en el caso de la región de procedencia del alumno. En este caso, creemos que realmente la región está enmascarando factores de renta que sí han resultado relevantes.

El hallazgo de factores cuyo impacto depende del curso objeto de análisis es coherente con aquellos estudios que manifiestan que los resultados pueden depender del curso en consideración. Este hecho se refleja de forma clara si se analizan diferentes compilaciones (Hachey et al., 2022; Muljana & Luo, 2019). En ellas se observa disparidad de resultados en lo que respecta al impacto de aspectos como la edad o el género que hemos encontrado en nuestro estudio.

En cuanto a los factores que sí han resultado relevantes en todos los cursos analizados, hay un aspecto que creemos de particular interés: hay una correlación entre un nivel económico preexistente por debajo de un umbral y una situación inicial desventajosa a la hora de abordar los estudios. Recordemos que una de las motivaciones tras nuestro análisis era contribuir a la reflexión de la Comisión Europea (Commission, 2017): la detección temprana es crucial. Fijémonos que un resultado de este tipo contribuye a la revisión de políticas educativas a nivel macro y a su materialización en acciones específicas en las instituciones de educación superior (por ejemplo, la extensión de los periodos de reembolso gratuito de matrícula para colectivos económicamente más vulnerables). Es de subrayar que cuando la renta supera un cierto umbral, dicho incremento no supone una influencia estadísticamente significativa en el abandono.

Queremos hacer notar también que estos resultados, y en particular por lo que respecta a la renta son coherentes con estudios que analizan el impacto en el abandono a nivel de curso ((Cochran et al., 2014) por ejemplo). Este mismo estudio indica también sin cuantificarlo que otro de los factores analizados – en particular, un bajo rendimiento académico previo – condiciona el resultado académico. Este factor es tratado como un indicador predictivo en algunos trabajos (Casanova et al., 2018; Cassidy, 2012; McClelland, 2015; Muljana & Luo, 2019; Nortvig et al., 2018).

Otros factores existentes en la literatura no figuran como relevantes en nuestra investigación. A modo de ejemplo, la región de procedencia del estudiante figura como relevante en entornos *online* en alguna referencia reciente (Rizvi et al., 2019). No hemos detectado tal relación, aun cuando sí es cierto que detrás de la región pueden enmascararse el efecto de mediación de otros factores, como la situación económica.

Es también interesante subrayar la influencia de la discapacidad en el abandono académico, no solo por su impacto a largo plazo, sino porque los estudiantes con discapacidad pueden mostrar patrones con incrementos notables en las tasas de abandono en etapas tempranas respecto a aquellos estudiantes que no presentan este condicionante. La discapacidad es un factor poco explorado en lo que a impacto en el abandono se refiere. Nuestras referencias se han centrado en los trabajos de (Shah & Cheng, 2019) que manifiestan también un potencial impacto (en este caso sobre cursos de acceso abierto).

Respecto al rendimiento previo, otro de los artículos publicados (Martínez-Carrascal, Márquez Cebrián, et al., 2020) muestra el impacto de haber cursado la asignatura con anterioridad. Creemos que este hecho es particularmente relevante, y no siempre considerado a la hora de plantear intervenciones. El estudio revela también la relevancia de la nota de acceso a la universidad como factor determinante predictivo. Este hecho debería ser considerado específicamente en el diseño pedagógico de los cursos universitarios, y en especial, ser analizado para el diseño de cursos propedéuticos.

A nivel comparativo, y en términos generales, los resultados obtenidos son coherentes con los modelos teóricos del abandono, y en concreto con las propuestas de (Bean, 1985; Cabrera et al., 1992; Rovai, 2003; Tinto, 1975) que indican los factores preexistentes como condicionantes del abandono.

Sin embargo, más allá de la coherencia con los modelos, y de los factores detectados, queremos remarcar la relevancia de la metodología planteada, y en particular, de la aproximación dinámica. El uso de un modelo basado en el análisis de supervivencia (Martínez-Carrascal, Hlosta, et al., 2023) plantea una aproximación novedosa en lo que se refiere al análisis de cursos reglados universitarios *online*. Es una aproximación que consideramos muy valiosa desde el punto de vista de su potencial interpretativo. En línea con el enfoque de la tesis, y tal y como se razona explícitamente en dicho artículo, las aproximaciones temporales son esenciales para analizar el problema del abandono cuando el objetivo es el diseño de intervenciones. Este análisis conviene realizarse a nivel de curso, pues como se ha visto, el impacto de los diferentes factores no puede asumirse como transversal e independiente del curso objeto de análisis.

Como se puso de manifiesto en el Capítulo 1, los análisis existentes suelen focalizar en el aspecto predictivo – aumentar ratios de predicción – más que en la comprensión de las causas subyacentes. Creemos que las técnicas longitudinales en su conjunto, y el análisis de supervivencia en particular, pueden aportar visiones de alto interés desde el punto de vista conceptual.

A modo de ejemplo, en (Martínez-Carrascal, Hlosta, et al., 2023) se muestra no solo la influencia del nivel inicial de renta, sino que se sugieren posibles intervenciones. En particular, la extensión del período de desistimiento de matrícula con el abono del pago realizado para los estudiantes con bajo nivel económico. El uso del método permite no solo establecer una relación entre bajo nivel de renta y abandono, sino fundamentar que una intervención adecuada es alargar el período de anulación de matrícula sin penalización económica para determinados estudiantes. Conceptualmente, permitimos que el estudiante siga el proceso de aprendizaje, y evite el abandono por miedo a las consecuencias económicas del fracaso, que se traducen en el abandono de los estudios.

Finalmente, y como sucede en el caso médico – disciplina en la que esta técnica es de uso común –, determinar qué grupo de estudiantes constituye una población de riesgo para un determinado factor permite establecer tratamientos adecuados que ataquen la raíz de dicho factor de riesgo. Esta intervención selectiva permite tratamientos más eficaces, a la vez que evita acciones que no tendrán efectos positivos sobre determinados colectivos. Enlazando con la siguiente pregunta de investigación, la importancia no será solo establecer un factor como relevante, sino su impacto en un grupo específico.

5.3. ¿Cuál es el impacto específico de dichos factores en el abandono? (RQ2)

La forma de determinar el impacto de los factores considerados relevantes se ha analizado explícitamente en (Martínez-Carrascal, Hlosta, et al., 2023). El análisis de las curvas de Kaplan-Meier proporciona no solo la cuantificación del impacto sino la visión temporal de la evolución del abandono académico a lo largo del curso. A modo de ejemplo, y en el citado artículo, la Figura 2 del mismo se centra en factores individuales específicos, mientras que la Figura 3 analiza el efecto combinado.

La primera conclusión es que el método planteado permite la cuantificación del impacto de un factor o la combinación de ellos en el abandono. El análisis de las 22 ediciones de 7 cursos diferentes presentes en el estudio proporcionan la visión de la relevancia estadística de los factores críticos detectados en la cuestión anterior (RQ1). Pero más allá de la relevancia, el método de Cox ha permitido establecer la cuantificación específica del impacto.

De forma sintética, por lo que respecta al **nivel académico previo**, un **bajo nivel preexistente multiplica el riesgo** de abandono por un **factor entre el 1.30 y el 1.42** para estudiantes que provenían de unas calificaciones bajas. Una **renta preexistente baja incrementa el riesgo de abandono entre un 1.18 y un 1.56** según el curso. Finalmente, el efecto de la **discapacidad se cuantifica entre un 1.14 y un 1.49**. Estos resultados resultan de los datos mostrados en las Tablas 6 y 7 del artículo citado.

Según se ha comentado en el apartado anterior, estos resultados son coherentes en términos generales con los estudios preexistentes. Como factor diferencial, nuestro estudio se ha focalizado no solo en la determinación del factor, sino en la cuantificación de su impacto a lo largo del tiempo. El método planteado no solo establece significatividad estadística, sino que **cuantifica el aumento del riesgo, y proporciona la visión evolutiva** de cuándo se puede materializar.

Los cursos analizados presentan una característica que resulta de interés para profundizar en el estudio del nivel de educación previo: son cursos universitarios *online* reglados sin prerrequisitos académicos. Ello significa que puede haber estudiantes sin bagaje académico previo. Si se analiza el impacto de no poseer calificación previa alguna demostrable, el factor de riesgo de abandono se incrementa en hasta un 2.38. Notar también – enlazando con los comentarios finales de la sección previa – que el método propuesto permite la cuantificación del impacto para diferentes grupos objeto de estudio, ya que el impacto de un determinado factor puede estar condicionado por las características de la subpoblación analizada.

Conviene destacar también que en lo que respecta a la educación previa, niveles de educación por encima del umbral mínimo no suponen una reducción significativa del riesgo. De hecho, y como punto relevante, el nivel alto de estudios indicó un aumento de riesgo en uno de los cursos analizados. El análisis de las especificidades de dicho curso indicó que se trataba de un curso propedéutico, cosa que explicaría que estudiantes con buen nivel académico realmente abandonen el curso.

Este hallazgo supone una contrastación empírica del método, y ahonda en la necesidad del conocimiento en profundidad del curso y su contexto para una correcta interpretación del resultado. Se constata la dificultad de profundizar en la interpretación de resultados si no se tiene conocimiento del diseño pedagógico que hay detrás del curso, y refuerza la idea de explorar los problemas en clave interpretativa y con una perspectiva de *learning analytics*. Conviene analizar los factores condicionantes del abandono contextualizándolos como vinculados a cursos específicos, ya que los hallazgos pueden no resultar relevantes con carácter general. La diferencia de impacto puede derivar de aspectos intrínsecos del curso.

De igual forma, pero en relación a otra variable explicativa, en el curso anonimizado como EEE la discapacidad no ha presentado una relevancia significativa, pese a resultar relevante a nivel global. Convendría para dicho curso conocer los detalles internos que pueden explicar esa diferencia. Como autores del trabajo no hemos tenido acceso a información específica sobre dicho curso que hubiera permitido establecer (o al menos, explorar) las causas de esta divergencia. De nuevo se pone de manifiesto la necesidad de interpretación pedagógica de los resultados más allá de la eficiencia del algoritmo subyacente.

El método revela también que un factor puede resultar relevante cuando presenta un valor en determinadas franjas, aun cuando en otros valores no resulte relevante. En particular, este efecto se observa al analizar el nivel de renta. Aquellos estudiantes que declaran un bajo nivel de renta presentan un incremento relevante de riesgo, mientras que aquellos que se sitúan en rentas por encima del 50% del valor medio no presentan una reducción de riesgo. Es importante notar también que más allá del incremento en el riesgo, hay diferencias temporales. En particular, los alumnos de bajo nivel de renta son más propensos al abandono temprano. Según hemos comentado, nuestra hipótesis es que se debe a la relevancia para este colectivo del período de abandono sin efectos económicos. El estudiante que se encuentra en dicha situación tendría más presente la relevancia del coste, y prefiere abandonar para no verse penalizado económicamente en caso de no superar la asignatura.

Finalmente queremos indicar que las bases metodológicas de este proceso fueron exploradas de forma preliminar en (Martínez-Carrascal & Sancho-Vinuesa, 2021). El hecho de no haberlo incluido explícitamente en este apartado se debe a que la pregunta planteada pone el foco en los factores preexistentes al inicio del curso, mientras que en el artículo se utilizan resultados académicos tempranos del mismo curso bajo estudio. Aun así, dicho artículo establece los aspectos metodológicos que se trabajan en (Martínez-Carrascal, Hlosta, et al., 2023). En una línea similar, en (Martínez-Carrascal & Sancho-Vinuesa, 2023) se aplica el método para extender los resultados obtenidos más allá de la parte inicial de curso, trasladando además en análisis del plano del comportamiento del estudiante al plano del diseño de la evaluación.

5.4. ¿Cómo influyen los resultados académicos iniciales de un determinado curso en la posibilidad de abandonarlo? (RQ3)

Varios de los artículos presentados incluyen reflexiones sobre el impacto de la actividad temprana, tanto en el abandono como en el rendimiento académico en sentido amplio (en particular (Martínez-Carrascal, Márquez Cebrián, et al., 2020; Martínez-Carrascal & Sancho-Vinuesa, 2021). El último de estos artículos (Martínez-Carrascal & Sancho-Vinuesa, 2021) aborda específicamente la cuestión planteada en este apartado, y muestra que **un resultado académico negativo en la primera prueba evaluativa** de un curso **incrementa el riesgo de abandono en un factor 2.57 en promedio**, y que puede llegar a ser del 4.22 dependiendo del curso analizado.

Incidimos de nuevo no solo en la respuesta a la pregunta sino en la relevancia de la metodología. Si bien la relevancia de esta actividad temprana puede obtenerse por medios estadísticos clásicos, el análisis de supervivencia aporta información temporal relevante. Más allá de las ratios finales, muestra que los estudiantes que se fracasan en las etapas iniciales del curso son más propensos a presentar un abandono temprano. Este hecho se ha manifestado en todos los cursos analizados, y puede verse de forma gráfica en la Figura 2 del artículo específico sobre este tema (Martínez-Carrascal & Sancho-Vinuesa, 2021).

Estos resultados establecen una relación clara entre la evaluación del rendimiento en etapas iniciales del curso y su posibilidad de superación. Constituye por tanto un factor predictivo relevante. Aun cuando el foco de la pregunta estaba centrado en el abandono, nuestra exploración en trabajos previos permite también contextualizar el impacto de la actividad inicial en el rendimiento.

Un artículo previo (Martínez-Carrascal, Márquez Cebrián, et al., 2020) anticipaba ya hasta qué punto la actividad temprana en su conjunto es relevante. Nuestros resultados muestran que el análisis de la actividad temprana, incluso obviando los resultados de posibles evaluaciones, presenta una capacidad predictiva. A la vez, pone de manifiesto que esta capacidad predictiva es limitada, y que de nuevo requiere de interpretación pedagógica. A modo de ejemplo, dicho artículo pone de manifiesto que solo una diferenciación clara entre estudiantes repetidores y no repetidores permite validar la relevancia estadística de aspectos específicos de la actividad temprana (mediante la obtención de modelos con valores de Kappa en intervalos razonables para considerarlos como válidos). La combinación de actividad y resultado académico temprano sí que da lugar a modelos con mayor capacidad predictiva. Remarcar que este estudio no se focaliza en el abandono sino en considerar el rendimiento a nivel de curso en términos de superación.

Este mismo artículo muestra la relevancia de plantear diseños pedagógicos y métodos orientados a motivar el compromiso del estudiante. Estos diseños constituyen los caminos de aprendizaje que el docente sugiere seguir al estudiante, y que han sido analizados en las publicaciones presentadas en el Capítulo 4. Este concepto de camino de aprendizaje orienta también hacia la consideración no sólo de visiones estáticas del proceso, sino del análisis de la trayectoria de aprendizaje seguida a lo largo del curso.

5.5. ¿Cómo puede validarse la bondad de un camino de aprendizaje a la hora de superar la asignatura? (RQ4)

Antes de abordar las dos cuestiones restantes conviene sintetizar algunas de las reflexiones que se han derivado de las preguntas de investigación ya abordadas. Los trabajos planteados han mostrado la relevancia de ciertos factores en el bajo rendimiento o el abandono. Se ha evolucionado desde el análisis del impacto en visiones estáticas hacia la consideración de la evolución temporal. Se ha puesto de manifiesto la importancia de la contextualización pedagógica. Finalmente, y como colofón del apartado previo, se ha determinado la conveniencia de establecer caminos de aprendizaje que ayuden al estudiante a superar la asignatura.

Este razonamiento sugiere que el docente debe establecer caminos de aprendizaje adecuados así como disponer de métodos para determinar hasta qué punto un estudiante sigue el camino pautado o para corregir la estrategia definida inicialmente. Los caminos de aprendizaje ya están presentes en la literatura revisada (Govindarajan et al., 2017; Maciej Serda et al., 2011; Nabizadeh et al., 2020; Ramos et al., 2021; Rodríguez-Muñiz et al., 2019). Sin embargo, y como muestra nuestro artículo (Martínez-Carrascal, Munoz-Gama, et al., 2023), no existen métodos que establezcan de forma cuantitativa hasta qué punto existe una relación entre seguir un determinado camino – en concreto, el pautado por el equipo docente – y obtener un mejor resultado académico. Tampoco existen métodos para cuantificar el seguimiento que realiza el estudiante de dicho camino.

Por estas razones, segmentamos el análisis del proceso de aprendizaje y su impacto en el rendimiento académico en dos cuestiones. En primer lugar, convendrá **establecer que el camino planteado efectivamente ayuda a llegar al destino** (superar la asignatura en nuestro caso). Esto ayudará a la definición de caminos efectivos en lo que se refiere a la correlación entre seguirlo y llegar al resultado esperado. Una vez establecida la bondad del camino, será el momento de analizar **cómo pueden detectarse las divergencias**, y hasta qué punto conviene intervenir ante divergencias relevantes del camino pautado.

Hecha esta reflexión, la pregunta abordada en este apartado focaliza en el primero de los aspectos. La aproximación preliminar al problema se ha realizado en dos artículos (Martínez-

Carrascal, Valderrama Vallés, et al., 2020; Martínez-Carrascal & Sancho-Vinuesa, 2022). Partiendo de la exploración de secuencias, se ha buscado la validación de resultados de investigaciones previas (Figueroa-Cañas & Sancho-Vinuesa, 2021) mediante nuevos métodos. En particular, se ha establecido la relevancia de la realización de cuestionarios antes de abordar las pruebas evaluativas en un curso de matemáticas *online*.

Sin embargo, el gran avance es el salto desde el análisis de un aspecto concreto o una secuencia en particular al modelado del proceso global de aprendizaje. Realizada la aproximación preliminar, (Martínez-Carrascal, Munoz-Gama, et al., 2023) propone un método de modelado de caminos de aprendizaje que permite evaluar la relevancia de seguir dicho camino. La propuesta metodológica responde a la pregunta planteada en ese apartado, pero constituye además una metodología muy apropiada para mejorar la comprensión de los procesos de aprendizaje y guiar la revisión de estrategias docentes existentes.

Las bases teóricas de la aproximación realizada se basan en **el uso de metodologías de minería de procesos (*process mining*).** La aproximación novedosa en lo que respecta al uso de *skeletons* permite extraer los puntos clave de un proceso – en nuestro caso, del proceso de aprendizaje – y una vez modelado, evaluar la relevancia de seguir un determinado camino. Nuestra investigación ha dado como resultado un método que consta de las siguientes fases, según se describe en (Martínez-Carrascal, Munoz-Gama, et al., 2023):

1. Preproceso del log con una visión de proceso de aprendizaje (*event abstraction*).
2. Extracción de trayectorias de los estudiantes que siguen el proceso sugerido por el docente desde el punto de vista pedagógico (*scoping*).
3. Obtención del *skeleton* asociado a las trayectorias de los estudiantes que siguen dicho proceso.
4. Cuantificación de las diferencias entre el proceso realizado y el modelo generado en el punto anterior (*conformance*).

El análisis de **la correlación entre el índice de conformidad del punto 4 y el resultado académico permite validar la bondad de un determinado camino de aprendizaje**. El proceso detallado se describe en la Sección 3 del artículo (Martínez-Carrascal, Munoz-Gama, et al., 2023).

Más allá de la propuesta metodológica, el artículo realiza la validación de un camino específico asociado a un curso *online*. Esta validación demuestra que el seguimiento del método descrito permite dar respuesta específica a la pregunta planteada en este apartado. En particular, y para el caso concreto de estudio, existe una correlación positiva entre el seguimiento del camino establecido por el equipo docente y el resultado académico obtenido por el estudiante.

Es particularmente relevante el hecho de poder cuantificar el ajuste a un determinado camino propuesto con un indicador cuyo valor absoluto de una visión clara del ajuste. Los enfoques tradicionales al *conformance* (basados en técnicas de *log matches model* o *model matches log*) permiten obtener un índice, pero el valor numérico de dicho índice depende de la complejidad del proceso subyacente, por lo que es difícil hacer una interpretación directa. Su valor depende finalmente de la complejidad del modelo, y de los estados por los que pasa la ejecución real que presenta una determinada ejecución. En cambio, el uso de *skeletons* para este propósito permite una interpretación clara, que además, indica cuán cerca se está del ajuste completo (índice con valor 1) o de ser completamente diferente (índice con valor 0),

El uso de este enfoque constituye una contribución relevante al análisis de procesos de aprendizaje. En este caso no se trata únicamente de la aplicación de una nueva metodología a un campo científico sino de una aproximación conceptualmente diferente. La comparación de procesos se ha demostrado particularmente complicada. Los procesos de aprendizaje son altamente complejos, ya que no solo presentan un alto número de estados, sino también una alta variabilidad en el orden de realización. Finalmente, cada estudiante sigue un camino, pero el método descrito permite cuantificar hasta qué punto ese camino refleja la idea sugerida por el equipo docente.

Como cierre a la pregunta planteada en este apartado, queremos indicar que el método expuesto permite la aplicación directa tanto en escenarios de *A/B testing* para evaluar el camino más conveniente, como en la evaluación de cuáles son los caminos óptimos en base a los resultados obtenidos por los estudiantes. Se contribuye de esta forma a la mejora de la calidad docentes y finalmente de resultados académicos, objetivo final del presente trabajo.

5.6. ¿Cómo puede establecerse una medida cuantitativa de hasta qué punto el estudiante está siguiendo el camino pautado por el docente? (RQ5)

La respuesta sintética a esta pregunta está en el uso de *skeletons*, y ha quedado parcialmente cubierta en el punto anterior. Desde nuestro punto de vista, y más allá de lo comentado en la pregunta previa, conviene subrayar el potencial de este concepto.

La posibilidad de cuantificar la proximidad a un camino pautado en una escala claramente comprensible (0 indicando alejamiento completo, 1 ajuste perfecto) es una aportación del *skeleton* que no presentan los algoritmos de *conformance* comúnmente utilizados (y en particular, los de '*match*' del *log vs model* o *model vs log*).

Estudios recientes han establecido el poder de este método como clasificador (Verbeek & de Carvalho, 2018). Basándonos en nuestro artículo (Martínez-Carrascal, Munoz-Gama, et al., 2023) consideramos que esta potencialidad es trasladable al análisis de caminos de aprendizaje en el ámbito educativo. Conviene para ello validar en primer lugar que seguir el camino pautado es relevante para superar el curso (según visto en RQ4).

Contando con la conveniencia de seguir el camino pautado, el seguimiento del camino por parte del estudiante constituye una potencial variable predictiva de ayuda en la selección de estudiantes susceptibles de una intervención.

Creemos que el uso de *skeletons* tiene también una importancia conceptual más allá de la respuesta a la pregunta. Los artículos en los que hemos trabajado la modelización de procesos – tanto los preliminares (Martínez-Carrascal, Valderrama Vallés, et al., 2020; Martínez-Carrascal & Sancho-Vinuesa, 2022) como el que completa el modelado de los caminos de aprendizaje (Martínez-Carrascal, Munoz-Gama, et al., 2023) – han mostrado la complejidad de los procesos educativos. Cada estudiante sigue un camino diferente, donde intervienen multitud de acciones y posibles derivaciones. El uso de *skeletons* contribuye de forma capital a simplificar un problema que de otra forma sería difícilmente resoluble. Podemos admitir de nuevo (como pasó con el caso del análisis de supervivencia) un símil con otras disciplinas científicas. El *skeleton* determina las características clave que presenta un proceso complejo, de forma similar a como ciertas características diferencian el genoma de las especies. Comparar las características específicas de un individuo con su genoma de referencia, permite determinar hasta qué punto pertenece a una especie. De la misma forma, la comparación del *skeleton* de un proceso y una materialización concreta (es decir, la traza de un estudiante) nos indicará hasta qué punto la actividad del estudiante es coherente con el proceso con el que se compara, que en nuestro caso de estudio, corresponde al camino de aprendizaje sugerido por el equipo docente.

5.7. Propuesta de acciones orientadas a mejorar los resultados académicos de los cursos analizados en base a la modelización realizada

Al margen de las preguntas de investigación, se planteó en la introducción la conveniencia de realizar una reflexión final focalizada en proporcionar recomendaciones orientadas a la mejora académica en base a los modelos planteados. Las cuestiones previas han contribuido a poner de manifiesto factores específicos que impactan en el abandono, a sugerir potenciales intervenciones y a proponer nuevas técnicas de análisis. Es el momento de ir más allá, buscando qué aportaciones adicionales derivan del proceso de investigación seguido.

Creemos en primer lugar, y como se ha mostrado, que las aproximaciones metodológicas planteadas son particularmente convenientes. En particular, consideramos que la componente dinámica aporta capacidad de interpretación y permite profundizar en aspectos que de otra forma quedarían ocultos. El tiempo es un factor clave en el aprendizaje, y tenerlo en cuenta en el análisis es fundamental. De la misma forma, no solo es relevante el punto en que se encuentra el estudiante en un momento del curso, sino cómo ha llegado a él.

Dicho esto, la exploración de técnicas realizada ha considerado un espectro amplio: desde aproximaciones más comunes como árboles de decisión, KNN o SVC a las que han constituido la parte troncal de este trabajo, análisis de supervivencia y modelado de procesos, sin olvidar las técnicas de secuenciación. Como se ha comentado, consideramos que dos de ellas aportan aspectos diferenciales en potencial mejora del rendimiento a nivel de curso: el análisis de supervivencia y el modelado de procesos en base a *skeletons* para el modelado de caminos de aprendizaje.

Nuestra investigación sugiere aplicar el análisis de supervivencia para tratar el abandono. Entre las ventajas del método se incluyen la validación estadística del impacto de un factor, la cuantificación del impacto y la posibilidad de precisar el impacto combinado. Como aspectos particularmente relevantes cuando se considera como fin último una potencial intervención, el establecimiento claro de grupos objeto de la intervención, así como los momentos clave en los que llevarla a cabo.

Conceptualmente, la aplicación de esta técnica deriva del hecho que puede establecerse de forma simple el *evento* de estudio, requisito del análisis de supervivencia. En particular, lo que preocupa no es sólo el abandono en términos binarios – abandona, no abandona –, sino el momento de abandono, ya que debe actuarse antes de dicho momento. Remarcamos el aspecto diferencial que supone no ya abandonar o no, sino tener una visión de la evolución estimada de la curva de abandono.

Sin duda, poder intervenir para reducir el abandono, es clave, pero a la vez, se pretende una mejora global del rendimiento. A nivel pedagógico se ha constatado la relevancia que la investigación actual pone en el proceso, y en particular en los caminos de aprendizaje. Se ha detectado un aspecto no cubierto, puesto que pese al volumen de investigación que analiza la conceptualización y el diseño de caminos de aprendizaje, no nos consta literatura previa que permita modelizar, o cuantificar la bondad o el seguimiento del camino pautado de forma sistemática.

Es esta una contribución clave a la mejora del proceso de aprendizaje a partir de su análisis. El uso de *process mining* es claramente trasladable a entornos educativos. Y remarcamos este hecho ya que sus aproximaciones iniciales estaban en procesos altamente estructurados, donde se esperan pocas divergencias. No es este el caso de los procesos educativos. La literatura existente en este ámbito mostraba la complejidad de aplicación, y con frecuencia, la difícil interpretación de ciertos resultados. Gracias a la contribución de los *skeletons*, se han podido obtener aquellas características clave que permiten modelar el proceso, y finalmente, identificar si el estudiante está siguiendo el camino deseado o no.

Adicionalmente, y como punto de ayuda al docente, la propuesta metodológica permite validar empíricamente si seguir un camino ayuda o no a superar la asignatura. Aun cuando no hemos analizado escenarios *A/B testing*, sí se ha apuntado cómo el método puede servir de base comparativa a la hora de establecer la conveniencia de uno u otro camino.

Como muestra esta exposición, son precisamente estos resultados del análisis los que contribuyen a la mejora planteada en el rendimiento. Todos los estudios presentados en esta tesis se han realizado sobre cursos reales. Ya desde el primero de los artículos (Martínez et al., 2019) se focalizó en no limitarse al análisis, sino en mostrar como el análisis se traduce en acciones específicas. A modo de ejemplo, hemos visto cómo la aplicación de las técnicas expuestas ha resultado en recomendaciones como:

- Establecer cursos propedéuticos para aquellos estudiantes que abordan los cursos con niveles académicos por debajo del deseable. Esta recomendación deriva particularmente del efecto negativo que demuestra en términos de rendimiento una baja nota de entrada o bien un nivel académico preexistente bajo. Dos de los estudios planteados (Martínez-Carrascal, Hlosta, et al., 2023; Martínez-Carrascal, Márquez Cebrián, et al., 2020) sugieren esta aproximación.
- Definir y proponer caminos diferentes de aprendizaje para estudiantes repetidores y no repetidores. Como se ha visto (Martínez-Carrascal, Márquez Cebrián, et al., 2020), la forma de abordar el aprendizaje de ambos grupos es diferente. Nuestra propuesta, y pensando en particular en mejorar el rendimiento del grupo repetidor, sería explorar el camino de aprendizaje de los estudiantes repetidores que finalmente superan el curso, para establecer - basado en el método de análisis sugerido en el punto 5.4 - un camino diferente para los estudiantes repetidores.
- Establecer modelos pedagógicos que fuercen el compromiso académico durante el período lectivo del curso. La relevancia de mantener la actividad de ha mostrado en específicamente en (Martínez et al., 2019; Martínez-Carrascal, Márquez Cebrián, et al., 2020; Martínez-Carrascal, Munoz-Gama, et al., 2023; Martínez-Carrascal, Valderrama Vallés, et al., 2020). De forma colateral en (Martínez-Carrascal & Sancho-Vinuesa, 2023) ya que muestra que mantener períodos largos sin evaluación deriva en mayor abandono.
- En línea con el punto anterior, y dado el impacto del tiempo sin procesos evaluativos, incentivar los métodos de evaluación continua. Se trata finalmente de mantener el

compromiso del estudiante con el curso. En esta misma línea, diseñar modelos de evaluación que no dejen intervalos largos entre actividades, pues se ha demostrado su relación con una mayor tasa de abandono (Martínez-Carrascal & Sancho-Vinuesa, 2021).

- Usar el valor predictivo de la primera evaluación. Para ello, se sugiere avanzar el período de la primera actividad evaluativa, tanto para informar al estudiante de su posible riesgo, como para combinarlo con acciones que busquen compensar ya de forma temprana posibles indicadores de abandono o bajo rendimiento.
- Forzar en la medida de lo posible el uso de cuestionarios de entrenamiento no evaluativos. Este punto ha sido detectado en la investigación previa (Figueroa-Cañas & Sancho-Vinuesa, 2021) y confirmado aplicando una metodología totalmente diferente en uno de los trabajos presentados (Martínez-Carrascal & Sancho-Vinuesa, 2022). Los casos de estudio en los que se basa esta recomendación son del ámbito de las ingenierías, por lo que creemos que esto es particularmente relevante en las disciplinas cuantitativas. Aun así, conviene revisar los diseños pedagógicos maximizando la práctica previa a la evaluación. Creemos además que este hecho va en la línea apuntada de refuerzo del compromiso del estudiante a lo largo del curso.
- Extender los períodos para ejercer el derecho de desistimiento de la matrícula para aquellos estudiantes que presentan una situación económica desfavorable. Creemos que este hecho, derivado de (Martínez-Carrascal, Hlosta, et al., 2023) reduciría en particular el abandono de este colectivo, y en particular, el abandono temprano derivado potencialmente de motivos no estrictamente académicos.
- Establecer caminos de aprendizaje en una asignatura, modelarlos y evaluar de forma periódica la divergencia de los estudiantes respecto a dichos modelos. Informar a los estudiantes en base a divergencias umbral que, de acuerdo con el modelo, están en el grupo de estudiantes en riesgo de no superar la asignatura.
- Más allá de definir caminos de aprendizaje, evaluarlos. La investigación planteada ha revelado que son muchos los estudios que versan sobre el diseño de caminos de aprendizaje, e incluso de caminos personalizados. Sin embargo, no siempre se evalúa hasta qué punto seguir el camino es realmente beneficioso, y raramente por técnicas cuantitativas reproducibles en cualquier otro curso. La propuesta realizada en

(Martínez-Carrascal, Munoz-Gama, et al., 2023) de modelar el camino propuesto, evaluar su seguimiento, y su relación con el rendimiento académico podría aplicarse a prácticamente cualquier curso, orientado a un ciclo de mejora continua que permita definir caminos más adecuados en relación a la superación del curso.

Elevando el plano de análisis, y buscando sintetizar más allá de las acciones concretas, descubrimos que las acciones planteadas pueden agruparse alrededor de tres ejes. Aquellas que buscan compensar el efecto negativo de factores preexistentes. Aquellas que se centran en evitar el '*disengagement*' y finalmente aquellas que se realizan a-priori buscando la mejora de los métodos pedagógicos, mediante el diseño de mejores caminos de aprendizaje. Hemos llegado así desde el análisis de método al planteamiento de intervenciones. Una intervención pertinente y realizada a tiempo es clave en la mejora del proceso de aprendizaje, y el trabajo planteado proporciona herramientas para diseñarla.

No queremos acabar sin remarcar un aspecto del título que no debe pasar por alto. Nuestro foco – como se justificó en la introducción – se ha centrado en el análisis de cursos en educación superior *online*. Es este un aspecto fundamental, ya que el componente *online* comporta de forma intrínseca la posibilidad de disponer de datos asociados al comportamiento del estudiante para su análisis. No hay inconveniente en utilizar las aproximaciones planteadas a entornos que no hagan uso de esta metodología, siempre y cuando se disponga de formas de obtener los datos. Por otra parte, y en lo que respecta a la educación superior, los métodos de análisis planteados son perfectamente trasladables a otros niveles educativos. No lo son necesariamente los resultados específicos que derivan de la aplicación de las técnicas planteadas, que requerirían una validación formal en cursos asociados a dichos niveles.

Aun así, la contribución del factor *online* va más allá de los datos que permite obtener. Datos de actividad de docentes y estudiantes, de qué y cuándo se realizan diferentes acciones, y cuyo análisis proporciona datos valiosísimos. Sin embargo, esto no debe ocultar la flexibilidad que aportan estos entornos, la contribución a la democratización en el acceso al conocimiento y la creciente penetración en casi cualquier nivel educativo bien como aspecto nuclear o de soporte. Todos estos factores hacen que los estudios *online* deban ser considerados fundamentales en la educación del siglo XXI.

Nuestro trabajo ha llevado a cabo una exploración de técnicas, pero no querríamos clasificarlo – o no solo – dentro del marco del *educational data mining*. Cuanto más se ha ahondado en la reflexión, más se ha necesitado conocer los detalles intrínsecos de los cursos objeto de estudio. Por poner un ejemplo simple y paradigmático, no parece razonable desde un punto de *data mining* que los mejores estudiantes presenten mayores tasas de abandono en un curso. Lo ha sido cuando se ha contextualizado como curso propedéutico. Valga este ejemplo para justificar la diferencia entre la minería de datos – incluida la educativa – y el enfoque de *learning analytics* que ha guiado el presente estudio.

5.8. Reflexión final

Los puntos anteriores han dado respuesta a las preguntas que se plantearon al inicio de la presente tesis, incluyendo propuestas de acción que derivan no de percepciones sino del análisis sistemático de los datos contextualizados. Creemos aun así que la contribución del presente trabajo va más allá de dichas propuestas, y que abre interesantes líneas de investigación futura.

Ciertamente, la aplicación de las técnicas exploradas contribuye a la mejora académica a nivel de curso. Constituyen herramientas de ayuda al diseño de intervenciones desde la perspectiva de *learning analytics*. En particular, ayudando a la selección de grupos objeto de la intervención, definiendo medidas específicas y estableciendo el momento en que deben realizarse. Igualmente, queremos subrayar la relevancia de analizar el aprendizaje como proceso, siempre buscando la visión evolutiva. Tanto el análisis de supervivencia como las técnicas de minería de procesos van en esta línea.

La contribución metodológica se ha validado siempre con casos de estudio, en particular, de asignaturas. Esto permite por un lado dar validez a los métodos planteados, pero adicionalmente ha permitido definir acciones concretas para reducir el abandono y mejorar el rendimiento en los cursos objeto de estudio. Es remarcable que en estas acciones figuran los tres ejes clave que se detectaron en los modelos iniciales – en particular del abandono –: el estudiante (por el impacto de factores preexistentes y su compromiso con el curso), la institución (por el impacto de aspectos como la política de evaluación o la de abandono sin penalización económica), y la docencia (con el diseño de caminos de aprendizaje adecuados). Mejorar el rendimiento comporta un enfoque holístico que incluya a todos ellos.

Queremos poner en valor e incentivar la portabilidad de técnicas entre disciplinas científicas. Las técnicas exploradas en este trabajo no son nuevas. Sí lo es el cómo aplicarlas. En nuestro caso, la idea del análisis de supervivencia, más allá de ser explorado en algunos estudios, derivó del planteamiento del problema con una perspectiva médica. El análisis de procesos por su parte es comúnmente analizado en entornos empresariales. Su traslación a procesos educativos produjo inicialmente resultados de difícil interpretación, pero a su vez, provocó la consideración del uso de *skeletons* para soslayar este hecho, derivando en una propuesta metodológica innovadora para el análisis de caminos de aprendizaje. Por este motivo, alentamos la aplicación de herramientas existentes a ámbitos diferentes, o con matices específicos. Como se ha visto, puede dar lugar a propuestas metodológicas novedosas para resolver los problemas derivados de la portabilidad entre disciplinas.

Se abren además líneas futuras de particular interés, como se ha reflejado en los diferentes artículos. De forma sintética, proponemos dos ámbitos específicos para futuras extensiones de esta investigación. El primero, la aplicación de los métodos planteados más allá del nivel de curso, y en particular a nivel de titulaciones universitarias. El segundo, el uso de la metodología sugerida para el análisis de caminos de aprendizaje en el aula, focalizada en el diseño de intervenciones. Visionamos específicamente la evaluación comparativa de caminos de aprendizaje en escenarios A/B *testing*, cuantificando la mejora académica que deriva de la aplicación de la metodología planteada.

Y es con esta idea de mejora académica con la que cerramos el presente trabajo. La misma que nos motivó a iniciarlo y la que ha estado detrás de todos los trabajos presentados. Una idea que entronca con la reflexión que inició la presente tesis: la importancia de la educación para nuestro progreso como individuos y como sociedad. Una importancia de la que todos y siempre deberíamos ser conscientes.

Referencias

Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. *Proceedings - International Conference on Data Engineering*, 3–14. https://doi.org/10.1109/icde.1995.380415

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and *online* learning. *Computers in Human Behavior*, *31*(1), 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016). Survival analysis based framework for early prediction of student dropouts. *International Conference on Information and Knowledge Management, Proceedings*, 903–912. https://doi.org/10.1145/2983323.2983351

Antoninis, M., April, D., Barakat, B., Bella, N., D'Addio, A. C., Eck, M., Endrizzi, F., Joshi, P., Kubacka, K., McWilliam, A., Murakami, Y., Smith, W., Stipanovic, L., Vidarte, R., & Zekrya, L. (2020). All means all: An introduction to the 2020 Global Education Monitoring Report on inclusion. *Prospects*, *49*(3–4). https://doi.org/10.1007/s11125-020-09505-x

Bawa, P. (2016). Retention in *online* courses: exploring issues and solutions— A literature review. *SAGE Open*, *6*(1), 1–11. https://doi.org/10.1177/2158244015621777

Bean, J. P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American Educational Research Journal*, *22*(1), 35–64. https://doi.org/10.3102/00028312022001035

Bloom, B. S. (1974). Time and learning. *American Psychologist*, *29*(9), 682.

Bloom, B. S., & others. (1971). *Handbook on formative and summative evaluation of student learning.*

Blundell, R., Dearden, L., & Sianesi, B. (2005). Evaluating the effect of education on earnings: Models, methods and results from the national child development survey. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *168*(3). https://doi.org/10.1111/j.1467-985X.2004.00360.x

Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1), e1230. https://doi.org/10.1002/widm.1230

Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do We Know Who Will Drop Out?: A Review of the Predictors of Dropping out of High School: Precision, Sensitivity, and Specificity. *The High School Journal*. https://doi.org/10.1353/hsj.2013.0000

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003a). Survival analysis part II: Multivariate data analysis- An introduction to concepts and methods. In *British Journal of Cancer* (Vol. 89, pp. 431–436). https://doi.org/10.1038/sj.bjc.6601119

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003b). Survival analysis part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit. In *British Journal of Cancer* (Vol. 89, pp. 605–611). https://doi.org/10.1038/sj.bjc.6601120

Brunello, G., & Comi, S. (2004). Education and earnings growth: Evidence from 11 European countries. *Economics of Education Review*, *23*(1). https://doi.org/10.1016/S0272-7757(03)00048-7

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity-propensity analysis. *Contemporary Educational Psychology*, *32*(4). https://doi.org/10.1016/j.cedpsych.2006.09.002

Cabrera, A. F., Castaneda, M. B., Nora, A., & Hengstler, D. (1992). The convergence between two theories of college persistence. *The Journal of Higher Education*, *63*(2), 143–164. https://doi.org/10.2307/1982157

Card, D. (1999). Chapter 30 The causal effect of education on earnings. In *Handbook of Labor Economics: Vol. 3 PART* (Issue 1). https://doi.org/10.1016/S1573-4463(99)03011-4

Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, *64*(8), 1–9.

Casanova, J. R., Cervero, A., Núñez, J. C., Almeida, L. S., & Bernardo, A. (2018). Factors that determine the persistence and dropout of university students. *Psicothema*, *30*(4), 408–414. https://doi.org/10.7334/psicothema2018.155

Cassidy, S. (2012). Exploring individual differences as determining factors in student academic achievement in higher education. *Studies in Higher Education*, *37*(7). https://doi.org/10.1080/03075079.2010.545948

Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. *Studies in Computational Intelligence*. https://doi.org/10.1007/978-3-540-71974-8_8

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers and Education*, *96*, 42–54. https://doi.org/10.1016/j.compedu.2016.02.006

Cho, M.-H., & Shen, D. (2013). Self-regulation in *online* learning. *Distance Education*, *34*(3), 290–301.

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003a). Survival analysis part I: Basic concepts and first analyses. In *British Journal of Cancer* (Vol. 89, pp. 232–238). https://doi.org/10.1038/sj.bjc.6601118

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003b). Survival analysis part IV: Further concepts and methods in survival analysis. In *British Journal of Cancer* (Vol. 89, pp. 781–786). https://doi.org/10.1038/sj.bjc.6601117

Cochran, J. D., Campbell, S. M., Baker, H. M., & Leeds, E. M. (2014). The Role of Student Characteristics in Predicting Retention in *Online* Courses. *Research in Higher Education*, *55*(1), 27–48. https://doi.org/10.1007/s11162-013-9305-8

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Commission, E. (2017). 2017.05.30. A renewed EU agenda for higher education. *Official Journal of the European Union ISSN*.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2016.2616312

Dutt, A., Aghabozrgi, S., Ismail, M. A. B., & Mahroeian, H. (2015). Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering*, *5*(2), 112.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*. https://doi.org/10.1109/ACCESS.2017.2654247

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*. https://doi.org/10.1146/annurev.psych.53.100901.135153

Emmert-Streib, F., & Dehmer, M. (2019). Introduction to survival analysis in practice. *Machine Learning and Knowledge Extraction*, *1*(3), 1013–1038. https://doi.org/10.3390/make1030058

European Commission. (2020). Education and Training Monitor 2020 Teaching and learning in a digital age. *Brussels: European Commission*.

Figueroa-Cañas, J., & Sancho-Vinuesa, T. (2021). Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous *online* calculus module. *Interactive Learning Environments*, *29*(1). https://doi.org/10.1080/10494820.2018.1559864

Findley, M. J., & Cooper, H. M. (1983). Locus of control and academic achievement: A literature review. *Journal of Personality and Social Psychology*, *44*(2). https://doi.org/10.1037/0022-3514.44.2.419

Fournier-Viger, P., Chun, J., Lin, -Wei, Kiran, R. U., Koh, Y. S., Thomas, R., Lin, W., Kiran, R. U., Koh, Y. S., & Thomas, R. (2017). A Survey of Sequential Pattern Mining. *Ubiquitous International*, *1*(1).

Fournier-Viger, P., Lin, J. C. W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., & Lam, H. T. (2016). The SPMF open-source data mining library version 2. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-46131-1_8

Gage, N. L., & Berliner, D. C. (1992). Educational psychology, 5th ed. In *Educational psychology, 5th ed.* Houghton, Mifflin and Company.

Ghazal, M. A., Ibrahim, O., & Salama, M. A. (2018). Educational process mining: A systematic literature review. *Proceedings - 2017 European Conference on Electrical Engineering and Computer Science, EECS 2017*, 198–203. https://doi.org/10.1109/EECS.2017.45

Gough, H. G. (1953). What determines the academic achievement of high school students. *Journal of Educational Research*, *46*(5), 321–332. https://doi.org/10.1080/00220671.1953.10882027

Govindarajan, K., Kumar, V. S., & Kinshuk. (2017). Dynamic Learning Path Prediction - A Learning Analytics Solution. *Proceedings - IEEE 8th International Conference on Technology for Education, T4E 2016*. https://doi.org/10.1109/T4E.2016.047

Hachey, A. C., Conway, K. M., Wladis, C., & Karim, S. (2022). Post-secondary *online* learning in the U.S.: an integrative review of the literature on undergraduate student characteristics. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-022-09319-0

Hai-Jew, S. (2017). Using Decision Trees to Analyze *Online* Learning Data. *International Symposium for Innovative Teaching and Learning*. https://doi.org/10.4148/2573-4911.1005

Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. v, Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., & Liao, S. N. (2018). Predicting Academic Performance: A Systematic Literature Review. *Proceedings Companion of the 23rd Annual ACM*

Conference on Innovation and Technology in Computer Science Education, 175–199. https://doi.org/10.1145/3293881.3295783

Higher education initiatives | European Education Area. (n.d.). Retrieved December 27, 2022, from https://education.ec.europa.eu/education-levels/higher-education/about-higher-education

Hoge, R. D., & Luce, S. (1979). Predicting academic achievement from classroom behavior. *Review of Educational Research*, *49*(3), 479–496.

Holder, B. (2007). An investigation of hope, academics, environment, and motivation as predictors of persistence in higher education *online* programs. *Internet and Higher Education*. https://doi.org/10.1016/j.iheduc.2007.08.002

Huitt, W. (2000, August). *Educational Psychology Interactive: Teacher efficacy*. https://www.edpsycinteractive.org/topics/teacher/tcheff.html

Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, *26*(1). https://doi.org/10.1007/s10639-020-10230-3

Kudo, M., Ishida, A., & Sato, N. (2013). Business process discovery by using process skeletonization. *Proceedings - 2013 International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2013*. https://doi.org/10.1109/SITIS.2013.158

Landis, J. R., & Koch, G. G. (1977). No Title. *Biometrics*, *33*(1). https://doi.org/10.2307/2529310

Larsen, M. S., Kornbeck, K. P., Kristensen, R. M., Larsen, M. R. &, & Sommersel, H. B. (2013). *Dropout phenomena at universities: What is dropout? why does dropout occur? what can be done by the universities to prevent or reduce it? : A systematic review*. Danish Clearinghouse for Educational Research.

Lee, Y., & Choi, J. (2011). A review of *online* course dropout research: Implications for practice and future research. In *Educational Technology Research and Development* (Vol. 59, Issue 5, pp. 593–618). https://doi.org/10.1007/s11423-010-9177-y

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers and Education*. https://doi.org/10.1016/j.compedu.2009.09.008

Maciej Serda, Becker, F. G., Cleary, M., Team, R. M., Holtermann, H., The, D., Agenda, N., Science, P., Sk, S. K., Hinnebusch, R., Hinnebusch A, R., Rabinovich, I., Olmert, Y., Uld, D. Q. G. L. Q., Ri, W. K. H. U., Lq, V., Frxqwu, W. K. H., Zklfk, E., Edvhg, L. V, … (2011) .فاطمی, ح. Evaluation of the Learning Path Specification. *Educational Technology & Society*, *14*(3), 218–230. https://doi.org/10.2/JQUERY.MIN.JS

Martínez, J. A., Campuzano, J., Sancho-Vinuesa, T., & Valderrama, E. (2019). Predicting student performance over time. A case study for a blended-learning engineering course. *CEUR Workshop Proceedings*, *2415*.

Martínez-Carrascal, J. A., Hlosta, M., & Sancho-Vinuesa, T. (2023). Using Survival Analysis to Identify Populations of Learners at Risk of Withdrawal: Conceptualization and Impact of Demographics. *The International Review of Research in Open and Distributed Learning*, *24*(1), 1–21. https://doi.org/10.19173/irrodl.v24i1.6589

Martínez-Carrascal, J. A., Márquez Cebrián, D., Sancho-Vinuesa, T., & Valderrama, E. (2020). Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course. *Computer Applications in Engineering Education*, *28*(3), 590–605. https://doi.org/10.1002/cae.22229

Martínez-Carrascal, J. A., Munoz-Gama, J., & Sancho-Vinuesa, T. (2023). Evaluation of Recommended Learning Paths using Process Mining and Log Skeletons: Conceptualization and Insight into an *Online* Mathematics Course. *IEEE Transactions on Learning Technologies*, doi: 10.1109/TLT.2023.3298035.

Martínez-Carrascal, J. A., & Sancho-Vinuesa, T. (2021). *Does failing the first assessment affect withdrawal decisions at course level? An empirical evidence* (Á. Hernández-García, D.

Hernández-Leo, M. Caeiro-Rodríguez, & T. Sancho-Vinuesa, Eds.). CEUR-WS.org. https://ceur-ws.org/Vol-3029/paper01.pdf

Martínez-Carrascal, J. A., & Sancho-Vinuesa, T. (2022). *Using Process Mining to determine the relevance and impact of performing non-evaluative quizzes before evaluative assessments*. https://ceur-ws.org/Vol-3238/paper7.pdf

Martínez-Carrascal, J. A., & Sancho-Vinuesa, T. (2023). Exploring the impact of assessment characteristics on course withdrawal: a survival analysis approach. Proceedings of the Learning Analytics Summer Institute Spain 2023 (*publication pending as part of the conference Proceedings)*.

Martínez-Carrascal, J. A., Valderrama Vallés, E., Sancho-Vinuesa, T., Martinez-Carrascal, J. A., & Sancho-Vinuesa, T. (2020). Combining clustering and sequential pattern mining to detect behavioral differences in log data: conceptualization and case study. In A. Martínez-Monés, A. Álvarez, M. Caeiro-Rodríguez, & Y. Dimitriadis (Eds.), *Learning Analytics Summer Institute*. http://ceur-ws.org/Vol-2671/paper04.pdf

McClelland, T. J. (2015). *Why do they leave? an exploration of situational, dispositional, institutional, technological, and epistemological factors on undergraduate student withdrawal from online studies at an institute of technology in New Zealand* [Northeastern University]. https://doi.org/10.17760/d20128397

McGregor, H. A., & Elliot, A. J. (2002). Achievement goals as predictors of achievement-relevant processes prior to task engagement. *Journal of Educational Psychology*, *94*(2), 381–395. https://doi.org/10.1037/0022-0663.94.2.381

McIlrath, D., & Huitt, W. (1995). The teaching-learning process: A discussion of models. *Educational Psychology Interactive*, 45–49.

Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., & Kloos, C. D. (2019). Prediction in MOOCs: A review and future research directions. *IEEE Transactions on Learning Technologies*, *12*(3), 384–401. https://doi.org/10.1109/TLT.2018.2856808

Muljana, P. S., & Luo, T. (2019). Factors contributing to student retention in *online* learning and recommended strategies for improvement: A systematic literature review. *Journal of Information Technology Education: Research*, *18*, 19–57. https://doi.org/10.28945/4182

Nabizadeh, A. H., Gonçalves, D., Gama, S., Jorge, J., & Rafsanjani, H. N. (2020). Adaptive learning path recommender approach using auxiliary learning objects. *Computers and Education*, *147*. https://doi.org/10.1016/j.compedu.2019.103777

Nebot, A., Castro, F., Mugica, F., & Vellido, A. (2006). Identification of fuzzy models to predict students performance in an E-learning environment. *Proceedings of the Fifth IASTED International Conference on Web-Based Education*.

Nistor, N., & Neubauer, K. (2010). From participation to dropout: Quantitative participation patterns in *online* university courses. *Computers and Education*, *55*(2), 663–672. https://doi.org/10.1016/j.compedu.2010.02.026

Nortvig, A. M., Petersen, A. K., & Balle, S. H. (2018). A literature review of the factors influencing e-learning and blended learning in relation to learning outcome, student satisfaction and engagement. *Electronic Journal of E-Learning*.

Nota, L., Soresi, S., & Zimmerman, B. J. (2004). Self-regulation and academic achievement and resilience: A longitudinal study. *International Journal of Educational Research*, *41*(3). https://doi.org/10.1016/j.ijer.2005.07.001

Organisation for Economic Co-Operation and Development (OECD). (2020). PISA 2018 Results (Volume V): Effective Policies, Successful Schools. In *OECD Publishing: Vol. V* (Issue 425).

Payandeh Najafabadi, A. T., Najafabadi, M. O., & Farid-Rohani, M. R. (2013). Factors contributing to academic achievement: A Bayesian structure equation modelling study. *International Journal of Mathematical Education in Science and Technology*, *44*(4). https://doi.org/10.1080/0020739X.2012.742149

Proctor, C. P. (1984). Teacher Expectations: A Model for School Improvement. *The Elementary School Journal*, *84*(4). https://doi.org/10.1086/461377

Ramos, D. B., Ramos, I. M. M., Gasparini, I., & De Oliveira, E. H. T. (2021). A new learning path model for e-learning systems. *International Journal of Distance Education Technologies*, *19*(2). https://doi.org/10.4018/IJDET.20210401.oa2

Reimann, P., Markauskaite, L., & Bannert, M. (2014). E-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540. https://doi.org/10.1111/bjet.12146

Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a learning analytics intervention and evaluation framework: What works? In B. Kei Daniel (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (pp. 147–166). Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5_10

Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in *online* learning; A decision tree based approach. *Computers and Education*, *137*(2). https://doi.org/10.1016/j.compedu.2019.04.001

Rodríguez-Muñiz, L. J., Bernardo, A. B., Esteban, M., & Díaz, I. (2019). Dropout and transfer paths: What are the risky profiles when analyzing university persistence with machine learning techniques? *PLOS ONE*, *14*(6), e0218796. https://doi.org/10.1371/journal.pone.0218796

Rosset, S. (2004). Model selection via the AUC. *Machine Learning, Proceedings of the 21st International Conference*.

Rovai, A. P. (2003). In search of higher persistence rates in distance education *online* programs. *Internet and Higher Education*, *6*(1), 1–16. https://doi.org/10.1016/S1096-7516(02)00158-6

Scheerens, J. (1990). School Effectiveness Research and the Development of Process Indicators of School Functioning. *School Effectiveness and School Improvement*, *1*(1), 61–80. https://doi.org/10.1080/0924345900010106

Seel, N. M. (2012). Bloom's Model of School Learning. In *Encyclopedia of the Sciences of Learning*. https://doi.org/10.1007/978-1-4419-1428-6_979

Setiawan, R., Aprillia, A., & Magdalena, N. (n.d.). *Analysis of antecedent factors in academic achievement and student retention*. https://doi.org/10.1108/AAOUJ-09-2019-0043

Shah, M., & Cheng, M. (2019). Exploring factors impacting student engagement in open access courses. *Open Learning*, *34*(2), 187–202. https://doi.org/10.1080/02680513.2018.1508337

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/j.procs.2015.12.157

Simpson, O. (2010). *'22% -can we do better?'-The CWP Retention Literature Review*. https://doi.org/10.13140/RG.2.2.15450.16329

Simpson, O. (2013). Student retention in distance education: are we failing our students? *Open Learning*, *28*(2). https://doi.org/10.1080/02680513.2013.847363

Spinath, B. (2012). Academic Achievement. In *Encyclopedia of Human Behavior: Second Edition* (pp. 1–8). https://doi.org/10.1016/B978-0-12-375000-6.00001-X

Steinmayr, Ricarda., Meißner, Anja., Weidinger, A. F., & Wirthwein, L. (2014). Academic Achievement - Education - Oxford Bibliographies. In *Oxford Bibliographies Online*.

Tamborini, C. R., Kim, C. H., & Sakamoto, A. (2015). Education and Lifetime Earnings in the United States. *Demography*, *52*(4). https://doi.org/10.1007/s13524-015-0407-0

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89–125. https://doi.org/10.3102/00346543045001089

Tsiakmaki, M., Pierrakeas, C., Kostopoulos, G., Kotsiantis, S., Koutsonikos, G., & Ragos, O. (2019). Predicting university students' grades based on previous academic achievements. *2018 9th International Conference on Information, Intelligence, Systems and Applications, IISA 2018*. https://doi.org/10.1109/IISA.2018.8633618 van der Aalst, W. (2012). Process mining: Overview and opportunities. In *ACM Transactions on Management Information Systems* (Vol. 3, Issue 2, pp. 1–17). https://doi.org/10.1145/2229156.2229157

Verbeek, H. M. W., & de Carvalho, R. M. (2018). *Log Skeletons: A Classification Approach to Process Discovery*. https://doi.org/10.48550/arxiv.1806.08247

Villano, R., Harrison, S., Lynch, G., & Chen, G. (2018). Linking early alert systems and student retention: A survival analysis approach. *Higher Education*, *76*(4), 903–920. https://doi.org/10.1007/s10734-018-0249-y

von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, *6*(6), 574–588. https://doi.org/10.1177/1745691611421204

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership*, *41*(8).

Watley, D. J., & Merwin, J. C. (1964). The effectiveness of variables for predicting academic achievement for business students. *The Journal of Experimental Education*, *33*(2), 189–192.

Wolter, S. C., Diem, A., & Messer, D. (2014). Drop-outs from Swiss Universities: An empirical analysis of data on all students between 1975 and 2008. *European Journal of Education*. https://doi.org/10.1111/ejed.12096

Xavier, M., & Meneses, J. (2020). *Dropout in online higher education: A scoping review from 2014 to 2018*. https://doi.org/10.7238/uoc.dropout.factors.2020

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in *online* learning. *Internet and Higher Education*, *29*, 23–30. https://doi.org/10.1016/j.iheduc.2015.11.003

Zimmerman, B. J., & Pons, M. M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal*, *23*(4). https://doi.org/10.2307/1163093