

## Analítica de Aprendizaje y Retención Estudiantil: un Estudio Longitudinal en Educación Superior Virtual

### Learning Analytics and Student Retention: A Longitudinal Study in Virtual Higher Education

 Ester Martín Caro Alamo<sup>1</sup>, Instituto Europeo de Posgrado (España)

 Nuria Segovia-García, Corporación Universitaria de Asturias (Colombia)

#### Resumen

La analítica de aprendizaje (AA) se ha destacado como una herramienta eficaz para abordar retos como la retención estudiantil y el rendimiento académico en programas de educación superior virtual. Este estudio tiene como objetivo analizar la relación entre los patrones de participación en actividades académicas digitales y el rendimiento académico de estudiantes de posgrado virtual, así como su vinculación con la permanencia en el sistema. Para ello, se abordan tres ejes fundamentales: los patrones y agrupaciones de participación a lo largo del tiempo; los factores que predicen el rendimiento y la retención; y el potencial de los modelos predictivos para personalizar estrategias pedagógicas. Se empleó un diseño cuantitativo, longitudinal y exploratorio, con una muestra de 393 estudiantes matriculados en una universidad colombiana durante 2024. Los datos, recolectados a través de sistemas institucionales, fueron analizados mediante modelos lineales mixtos, algoritmos de aprendizaje automático y análisis de series temporales. Las actividades académicas evaluadas incluyeron proyectos de aplicación, casos prácticos, exámenes, foros y participación general. Los resultados indican que las actividades prácticas, como los proyectos de aplicación y los casos prácticos, generan los mayores niveles de participación y tienen un impacto significativo en el rendimiento académico. En contraste, los foros y los exámenes muestran menores niveles de interacción. El análisis de clustering identificó tres grupos de estudiantes con patrones diferenciados de participación, lo que permite personalizar las estrategias pedagógicas. Se concluye que la AA ofrece una base sólida para diseñar intervenciones orientadas a mejorar la retención estudiantil y optimizar el rendimiento académico en la educación superior virtual. Este enfoque refuerza el valor de la personalización educativa basada en datos.

#### Abstract

Learning analytics (LA) has emerged as an effective tool to address challenges such as student retention and academic performance in virtual higher education programs. This study aims to analyze the relationship between participation patterns in digital academic activities and the academic performance of online postgraduate students, as well as their connection to student persistence. Three key dimensions are examined: participation trends and clusters over time; the factors that predict academic performance and retention; and the potential of predictive models to personalize pedagogical strategies. A quantitative, longitudinal, and exploratory design was employed, using a sample of 393 students enrolled at a Colombian university in 2024. Data was collected from institutional systems and analyzed through mixed linear models, machine learning algorithms, and time series analysis. The academic activities evaluated included application projects, practical cases, exams, discussion forums, and overall participation. Results indicate that practical activities, such as application projects and case studies, yield the highest levels of engagement and have a significant impact on academic performance. In contrast, forums and exams show lower levels of interaction. The clustering analysis identified three distinct student groups based on participation patterns, enabling more personalized pedagogical strategies. The findings suggest that LA provides a solid foundation for designing targeted interventions aimed at improving student retention and optimizing academic performance in virtual higher education. This approach reinforces the value of data-driven educational personalization.

#### Palabras clave / Keywords

Educación superior, Analítica de aprendizaje, Evaluación del estudiante, Educación a distancia, Participación estudiantil, Indicadores de rendimiento académico, Tecnología educativa, Deserción estudiantil.

Higher education, Learning analytics, student assessment, Distance education, Student engagement, Academic performance metrics, Educational technology, Student dropout.

<sup>1</sup> Autor de correspondencia: emartincaro@iep.edu.es

## 1. Introducción

La analítica de aprendizaje (AA) se define como el uso de datos y herramientas tecnológicas para comprender y optimizar los procesos educativos (Carmean et al., 2021). Este enfoque se ha consolidado como esencial en la educación superior, especialmente para abordar desafíos como la retención estudiantil y la personalización del aprendizaje (Quadri & Shukor, 2021). La AA permite diseñar intervenciones específicas mediante el análisis de patrones de comportamiento estudiantil, lo cual facilita la identificación de estudiantes en riesgo y la provisión de apoyo oportuno (Li et al., 2022).

Uno de los pilares de la AA son los modelos predictivos, los cuales emplean técnicas avanzadas —como el aprendizaje automático— para anticipar eventos como el abandono escolar o el éxito académico. Según McKinsey and Company (2022), estas herramientas no solo mejoran el rendimiento estudiantil, sino que también optimizan el uso de recursos y permiten personalizar las intervenciones educativas. Asimismo, técnicas como las redes bayesianas y la minería de datos han demostrado ser eficaces para predecir comportamientos asociados al abandono y al rendimiento (Nimy & Mosia, 2024). Un ejemplo destacado es el estudio de Arnold and Pistilli (2012) en Purdue, que evidenció cómo las alertas tempranas basadas en analítica predictiva lograron reducir significativamente las tasas de deserción.

Además de los modelos predictivos, herramientas como Power BI juegan un papel crucial en la implementación de la AA, al facilitar la visualización de tendencias y el diseño de estrategias adaptativas basadas en datos en tiempo real (IBM, 2022; Anthology, 2024). Estas soluciones convierten los datos cuantitativos en activos estratégicos, posibilitando una toma de decisiones ágil, informada y precisa (Segovia-García & Segovia-García, 2024).

Lo anterior subraya la importancia de trabajar con información de alta calidad que garantice la precisión de los análisis y predicciones. Para obtener resultados significativos que orienten la toma de decisiones, es fundamental que los datos sean consistentes y confiables (The Big Data Framework, 2024). Esto cobra especial relevancia al abordar fenómenos complejos como la retención estudiantil, entendida como la capacidad institucional para asegurar que los estudiantes completen su trayectoria académica, desde el ingreso hasta la titulación. La retención constituye un indicador clave del éxito institucional y de la calidad educativa. Además, implica no solo la permanencia en el sistema, sino también el progreso continuo y satisfactorio del estudiante, lo cual incide directamente en la reputación institucional, su estabilidad financiera y la percepción de calidad educativa (Villegas-Ch et al., 2023).

La complejidad de la retención estudiantil ha sido abordada en múltiples investigaciones, que consideran variables individuales, psicosociales, institucionales, académicas y contextuales. Estos trabajos han fundamentado teorías clásicas como el modelo de persistencia de Tinto (1987), el modelo de integración psicológica de Bean and Eaton (2000), y la teoría del involucramiento de Astin (1984). El auge de la educación virtual ha replanteado estos marcos, incorporando factores adicionales como la ruralidad, las condiciones socioeconómicas o la adaptación tecnológica, los cuales plantean nuevos desafíos y contribuyen al aumento de las tasas de deserción (Segovia-García et al., 2022).

Estos factores refuerzan la necesidad de implementar estrategias efectivas para garantizar la permanencia estudiantil (Guzmán et al., 2021), especialmente tras la pandemia de COVID-19, que aceleró los procesos de digitalización y obligó a las instituciones a replantear sus estrategias de retención. La AA ha demostrado su eficacia en este escenario. La experiencia de Gannon University evidencia que la integración de datos en tiempo real permitió desarrollar intervenciones efectivas para responder a las necesidades emergentes del estudiantado (Wong, 2021). Asimismo, la AA facilita el monitoreo del progreso académico y permite generar soluciones específicas orientadas a fortalecer la permanencia (Pulker, 2019).

El diseño de estrategias efectivas para mejorar la retención en entornos digitales requiere tanto herramientas tecnológicas avanzadas como un marco teórico robusto. Este estudio se apoya en el constructivismo y el conectivismo, teorías fundamentales para comprender el aprendizaje en línea. El constructivismo resalta la importancia de actividades significativas e interacción activa (Xu, 2019), mientras que el conectivismo destaca el papel de las redes digitales para fomentar el aprendizaje continuo (Downes, 2012). Ambas teorías respaldan el uso de la analítica de aprendizaje (AA) al permitir identificar patrones de comportamiento a partir de interacciones digitales, facilitando intervenciones educativas oportunas (Siemens & Long, 2011).

Esta relación es especialmente relevante en entornos virtuales, donde la participación es un fuerte predictor del éxito académico (Oudat et al., 2023). Revisiones sistemáticas confirman que una mayor frecuencia e intensidad en la participación —como en tareas, foros, evaluaciones y recursos asincrónicos— está vinculada con un mejor rendimiento y menor abandono (Li et al., 2022; Kaensar & Wongnин, 2023;

Nieuwoudt & Pedler, 2021). En esta línea, estudios recientes señalan que la interacción sostenida en actividades asincrónicas mejora el sentido de pertenencia y la persistencia del estudiante (Goicochea & Gómez, 2021; Horna, 2022; Horna & Seminario, 2023).

En este sentido, el análisis de la participación en plataformas de gestión del aprendizaje (LMS) incluye actividades como interacciones en foros, entrega de tareas, evaluaciones formativas y participación en sesiones sincrónicas. Estas actividades no solo constituyen indicadores clave de persistencia y éxito académico (Segovia-García et al., 2020), sino que también fomentan el compromiso estudiantil y reducen la sensación de aislamiento, un factor crítico en el aprendizaje a distancia (Segovia-García & Said-Hung, 2021). La retroalimentación derivada de estas interacciones fortalece la motivación y el progreso académico, consolidando la permanencia (ASEE Advances, 2020; Wang et al., 2024). Como señalan Tepgec et al. (2024), su efectividad no depende solo del acceso a los datos, sino de la capacidad del estudiante para interpretarlos y aplicarlos. Esto subraya la importancia de combinar el análisis predictivo con estrategias pedagógicas que promuevan el uso autónomo de la información, impulsando así una educación más personalizada y sostenida en el tiempo.

La AA, a través de modelos predictivos y técnicas analíticas avanzadas, permite anticipar conductas como el abandono y facilitar respuestas personalizadas. Esta información resulta clave para docentes y administradores al momento de tomar decisiones en tiempo real (Pulker, 2019; McKinsey & Company, 2022).

Asimismo, metodologías pedagógicas como el aula invertida, el aprendizaje basado en proyectos o la resolución de problemas, combinadas con estrategias de personalización educativa basadas en AA, promueven la participación y reducen el abandono (Digital Learning Edge, 2024; Macfadyen & Dawson, 2009). La integración de estas herramientas no solo optimiza el seguimiento académico, sino que también libera tiempo para diseñar experiencias de aprendizaje más significativas (Times Higher Education, 2024).

Esta relación entre participación y éxito académico ha sido ampliamente documentada en la literatura reciente. Revisiones sistemáticas vinculan directamente la frecuencia e intensidad de la participación en entornos virtuales con un mejor rendimiento académico y una menor probabilidad de abandono. Horna y Seminario Unzueta (2023) destacan que una participación constante en actividades asincrónicas, como tareas y evaluaciones, está positivamente relacionada con el desempeño académico. Otros autores como Goh (2025) o Horna (2022) subrayan que la interacción activa en plataformas virtuales, especialmente durante la pandemia, fue determinante para mantener el progreso académico. Cuanto más activa es la participación del estudiante —en tareas, foros, evaluaciones y recursos asincrónicos—, mayor es su sentido de pertenencia y persistencia dentro del entorno digital (Goicochea & Gómez, 2021). En esta misma línea, la analítica de aprendizaje ha demostrado ser una herramienta clave para personalizar la experiencia educativa en función de estos patrones de participación, adaptando contenidos y rutas formativas dentro de los LMS (Halkiopoulos & Gkintoni, 2024). Esta capacidad de adaptación no solo mejora el compromiso, sino que también contribuye a reducir el abandono, especialmente en entornos virtuales marcados por el aislamiento o la falta de retroalimentación significativa (Uzun et al., 2025).

Desde esta perspectiva, la retención estudiantil se aborda mediante tres enfoques interrelacionados: (1) factores predictivos como el rendimiento previo y el contexto familiar (Nieuwoudt & Pedler, 2021), (2) estrategias institucionales como tutorías o asesoría académica (Pulker, 2019), y (3) tecnologías avanzadas como la AA y la inteligencia artificial, que permiten personalizar la experiencia educativa (Martín-Caro, 2024).

A diferencia de estudios previos centrados en modelos descriptivos o variables aisladas, este trabajo adopta un enfoque longitudinal con análisis multinivel, técnicas de machine learning y una muestra amplia de estudiantes de posgrado. La combinación de clustering, series temporales y modelos predictivos permite abordar el objetivo de identificar patrones de participación y anticipar el rendimiento académico, ofreciendo bases sólidas para intervenciones educativas más personalizadas, precisas y escalables.

Para alcanzar este objetivo, se plantean las siguientes preguntas de investigación:

- ¿Cuáles son los patrones de participación digital predominantes entre los estudiantes a lo largo del tiempo, y cómo se agrupan según su comportamiento y características sociodemográficas?
- ¿Qué variables de participación y atributos del estudiantado predicen significativamente el rendimiento académico y la retención, y cómo varía esta relación en función del tipo de actividad desarrollada?
- ¿De qué manera pueden aprovecharse los resultados de los modelos predictivos y de segmentación para diseñar estrategias pedagógicas personalizadas e intervenciones estratégicas que fomenten la permanencia estudiantil?

## 2. Metodología

### 2.1. Muestra

La población del estudio estuvo conformada por 1,971 estudiantes matriculados en programas de especialización virtual de la Corporación Universitaria de Asturias durante el año 2024. A partir de esta población, se seleccionó una muestra representativa de 393 estudiantes, con un nivel de confianza del 95 % y un margen de error de  $\pm 5\%$ , asegurando diversidad en variables sociodemográficas como edad, género, ubicación geográfica y estrato socioeconómico.

Los criterios de inclusión fueron: estudiantes matriculados en programas de especialización en modalidad 100 % virtual, que hubieran cursado la totalidad de las asignaturas programadas durante el año 2024 y que contaran con registro completo de participación en la plataforma académica. Se excluyeron aquellos con registros incompletos o con menos del 70 % de participación registrada en las actividades evaluadas.

La muestra fue extraída aleatoriamente y refleja características similares a las de cohortes en programas virtuales de otras instituciones colombianas, lo que fortalece la validez externa y la posibilidad de generalización de los resultados a contextos educativos equivalentes.

El análisis descriptivo reveló una mayor proporción de mujeres (59.5 %) frente al 40.5 % de hombres. La mayoría de los estudiantes (63.8 %) eran mayores de 31 años, lo cual coincide con el perfil de estudiante no tradicional descrito por Sánchez-Gelabert y Elías (2017). Asimismo, el 68.7 % residía en zonas rurales, lo que subraya la necesidad de considerar limitaciones tecnológicas y de conectividad en el diseño pedagógico. En cuanto al estrato socioeconómico, el 51.1 % pertenecía a los estratos 3 y 4; el 46.1 %, a los estratos 1 y 2; y solo el 2.8 % se encontraba en los estratos 5 y 6. Este perfil evidencia la necesidad de apoyos adicionales como recursos tecnológicos y tutorías personalizadas para garantizar la equidad educativa.

### 2.2. Instrumentos

Se utilizaron dos fuentes institucionales para la recolección de los datos:

- Sistema CRM institucional: proporcionó información sociodemográfica (edad, género, zona de residencia, estrato socioeconómico), recolectada al inicio del ciclo lectivo.
- Tablero de control analítico (dashboard) en Power BI: consolidó datos provenientes del entorno virtual de aprendizaje (LMS), incluyendo interacción en foros, entregas de tareas, realización de evaluaciones, avance en las asignaturas y calificaciones finales.

Ambas bases se vincularon mediante un identificador personal único que garantizó la trazabilidad de la información y el anonimato de los estudiantes, cumpliendo con las normativas éticas y legales de protección de datos.

Los datos fueron validados mediante eliminación de duplicados, verificación cruzada entre sistemas y tratamiento de valores atípicos. Para ello se utilizaron funciones de validación automática en Power BI y rutinas de limpieza avanzadas con la biblioteca pandas en Python, lo que aseguró la confiabilidad y calidad de las mediciones.

### 2.3. Análisis de datos

El análisis se llevó a cabo en Google Colab utilizando Python, con bibliotecas especializadas como pandas, matplotlib, seaborn, scikit-learn y statsmodels. Estas herramientas, ampliamente reconocidas en la literatura académica (Kaensar & Wongnин, 2023), permitieron implementar modelos robustos y reproducibles. El procedimiento analítico se estructuró en cinco niveles:

- Análisis descriptivo: caracterización de la muestra según edad, género, zona de residencia y estrato socioeconómico.
- Modelo lineal mixto: análisis de la relación entre participación acumulada, calificaciones finales y retención estudiantil, considerando efectos fijos (tipo de actividad) y aleatorios (estudiante), además de las interacciones con variables sociodemográficas.
- Random Forest: modelo de clasificación para identificar actividades académicas con mayor poder predictivo sobre el rendimiento académico. Se aplicó validación cruzada (5-fold) y optimización de hiperparámetros mediante GridSearchCV.
- Clustering (K-means): agrupación de estudiantes en función de sus patrones de participación, permitiendo la personalización de estrategias pedagógicas.
- Series temporales (ARIMA): análisis de la evolución de la participación durante el año académico, para identificar momentos críticos de baja interacción y diseñar intervenciones oportunas.

### 3. Resultados

#### 3.1 Estado académico y desempeño

Del total de estudiantes analizados ( $n = 393$ ), el 96.4 % ( $n = 379$ ) se encuentra activo en sus estudios, mientras que el 3.6 % ( $n = 14$ ) ha discontinuado su matrícula.

En cuanto a la carga académica cursada, el 54.7 % ( $n = 215$ ) ha completado entre 10 y 12 asignaturas; el 42 % ( $n = 165$ ), entre 7 y 9 asignaturas; el 1.8 % ( $n = 7$ ), entre 4 y 6; y el 1.5 % ( $n = 6$ ), entre 1 y 3 asignaturas.

Respecto al promedio acumulado de calificaciones (escala de 0 a 5), el 63 % ( $n = 246$ ) se ubica entre 3.1 y 4.5; el 34 % ( $n = 134$ ), entre 4.6 y 5.0; y solo el 3 % ( $n = 13$ ) registra promedios inferiores a 3.0.

#### 3.2 Participación en actividades académicas

El modelo lineal mixto analizó 35,262 observaciones y detectó diferencias significativas entre estudiantes (varianza intraindividual: 0.990). Los efectos de cada tipo de actividad se detallan en la Tabla 1.

**Tabla 1.**

*Impacto de las actividades académicas en la participación estudiantil*

Actividad	Coeficiente	Significancia (p)	Interpretación
<b>Mayor Participación</b>			
Proyecto de Aplicación	5.775	p < .001	Mayor asociación positiva
Caso Práctico 3	+0.660	p < .001	Asociación positiva significativa
Caso Práctico 2	+0.356	p < .001	Asociación positiva significativa
<b>Menor Participación</b>			
Foro de Dudas	-10.119	p < .001	Valor más bajo de participación
Foro Unidad 1	-6.372	p < .001	Asociación negativa significativa
Foro Unidad 2	-6.250	p < .001	Asociación negativa significativa
Foro Unidad 3	-6.250	p < .001	Asociación negativa significativa
Exámenes	-9.170	p < .001	Menor participación promedio

Variables sociodemográficas:

- Edad: coeficiente = +0.018, p = .005 (positivo y significativo).
- Sexo: coeficiente = +0.215, p = .051 (marginalmente significativo).
- Zona de residencia y estrato socioeconómico: sin efectos significativos.

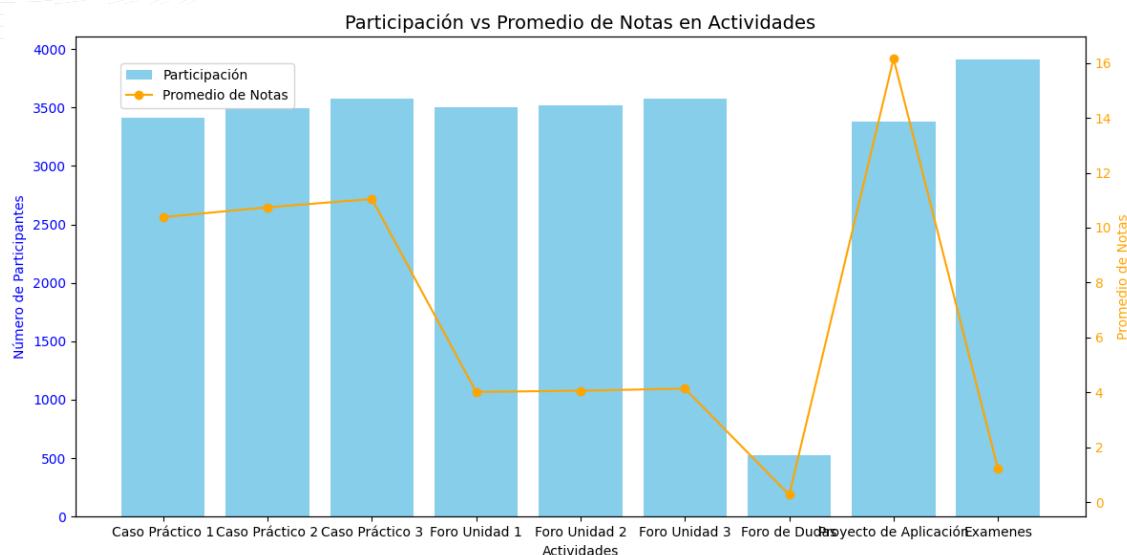
#### 3.3. Modelo predictivo (Random Forest)

El modelo Random Forest fue entrenado con variables de participación y calificaciones. Se implementó una validación cruzada de cinco pliegues (GridSearchCV) y ajuste de hiperparámetros:

- n\_estimators
- max\_depth
- min\_samples\_split
- min\_samples\_leaf

Las variables con mayor poder predictivo del rendimiento académico fueron:

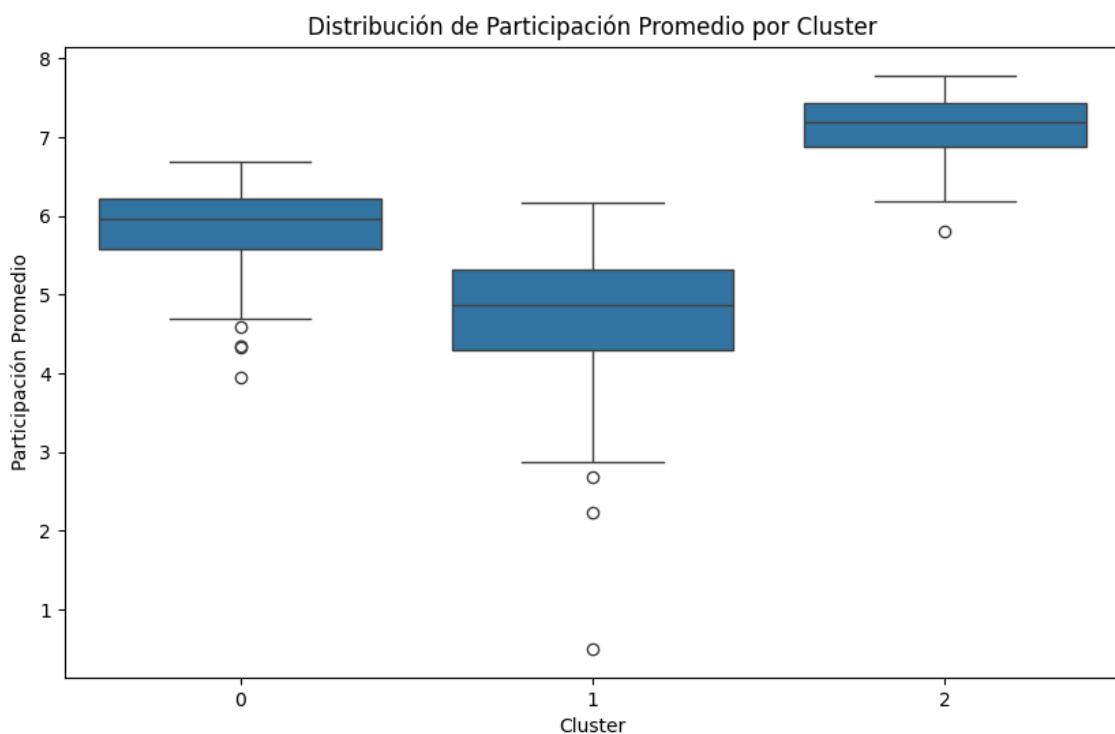
- Exámenes
  - Proyecto de Aplicación
  - Caso Práctico 3



**Figura 1. Participación en actividades y nota alcanzada**

### 3.4 Clustering de Trayectorias de Participación Académica

El análisis de agrupamiento (K-means) identificó tres grupos según patrones de participación (ver Figura 2). Las características de cada grupo se resumen a continuación:



**Figura 2. Distribución de Participación Promedio por Cluster**

**Tabla 2.***Segmentación de estudiantes por patrones de participación académica*

Cluster	Tamaño	Participación promedio	Dispersión intra-cluster
Cluster 0	150 estudiantes (38%)	Alta en todas las actividades.	Baja
Cluster 1	120 estudiantes (31%)	Moderada, baja en foros y exámenes.	Moderada
Cluster 2	123 estudiantes (31%)	Baja en todas las actividades.	Alta

El análisis ANOVA evidenció diferencias significativas en las calificaciones finales entre los clusters ( $p < 0.05$ ):

**Tabla 3.***Relación entre clusters y rendimiento académico*

Cluster	Nota Promedio	Interpretación
Cluster 0	4.5	Estudiantes más comprometidos, con un alto rendimiento académico, consistente con su alta participación.
Cluster 1	3.8	Grupo intermedio con una participación desigual, reflejado en una variabilidad moderada en el rendimiento académico.
Cluster 2	3.2	Estudiantes con menor participación y rendimiento, indicando barreras de compromiso o dificultades externas.

### 3.4 Análisis de Series Temporales (ARIMA)

Se implementó un modelo ARIMA ( $p = 1$ ,  $d = 1$ ,  $q = 0$ ) para evaluar la evolución de la participación a lo largo de 500 unidades de tiempo. Se validó su ajuste mediante los criterios AIC y BIC, observándose un buen desempeño predictivo.

Momentos clave identificados:

Los patrones observados en la Figura 3 pueden clasificarse en cuatro momentos clave:

- Crecimiento Inicial: se registra un aumento en la participación al inicio de la secuencia temporal, probablemente asociado a actividades iniciales del programa o a un mayor interés de los estudiantes al comenzar el curso.
- Fluctuaciones Medias: a lo largo del período, se observan variaciones recurrentes con picos en intervalos determinados, lo que sugiere que ciertos momentos o actividades generan mayor compromiso.
- Pico Máximo: entre las unidades 350 y 400, se detecta un incremento significativo en la participación, posiblemente relacionado con asignaturas específicas de alto impacto o mayor relevancia para los estudiantes.
- Descenso Final: hacia el final del período, la participación experimenta una disminución sostenida, lo que podría reflejar fatiga de los estudiantes o una pérdida progresiva de interés en las actividades académicas.

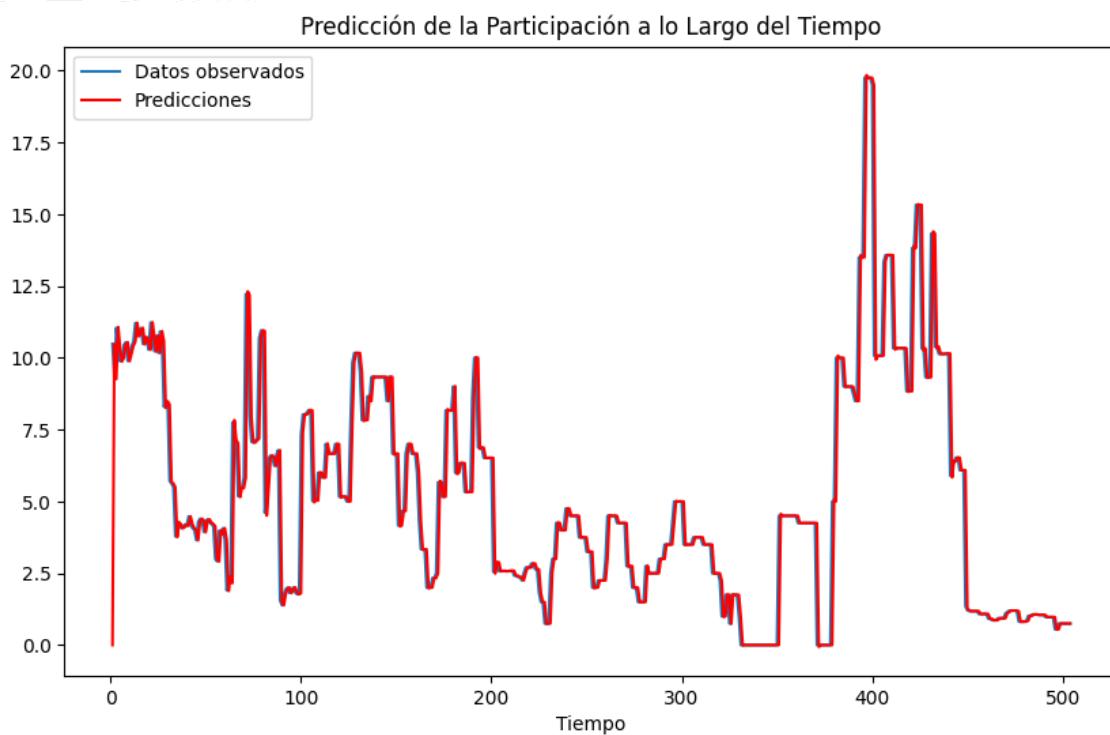


Figura 3. Predicción de la participación a lo largo del programa

#### 4. Discusión

El objetivo de esta investigación fue identificar patrones de participación estudiantil en entornos virtuales y analizar su relación con el rendimiento académico y la retención. A diferencia de otros trabajos, como el de De la Iglesia (2020), que emplearon diseños cuantitativos descriptivos y exploratorios, este estudio adoptó un enfoque cuantitativo, longitudinal y explicativo. Esta aproximación metodológica permitió observar la evolución de la participación a lo largo del tiempo, en línea con las recomendaciones de Miranda et al. (2021) y White et al. (2005), quienes destacan la relevancia de considerar las dinámicas temporales para diseñar intervenciones educativas basadas en evidencia.

En cuanto a las variables de participación que predicen el rendimiento y la retención, los resultados revelaron que las actividades prácticas—como los proyectos de aplicación y los casos prácticos—generaron mayores niveles de participación y se identificaron como predictores significativos del rendimiento académico. Estos hallazgos refuerzan la evidencia reportada por Díaz-García et al. (2023) y Mutanga (2024), quienes destacan la efectividad del aprendizaje basado en proyectos (ABP) para fomentar el compromiso y mejorar los resultados académicos.

Por el contrario, actividades como los foros evidenciaron niveles significativamente más bajos de participación, lo cual coincide con lo señalado por Donelan and Kear (2023), quienes argumentan que muchos estudiantes perciben estas actividades como poco relevantes o insuficientemente interactivas. No obstante, estudios como los de Candia et al. (2022) y Heryanto et al. (2024) advierten que, cuando están bien diseñados, los foros pueden beneficiar especialmente a estudiantes con bajo rendimiento previo, al facilitar el aprendizaje colaborativo, la inteligencia colectiva y la reflexión crítica. Estos resultados refuerzan la necesidad de rediseñar los foros para convertirlos en espacios más atractivos y con propósitos pedagógicos claros.

En relación con los patrones de participación digital y su agrupación, se encontró una asociación positiva entre la edad y la participación, lo cual concuerda con los hallazgos de Cruzado y Ortega (2022), quienes atribuyen esta relación a una mayor motivación intrínseca en los estudiantes adultos. Aunque se observaron diferencias marginales por género en actividades prácticas, no se identificaron brechas significativas en espacios colaborativos como los foros. Este hallazgo coincide con investigaciones previas (Canales & Rojas, 2020; González & López, 2024), que destacan el potencial inclusivo de la educación en línea al reducir las barreras tradicionales asociadas al género.

La aplicación del algoritmo Random Forest permitió identificar las actividades académicas con mayor capacidad predictiva del rendimiento estudiantil. Estos resultados se alinean con estudios como el de Brdnik et al. (2022), que reconocen el valor de estos modelos para construir sistemas de alerta temprana. En este caso, los exámenes y los proyectos de aplicación fueron los predictores más robustos del desempeño final, lo cual subraya la necesidad de priorizar estas actividades en el diseño curricular y en las estrategias de evaluación.

La implementación del análisis de clustering permitió segmentar a los estudiantes en tres grupos con trayectorias de participación diferenciadas: alta, moderada y baja. Esta segmentación evidenció una relación directa entre el nivel de compromiso y el rendimiento académico, confirmando lo reportado por Zhang et al. (2021), quienes argumentan que este tipo de análisis conductual contribuye a personalizar la enseñanza. De este modo, se evidencia cómo los modelos de segmentación pueden utilizarse para diseñar estrategias pedagógicas personalizadas e intervenciones orientadas a la retención estudiantil. Esta conclusión coincide con lo planteado por Sinchi et al. (2024), quien resalta que la personalización basada en datos puede mejorar significativamente tanto la retención como el éxito académico.

Asimismo, el análisis de series temporales (ARIMA) permitió observar patrones cíclicos en la participación a lo largo del ciclo académico, identificando momentos críticos de mayor y menor interacción estudiantil. Se observaron cuatro fases relevantes: crecimiento inicial, fluctuaciones medias, un pico de participación entre las unidades 350 y 400, y un descenso sostenido hacia el final del curso. Estos hallazgos ofrecen evidencia valiosa para anticipar períodos de desmotivación y planificar intervenciones proactivas, como sesiones de tutoría o actividades de cierre más motivadoras, con el fin de reducir la deserción.

Los resultados de este estudio ofrecen a las instituciones una base sólida para diseñar estrategias pedagógicas adaptativas y personalizadas, fundamentadas en patrones de participación estudiantil. Tal como destacan Tepgec et al. (2024), el valor de la analítica de aprendizaje no reside únicamente en la disponibilidad de datos, sino en su integración con estrategias pedagógicas que promuevan el uso reflexivo y autónomo de la información. En esta línea, investigaciones como las de Kaliisa et al. (2023) y Bergdahl et al. (2024) evidencian cómo los dashboards de analítica (LADs) favorecen el compromiso sostenido del estudiantado. Cabe destacar, además, que la articulación de estos recursos con metodologías activas —como el aprendizaje basado en proyectos, el aula invertida o la resolución de problemas— ha demostrado fortalecer la permanencia y el rendimiento académico (Digital Learning Edge, 2024; Martín-Caro, 2024), especialmente cuando se adaptan a los distintos perfiles de participación.

De este modo, se da respuesta a las preguntas planteadas al inicio del estudio, al identificar patrones de participación diferenciados, variables predictoras del rendimiento y del abandono, y oportunidades concretas para personalizar la enseñanza mediante el uso de la analítica de aprendizaje. Estos hallazgos ofrecen a las instituciones herramientas concretas para avanzar hacia una educación virtual más inclusiva, efectiva y centrada en el estudiante.

## 5. Conclusiones

Este estudio confirma que la analítica de aprendizaje (AA) representa una herramienta estratégica para mejorar la participación estudiantil y fortalecer la toma de decisiones en entornos virtuales. Su valor no se limita a describir patrones de comportamiento, sino que permite anticipar riesgos, identificar perfiles estudiantiles diferenciados y diseñar estrategias pedagógicas fundamentadas en evidencia.

Uno de los principales aportes prácticos es la validación del impacto positivo de las actividades prácticas —como los proyectos de aplicación— en la participación y el rendimiento académico. A partir de esta evidencia, se recomienda que las instituciones prioricen metodologías activas que integren la aplicación contextualizada del conocimiento. Asimismo, se sugiere rediseñar actividades con menor impacto participativo, como los foros o los exámenes, incorporando principios de retroalimentación formativa, co-creación de contenidos y mayor interactividad.

La implementación de modelos predictivos y técnicas de clustering proporciona información clave para el diseño de políticas de acompañamiento académico diferenciadas. La segmentación estudiantil permite avanzar hacia modelos de intervención personalizados: tutorías intensivas y recursos de apoyo para estudiantes con baja participación, y propuestas formativas más exigentes para aquellos con alta implicación. Esta perspectiva adaptativa puede contribuir a mejorar la equidad, la permanencia y la calidad del aprendizaje en programas de educación virtual.

Desde una perspectiva institucional, los hallazgos ofrecen criterios para optimizar los calendarios académicos y planificar intervenciones en función de los ciclos de participación estudiantil. El uso de series

temporales para identificar momentos críticos facilita la implementación de acciones preventivas más oportunas, lo cual fortalece tanto la experiencia del estudiante como la sostenibilidad y reputación de las instituciones.

En términos de política educativa, se recomienda integrar la AA como componente estructural en los sistemas de aseguramiento de la calidad en educación superior virtual. La incorporación de estas herramientas en la gestión institucional puede mejorar el seguimiento del progreso estudiantil, favorecer el uso eficiente de los recursos y promover procesos de mejora continua orientados al bienestar del estudiante.

Entre las principales limitaciones del estudio se encuentra el enfoque en una sola institución durante un periodo específico, lo que restringe la generalización. Tampoco se consideraron variables afectivas o contextuales, que pueden influir en la participación estudiantil. Futuros estudios podrían ampliar este enfoque explorando dimensiones emocionales, motivacionales y metacognitivas del compromiso estudiantil, así como sus interacciones con factores institucionales, socioculturales y tecnológicos. Esta línea de investigación contribuirá a construir una visión más holística, inclusiva y humana de la educación digital, alineada con los desafíos actuales de sostenibilidad e innovación en el ámbito universitario.

### Contribución de las autoras:

**Ester Martín-Caro:** Conceptualización y Metodología; Escritura del borrador original; Análisis de datos; Escritura (revisión y edición). **Nuria Segovia-García:** Recopilación y gestión de datos; Análisis de datos.

### Referencias

- Anthology. (2024). *Leveraging actionable data to increase student retention* [Sitio web]. Recuperado de <https://www.anthology.com/story/leveraging-actionable-data-to-increase-student-retention>
- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue. <https://doi.org/10.1145/2330601.2330666>
- ASEE Advances. (2020). *Active student engagement in online STEM classes: Approaches and recommendations. Advances in Engineering Education*. <https://advances.asee.org/active-student-engagement-in-online-stem-classes-approaches-and-recommendations/>
- Astin, A. W. (1984). Student Involvement: A Developmental Theory for Higher Education. *Journal of College Student Personnel*, 25, 297-308.
- Bean, J. P., & Eaton, S. B. (2000). A psychological model of college student retention. En J. M. Braxton (Ed.), *Reworking the departure puzzle: New theory and research on college student retention*. Vanderbilt University Press.
- Bergdahl, N., Bond, M., Sjöberg, J., Dougherty, M., & Oxley, E. (2024). Unpacking student engagement in higher education learning analytics: a systematic review. *International Journal Of Educational Technology In Higher Education*, 21(1). <https://doi.org/10.1186/s41239-024-00493-y>
- Brdnik, S., Šumak, B., & Podgorelec, V. (2022). Aligning Learners' Expectations and Performance by Learning Analytics System with a Predictive Model [Preprint]. arXiv. <https://arxiv.org/abs/2211.07729>
- Canales, R., & Rojas, L. (2020). Igualdad de género y rendimiento académico en entornos virtuales de educación superior. *Calidad en la Educación*, 40(5), 137-156. [https://www.scielo.cl/scielo.php?pid=S0718-5006202000500137&script=sci\\_arttext](https://www.scielo.cl/scielo.php?pid=S0718-5006202000500137&script=sci_arttext)
- Candia, C., Maldonado-Trapp, A., Lobos, K., Peña, F., & Bruna, C. (2022). Disadvantaged students increase their academic performance through collective intelligence exposure in emergency remote learning due to COVID-19 [Preprint]. arXiv. <https://arxiv.org/abs/2203.05621>
- Carmean, C., Kil, D., & Baer, L. (2021, agosto 10). Why data matters for student success in a post-pandemic world. Educause Review. <https://er.educause.edu/articles/2021/8/why-data-matters-for-student-success-in-a-post-pandemic-world>
- Cruzado, C., & Ortega, A. (2022). Relación entre motivación intrínseca y rendimiento académico en estudiantes adultos del Instituto de Educación Superior Pedagógico Público "Huancavelica". *Revista Iberoamericana de Educación Superior*, 18(2), 45-60. <https://doi.org/10.54943/rq.v%25vi%25i.75>
- De la Iglesia Villasol, M. C. (2020). Analítica de los usos digitales y rendimiento académico. Un estudio de caso con estudiantes universitarios. *REIRE Revista D Innovació i Recerca En Educació*, 13(2). <https://doi.org/10.1344/reire2020.13.229267>
- Digital Learning Edge. (2024, abril 15). *Enhancing online STEM course design for effective learning*. Recuperado de <https://digitallearningedge.com/about/>
- Donelan, H., & Kear, K. (2023). Online group projects in higher education: persistent challenges and implications for practice. *Journal Of Computing In Higher Education*, 36(2), 435-468. <https://doi.org/10.1007/s12528-023-09360-7>
- Downes, S. (2012). *Connectivism and connective knowledge. Essays on meaning and learning networks*. Recuperado de <http://www.downes.ca/me/mybooks.htm>

- Goh, T. (2025). Learning management system log analytics: the role of persistence and consistency of engagement behaviour on academic success. *Journal Of Computers In Education*. <https://doi.org/10.1007/s40692-025-00358-x>
- Goicochea, E. S., & Gómez, N. I. (2021). Competencias docentes en entornos virtuales de aprendizaje. *UCV - Scientia*, 13(1). <https://doi.org/10.18050/RevUcv-Scientia.v13n1a4>
- González, P., & López, A. (2024). Gender participation patterns in online learning environments [Preprint]. arXiv. <https://arxiv.org/abs/2406.11864>
- Guzmán, A., Barragán, S., & Cala Vitery, F. (2021). Ruralidad y deserción en programas de educación superior virtual en Colombia. *Sustainability*, 13(9), 4953. <https://doi.org/10.3390/su13094953>
- Halkiopoulos, C., & Gkintoni, E. (2024). Leveraging AI in E-Learning: Personalized Learning and Adaptive Assessment through Cognitive Neuropsychology—A Systematic Analysis. *Electronics*, 13(18), 3762. <https://doi.org/10.3390/electronics13183762>
- Heryanto, N., Firmansyah, F. H., & Rosmansyah, Y. (2024). Exploring Collaboration in Multiplayer Gamification: A Systematic Literature Review. *IEEE Access*, 1. <https://doi.org/10.1109/access.2024.3477465>
- Horna Li, L. E., & Seminario, R. J. (2023). Rendimiento académico en el entorno virtual de aprendizaje: una revisión sistemática. *Revista Conrado*, 19(91), 171–178. <https://conrado.ucf.edu.cu/index.php/conrado/article/view/2938>
- Horna, L. E. (2022). Plataforma virtual y rendimiento académico de los estudiantes durante la pandemia: una revisión sistemática. *UCV - Scientia*, 14(1). <https://doi.org/10.18050/RevUcv-Scientia.v14n1a6>
- IBM. (2022). *Data reliability*. IBM. Recuperado de <https://www.ibm.com/topics/data-reliability>
- Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4), 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>
- Kaensar, C., & Wongnin, W. (2023). Predicting new student performances and identifying important attributes of admission data using machine learning techniques with hyperparameter tuning. *Eurasia Journal Of Mathematics Science And Technology Education*, 19(12), em2369. <https://doi.org/10.29333/ejmste/13863>
- Kaliisa, R., Misiejuk, K., López-Pernas, S., Khalil, M., & Saqr, M. (2023). Have Learning Analytics Dashboards Lived Up to the Hype? A Systematic Review of Impact on Students' Achievement, Motivation, Participation and Attitude [Preprint]. arXiv. <https://arxiv.org/abs/2312.15042>
- Li, C., Herbert, N., Yeom, S., & Montgomery, J. (2022). Retention factors in STEM education identified using learning analytics: A systematic review. *Education Sciences*, 12(11), 781. <https://doi.org/10.3390/educsci12110781>
- Macfadyen, L. P., & Dawson, S. (2009). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2), 588–599. <https://doi.org/10.1016/j.compedu.2009.09.008>
- Martín-Caro, E. (2024). Análisis de estrategias innovadoras para retención estudiantil con inteligencia artificial: una perspectiva multidisciplinaria. *European Public & Social Innovation Review*, 9, 1–20. <https://doi.org/10.31637/epsir-2024-440>
- McKinsey & Company. (2022). *Using machine learning to improve student success in higher education*. Recuperado de <https://www.mckinsey.com/industries/education/our-insights/using-machine-learning-to-improve-student-success-in-higher-education>
- Miranda, H., Miranda-Zapata, E., Lara, L., & Saracostti, M. (2021). Análisis Longitudinal del Rendimiento Escolar según el Compromiso Escolar y Factores Contextuales: El Caso Chileno. *Revista Iberoamericana de Diagnóstico y Evaluación - E Avaliação Psicológica*, 163. <https://doi.org/10.21865/ridep59.2.13>
- Mutanga, M. B. (2024). Students' Perspectives and Experiences in Project-Based Learning: A Qualitative Study. *Trends In Higher Education*, 3(4), 903–911. <https://doi.org/10.3390/higheredu3040052>
- Nieuwoudt, J. E., & Pedler, M. L. (2021). Student retention in higher education: Why students choose to remain at university. *Journal of Student Affairs Research and Practice*, 25(2). <https://doi.org/10.1177/1521025120985228>
- Nimy, E., & Mosia, M. (2024). Modelling student retention in tutorial classes with uncertainty—A Bayesian approach to predicting attendance-based retention. *Education Sciences*, 14(8), 830. <https://doi.org/10.3390/educsci14080830>
- Oudat, M. S., Cheung, S. K. S., & Abu-Qdais, M. (2023). Enhancing sustainable learning in online education: The role of motivation, engagement, self-regulation, and perception. *Sustainability*, 16(2), 689. <https://doi.org/10.3390/su16020689>
- Pulker, H. (2019). Learning analytics to improve retention: Some critical questions. *Distances Et Médiations Des Savoirs*, 28(28). <https://doi.org/10.4000/dms.4602>
- Quadri, A. T., & Shukor, N. A. (2021). The benefits of learning analytics to higher education institutions: A scoping review. *International Journal of Emerging Technologies in Learning (iJET)*, 16(23), 4–15. <https://doi.org/10.3991/ijet.v16i23.27471>
- Sánchez-Gelabert, A. y Elías, M. (2017). Los estudiantes universitarios no tradicionales y el abandono de los estudios. *Estudios sobre Educación*, 32, 27–48. <https://doi.org/10.15581/004.32.27-48>
- Segovia-García, N., & Said-Hung, E. M. (2021). Factores de satisfacción de los alumnos en e-learning en Colombia. *Revista Mexicana de Investigación Educativa*, 26(89), 595–621. Recuperado de <https://cutt.ly/LbUPghi>
- Segovia-García, N., & Segovia-García, L. (2024). Chatbot de IA para prevenir el abandono de la educación superior: una revisión de la literatura. *Education In The Knowledge Society (EKS)*, 25, e31416. <https://doi.org/10.14201/eks.31416>

- Segovia-García, N., Said-Hung, E., & Aguilera, F. J. G. (2022). Educación superior virtual en Colombia: factores asociados al abandono. *Educación XXI*, 25(1), 197-218. <https://doi.org/10.5944/educxx1.30455>
- Segovia-García, N., Viñambres, D. O., & Álamo, E. M. (2020). La evaluación del aprendizaje social como herramienta de predicción del rendimiento académico y la permanencia en programas de Educación Superior virtual. En *Claves para la innovación pedagógica ante los nuevos retos: Respuestas en la vanguardia de la práctica educativa* (pp. 3798-3806). Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=7758458>
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30-32.
- Sinchi Pacurucu, V. Y., Morillo Revelo, W. P., López Velasco, J. E., Maldonado Palacios, I. A., & Vizcaíno Zúñiga, P. I. (2024). Evaluación de impacto de plataformas de aprendizaje en línea en el rendimiento académico: Assessment of the impact of online learning platforms on academic performance. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades*, 5(2), 727 – 739. <https://doi.org/10.56712/latam.v5i2.1912>
- Tepgec, M., Heil, J., & Ifenthaler, D. (2024). Feedback literacy matters: unlocking the potential of learning analytics-based feedback. *Assessment & Evaluation In Higher Education*, 1-17. <https://doi.org/10.1080/02602938.2024.2367587>
- The Big Data Framework. (2024). *Understanding data quality*. Recuperado de <https://www.bigdataframework.org/knowledge/understanding-data-quality>
- Times Higher Education. (2024). How digital tools can enhance STEM teaching [Sitio web]. Recuperado de <https://www.timeshighereducation.com/hub/digitalized/p/how-digital-tools-can-enhance-stem-teaching>
- Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. University Of Chicago Press
- Uzun, Y., Suraworachet, W., Zhou, Q., Gauthier, A., & Cukurova, M. (2025). Engagement with analytics feedback and its relationship to self-regulated learning competence and course performance. *International Journal Of Educational Technology In Higher Education*, 22(1). <https://doi.org/10.1186/s41239-025-00515-3>
- Villegas-Ch, W., Govea, J., & Revelo-Tapia, S. (2023). Improving student retention in institutions of higher education through machine learning: A sustainable approach. *Sustainability*, 15(19), 14512. <https://doi.org/10.3390/su151914512>
- Wang, X., Hall, A., & Wang, Q. (2024). Online learning environment and student engagement: The role of feedback and interaction. *The Australian Educational Researcher*, 51(1), 1-17. <https://doi.org/10.1007/s13384-024-00689-1>
- White, R. T., & Arzi, H. J. (2005). Longitudinal studies: Designs, validity, practicality, and value. *Research in Science Education*, 35(2), 137-149. <https://doi.org/10.1007/s11165-004-3437-y>
- Wong, W. (2021, 12 octubre). Higher Education Turns to Data Analytics to Bolster Student Success. *Technology Solutions That Drive Education*. <https://edtechmagazine.com/higher/article/2021/10/higher-education-turns-data-analytics-bolster-student-success>
- Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, 126(6), 841-864. <https://doi.org/10.1037/rev0000153>
- Zhang, M., Du, X., Rice, K., Hung, J., & Li, H. (2021). Revealing at-risk learning patterns and corresponding self-regulated strategies via LSTM encoder and time-series clustering. *Information Discovery And Delivery*, 50(2), 206-216. <https://doi.org/10.1108/idd-12-2020-0160>