

Web analytics para desarrollar un entorno de *learning analytics* y su relación con el rendimiento académico en cursos virtuales

Web Analytics to develop a Learning Analytics environment and its relationship with academic performance in virtual courses

Web Analytics para desenvolver um ambiente de Learning Analytics e sua relação com o desempenho académico em cursos virtuais

Guillermo Segundo Miñan-Olivos

Universidad Tecnológica del Perú
Lima, Perú

c20342@utp.edu.pe

 ORCID: <https://orcid.org/0000-0001-9523-8043>

Christian Abraham Dios-Castillo

Universidad Tecnológica del Perú
Lima, Perú

cdios@utp.edu.pe

 ORCID: <https://orcid.org/0000-0002-2469-9237>

Manuel Antonio Cardoza-Sernaqué

Universidad Tecnológica del Perú
Lima, Perú

e20185@utp.edu.pe

 ORCID: <https://orcid.org/0000-0001-6738-0683>

Luis Alexander Pulido-Joo

Universidad Tecnológica del Perú
Lima, Perú

e20188@utp.edu.pe

 ORCID: <https://orcid.org/0000-0002-4919-3558>

Recibido – Received – Recebido: 30 / 06 / 2021 Corregido – Revised – Revisado: 30 / 09 / 2021 Aceptado – Accepted – Aprovado: 19 / 10 / 2021

DOI: <https://doi.org/10.22458/ie.v23iespecial.3615>

URL: <https://revistas.uned.ac.cr/index.php/innovaciones/article/view/3615>

Resumen: El objetivo de la presente investigación fue demostrar cómo a partir de la *web analytics* se puede extraer información relevante para desarrollar un entorno de *learning analytics* en cursos virtuales e implementar métricas correlacionadas al rendimiento académico. En ese sentido, la metodología presentó un diseño de investigación no experimental, con alcance descriptivo correlacional y una recolección de datos longitudinal. La población estuvo conformada por estudiantes de primer ciclo de un curso de investigación académica en modalidad asincrónica y la muestra estuvo representada por 59 personas estudiantes. La información recopilada comprendió datos generados por cada estudiante en tres aspectos: visitas a páginas dentro del curso, participaciones en actividades del curso (foros, tareas, cuestionarios, etc.) y las calificaciones. Los resultados mostraron el siguiente comportamiento promedio de los estudiantes: 3809 visitas por semana, 563 visitas diarias, 143 participaciones por semana y 21 participaciones diarias. Al relacionar las métricas web con el rendimiento académico (calificaciones) se obtuvieron los siguientes valores, correspondientes a la rho de Spearman: 0.628 ($p < 0.00$) entre visitas y calificaciones y 0.638 ($p < 0.00$) entre participaciones y calificaciones. Finalmente, se pudo concluir que las métricas basadas en la *web analytics* son una fuente primaria para iniciar el desarrollo de un entorno asociado a la *learning analytics*; asimismo, se mostró una relación directamente proporcional de las métricas web con el rendimiento académico.

Palabras clave: educación a distancia, aprendizaje en línea, evaluación del estudiante, comportamiento del alumno, analítica web, analítica de aprendizaje

Summary: The object of this research was to demonstrate how relevant information can be extracted from Web Analytics to develop a Learning Analytics environment in virtual courses and implement metrics correlated to academic performance. In this sense, the methodology presented a non-experimental research design, with a correlational descriptive scope and a longitudinal data collection. The population was made up of first-cycle students from an asynchronous academic research course and the sample was represented by 59 students. The information collected included data generated by each student in three aspects: visits to pages within the course, participation in course activities (forums, assignments, questionnaires, etc.) and grades. The results showed the following average behavior of the students: 3809 visits per week, 563 visits per day, 143 participations per week and 21 participations per day. When relating the web metrics to academic performance (grades), the following values were obtained, corresponding to Spearman's Rho: 0.628 ($p < 0.00$) between visits and grades and 0.638 ($p < 0.00$) between participations and grades. Finally, it could be concluded that metrics based on Web Analytics are a primary source to start the development of an environment associated with Learning Analytics. Likewise, a directly proportional relationship of web metrics with academic performance was shown.

Key Words: distance education, online learning, student assessment, student behavior, web analytics, learning analytics

Resumo: O objetivo da pesquisa foi demonstrar como desde a Web Analytics é possível extrair informação relevante para desenvolver um ambiente de Learning Analytics em cursos virtuais e implementar métricas correlacionadas ao desempenho acadêmico. Nesse sentido, a metodologia apresentou um desenho de investigação não experimental, com escopo descritivo correlacional e uma coleta de dados longitudinal. A população esteve integrada por estudantes de primeiro ciclo de um curso de pesquisa acadêmica na modalidade assíncrona e a amostra representada por 59 estudantes. As informações coletadas continham dados gerados por cada estudante em três aspectos: visitas às páginas dentro do curso, participação em atividades do curso (fóruns, tarefas, questionários etc.) e as notas. Os resultados mostraram o seguinte comportamento médio dos estudantes: 3809 visitas por semana, 563 visitas diárias, 143 participações por semana e 21 participações por dia. Ao relacionar as métricas da web com o desempenho acadêmico (notas) foram obtidos os seguintes valores, correspondentes ao Rho de Spearman: 0.628 ($p < 0.00$) entre visitas e notas e 0.638 ($p > 0.00$) entre participações e notas. Finalmente, é possível concluir que as métricas baseadas na Web Analytics são uma fonte primária para iniciar o desenvolvimento de um ambiente associado à Learning Analytics, além de uma relação diretamente proporcional das métricas da Web com desempenho acadêmico.

Palavras-chave: Educação à distância, aprendizagem on-line, avaliação dos estudantes, comportamento do aluno, analítica web, analítica de aprendizagem.

INTRODUCCIÓN

En el contexto de la pandemia de COVID-19 se han presentado distintas dificultades y detectado falencias en los entornos de aprendizaje virtual (Salinas, 2020). A pesar de ello, las propuestas en sistemas educativos basados en tecnología digital han crecido exponencialmente. Los docentes y los estudiantes pueden trabajar e interactuar desde cualquier lugar y en cualquier momento. Además, esta forma de educación basada en computadora es prácticamente independiente de cualquier plataforma de hardware específica. Para ello, se han propuesto muchas plataformas de software, gratuitas o privadas, con fines educativos en forma de sistemas de gestión del aprendizaje (LMS por sus siglas en inglés), donde los estudiantes pueden interactuar frecuentemente con tutores y compañeros. Por ejemplo, Moodle es un entorno de aprendizaje interactivo gratuito donde los estudiantes aprenden interactuando con sus compañeros, con tutores y con materiales didácticos. Muchas instituciones públicas utilizan estos LMS para capacitar a las personas en programas de aprendizaje permanente (Sciarrone, 2018). Cada uno de estos sistemas está diseñado para almacenar una gran cantidad de información que podría usarse para mejorar los procesos de aprendizaje, enseñanza y tareas administrativas. En consecuencia, extraer información útil de estos datos representa un nuevo desafío que involucra el aprendizaje automático, la minería de datos y el análisis del aprendizaje. Entre estas herramientas, el aprendizaje automático resalta ya que se ocupa de una gran cantidad de algoritmos que mejoran su rendimiento con la experiencia, abordando problemas de regresión o clasificación en muchos campos de investigación, incluidos los contextos de aprendizaje donde los estudiantes dejan patrones de comportamiento útiles. Sin embargo, el presente estudio realizó una revisión de la literatura respecto a dos ejes temáticos: *learning analytics* como herramienta de vanguardia en el campo educativo y *web analytics* como herramienta

primaria para alcanzar un entorno de *learning analytics* que permita describir el comportamiento del estudiantado en cursos virtuales.

La *learning analytics* o analítica de aprendizaje representa un proceso complejo que tiene como objetivo comprender y optimizar todo el proceso de aprendizaje, junto con el entorno en el que se produce (Stewart, 2017). Asimismo, se define como un campo de investigación en ciencia de datos emergente y de rápido crecimiento que se centra en el desarrollo de métodos para analizar y detectar patrones dentro de estos datos, y así respaldar la experiencia de aprendizaje. La comunidad de análisis de aprendizaje ha madurado significativamente en los últimos años. Para avanzar aún más, la investigación debe avanzar hacia la provisión de analítica de aprendizaje a escala. La ampliación de entornos de aprendizaje virtuales requiere una visión holística de la analítica de aprendizaje en conjuntos de datos distribuidos a través de una variedad de entornos y contextos diferentes, mediante la aplicación de enfoques de métodos mixtos para abordar una amplia gama de participantes con diversos intereses, necesidades y objetivos. Un aspecto central de esta discusión es el concepto de análisis de aprendizaje abierto, que representa un cambio hacia un nuevo modelo de análisis de aprendizaje que tiene en cuenta la apertura. Todavía falta en la comunidad una comprensión común de la apertura en relación con *learning analytics*, y la investigación sobre plataformas de analíticas de aprendizaje abiertas se encuentra todavía en las primeras etapas de desarrollo. Además, hasta ahora se ha realizado relativamente escasa investigación para proporcionar una visión general sistemática del campo e identificar los diversos desafíos que se avecinan, así como las direcciones de investigación futuras (Pinkwart & Liu, 2020).

Learning analytics es un campo en desarrollo en el sector educativo: se utiliza para inspeccionar la información pertinente sobre los estudiantes, los docentes y sobre los cursos; es decir, se busca la comprensión del desempeño del alumno y de esa manera promover el rendimiento del estudiante con las técnicas analíticas de aprendizaje y otras herramientas de análisis avanzadas. El rendimiento del alumno y los resultados del estudio pueden mejorarse a partir de un enfoque de ayuda o tutoría, promoviendo de esta manera el aprendizaje y la educación. La investigación basada en *learning analytics* incluye el crecimiento, la utilización y la combinación de nuevos procedimientos y herramientas para mejorar el acto de aprender y educar para estudiantes y educadores (Ranjeeth, Latchoumi & Paul, 2020).

La analítica del aprendizaje se nutre de una serie de otros campos de estudio, como la *business intelligence*, la *web analytics*, la analítica académica, la minería de datos educativos y la analítica de la acción, y está estrechamente relacionada con cada uno de ellos. La analítica del aprendizaje puede ilustrarse como una aplicación informativa de la *web analytics* –analítica web o analítica de red– destinada a analizar a los alumnos mediante procedimientos de recogida de datos y del análisis de los mismos, a través del detalle de las distintas interacciones de los estudiantes en las acciones de aprendizaje virtual. La *learning analytics* consiste en herramientas analíticas utilizadas para desarrollar el aprendizaje y la educación (Bharara, Sabitha & Bansal, 2018). En ese sentido, las herramientas de *web analytics* pueden diseñarse con éxito, completamente integradas, en una plataforma digital y, por tanto, utilizarse con fines de *web analytics*. Una gran parte de las métricas relevantes puede recuperarse utilizando las capacidades de *web analytics*.

A pesar de los beneficios de la *web analytics*, existen algunas limitaciones, por lo que este método no puede utilizarse exclusivamente en este contexto. Entonces, debe establecerse con precisión el alcance que puedan tener las herramientas de *web analytics* para obtener información sobre el comportamiento de aprendizaje en las plataformas de *e-learning*. La *learning analytics* se apoya en la visualización de la información (Nguyen, Gardner & Sheridan, 2020); es por ello que un gran número de métricas de *learning analytics* corresponden o pueden ser mapeadas con métricas de *web analytics*. Por ejemplo, la duración de la sesión arroja datos sobre el tiempo en que los usuarios están aprendiendo de una sola vez, y las páginas vistas indican con qué frecuencia y en qué orden se visitan los elementos de aprendizaje. Además, el número de usuarios activos y las características de la audiencia, como los patrones de acceso temporal, los clientes y dispositivos utilizados y los orígenes geográficos, son relevantes para ambos

campos. Gracias al seguimiento genérico de eventos, que soportan la mayoría de las herramientas de *web analytics*, se puede rastrear cualquier tipo de interacción y, por tanto, analizarla. Como resultado, los indicadores clave, especialmente de *learning analytics*, pueden ser calculados fácilmente usando *web analytics*, ya que normalmente solo cuentan las ocurrencias de un determinado tipo de evento en un rango de fechas determinado. Al respecto, algunos antecedentes demuestran el interés de muchas entidades educativas por mejorar sus procesos educativos a través de la adquisición de herramientas digitales o soluciones informáticas que permitan recopilar información para mejorar la experiencia del estudiantado con cursos no presenciales desde un punto de vista pedagógico y de servicio o satisfacción; asimismo, otros trabajos previos han demostrado cómo datos digitales permiten describir comportamientos estudiantiles a través de métricas como la duración en minutos por cada visita, lecturas en línea, enlace a fuentes externas, entre otras (Chan, Botelho & Lam, 2019; Pérez, García y López, 2013).

En ese sentido, el presente estudio se justifica al contrastar ambas herramientas, pero al mismo tiempo explicar cómo funcionan de manera colaborativa. Por un lado, la *learning analytics* se centra en optimizar la experiencia de aprendizaje en las plataformas *online*. Esto incluye también la mejora de la experiencia del usuario, como hace la *web analytics*. Sin embargo, la intención difiere entre ambos campos. Mientras que la *web analytics* tiene como objetivo ayudar a las empresas en los procesos de toma de decisiones y pretende aumentar los ingresos, los usuarios están en el centro de la *learning analytics*, ya que deben ser apoyados mientras aprenden. Esto también incluye animar a los alumnos individualmente, por ejemplo, identificando a los usuarios con mayor probabilidad de abandonar pronto un curso y que, en concordancia con ello, podrían necesitar una ayuda especial. Este es un caso donde no solo se prioriza la gestión del aprendizaje sino también la administrativa y que, a su vez, demuestra el análisis del comportamiento individual al que se puede llegar. Es decir, no solo se puede acceder a las métricas relativas a los segmentos de usuarios (por ejemplo, usuarios móviles o usuarios de un determinado país) sino a distintos indicadores descriptivos (Rohloff, Oldag, Renz & Meinel, 2019).

Considerando lo expuesto, se planteó el siguiente problema de investigación: ¿De qué manera la *web analytics* puede utilizarse para desarrollar un entorno de *learning analytics* e identificar variables que se relacionen al rendimiento académico en cursos virtuales? Y en concordancia con el problema planteado, se estableció el siguiente objetivo general del estudio: Aplicar la *web analytics* para desarrollar un entorno de *learning analytics* e identificar variables que se relacionen al rendimiento académico en cursos virtuales. De la misma manera, se determinaron tres objetivos específicos: diseñar métricas para el seguimiento del estudiante utilizando *web analytics* a partir de la información registrada en el portal educativo de la universidad, implementar las métricas diseñadas y describir la interacción del estudiantado con el curso virtual, generar un entorno de *learning analytics* a través de la correlación de las métricas de *web analytics* y el rendimiento académico del curso virtual.

MATERIALES Y MÉTODOS

Hernández, Fernández y Baptista (2014) mencionan que las investigaciones pueden ser del tipo experimental o no experimental, mientras el alcance de las mismas puede llegar a niveles exploratorios, descriptivos, correlacionales o explicativos. Para el caso de la presente investigación, se estableció un estudio con enfoque cuantitativo y del tipo no experimental, ya que no se procedió a manipular ninguna de las variables de estudio. Se utilizó un diseño transversal y se logró tener un alcance tanto descriptivo como correlacional. La población del estudio estuvo conformada por estudiantes del primer ciclo que desarrollaban un curso de investigación en modalidad asincrónica. Como muestra se seleccionaron 59 estudiantes, para lo cual se utilizó un muestreo no probabilístico por conveniencia. Tal como se muestra en la figura 1, el procedimiento se inició con la descarga de los datos generados por el sistema de gestión del aprendizaje (LMS) denominado Canvas®.

Figura 1. Procedimiento para la obtención e importación de datos generados

a) Datos obtenidos en formato CSV de Canvas

Identificaci3n del estudiante	Nombre del estudiante	Nombre clasificable	Identificaci3n de SIS del estudiante	ID de la secci3n	Nombre de secci3n	ID del curso	Nombre del curso	Tipo de contenido	Nombre del contenido	Veces visto	Veces que ha participado	Fecha de inicio	Primeros 30 d3as	
9483E+16	MAICOL JUNIOR FARRO MATIENZO	FARRO MATIENZO, MAICOL JUNIOR	9483000000156789	INVESTIGACION ACADEMICA	(2026)	9483000000184025	INVESTIGACION ACADEMICA	(2026)	course.quiztes.quiz	Semana 10 - Tema 1: Autoevaluaci3n - Esquema de redacci3n	1	0	26/05/2021	26
9483E+16	MAICOL JUNIOR FARRO MATIENZO	FARRO MATIENZO, MAICOL JUNIOR	9483000000156789	INVESTIGACION ACADEMICA	(2026)	9483000000184025	INVESTIGACION ACADEMICA	(2026)	course.files.file	support-1.jpg	2	0	27/05/2021	27

b) Importaci3n de datos CSV mediante Microsoft Excel 365

The screenshot shows the 'Importar datos de texto/CSV' (Import text/CSV) dialog box in Microsoft Excel 365. The dialog is open to the 'Desde el texto/CSV' (From text/CSV) option. The 'Importe datos de archivos de texto, de valores delimitados por comas o de texto con formato (delimitado por espacios).' (Import data from text files, comma-separated values, or formatted text (delimited by spaces).) option is selected. The 'Hoja1' (Sheet1) is selected as the destination.

c) Generaci3n de tablas din3micas a partir de datos obtenidos en formato CSV

Identificaci3n d	Nombre del estudiante	Tipo de contenido	Nombre del contenido	Veces visto	Veces que ha participado	Fecha de inicio	Pr
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.quiztes.quiz	Semana 10 - Tema 1: Autoevaluaci3n - Esquema	1	0	26/05/2021	26
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.discussions.discussion	Ejemplo esquema de redacci3n	1	0	27/05/2021	27
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	2	0	29/05/2021	29
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	5	0	3/06/2021	3
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	2	0	28/05/2021	28
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	4	0	5/06/2021	5
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	4	0	4/06/2021	4
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	8	0	2/06/2021	2
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	2	0	26/05/2021	26
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	support-1.jpg	1	0	27/05/2021	27
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	2	0	29/05/2021	29
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	5	0	3/06/2021	3
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	2	0	28/05/2021	28
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	4	0	5/06/2021	5
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	4	0	4/06/2021	4
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	8	0	2/06/2021	2
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	2	0	26/05/2021	26
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	zoom-1.jpg	1	0	27/05/2021	27
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.assignments.assignment	(AC-311) Semana 11: Asesor3a 4 - Esquema de r	2	1	3/06/2021	3
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	book-1.jpg	2	0	29/05/2021	29
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	book-1.jpg	5	0	3/06/2021	3
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	book-1.jpg	2	0	28/05/2021	28
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	book-1.jpg	4	0	5/06/2021	5
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	book-1.jpg	4	0	4/06/2021	4
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	book-1.jpg	8	0	2/06/2021	2
9483E+16	MAICOL JUNIOR FARRO MATIENZO	course.files.file	book-1.jpg	2	0	26/05/2021	26

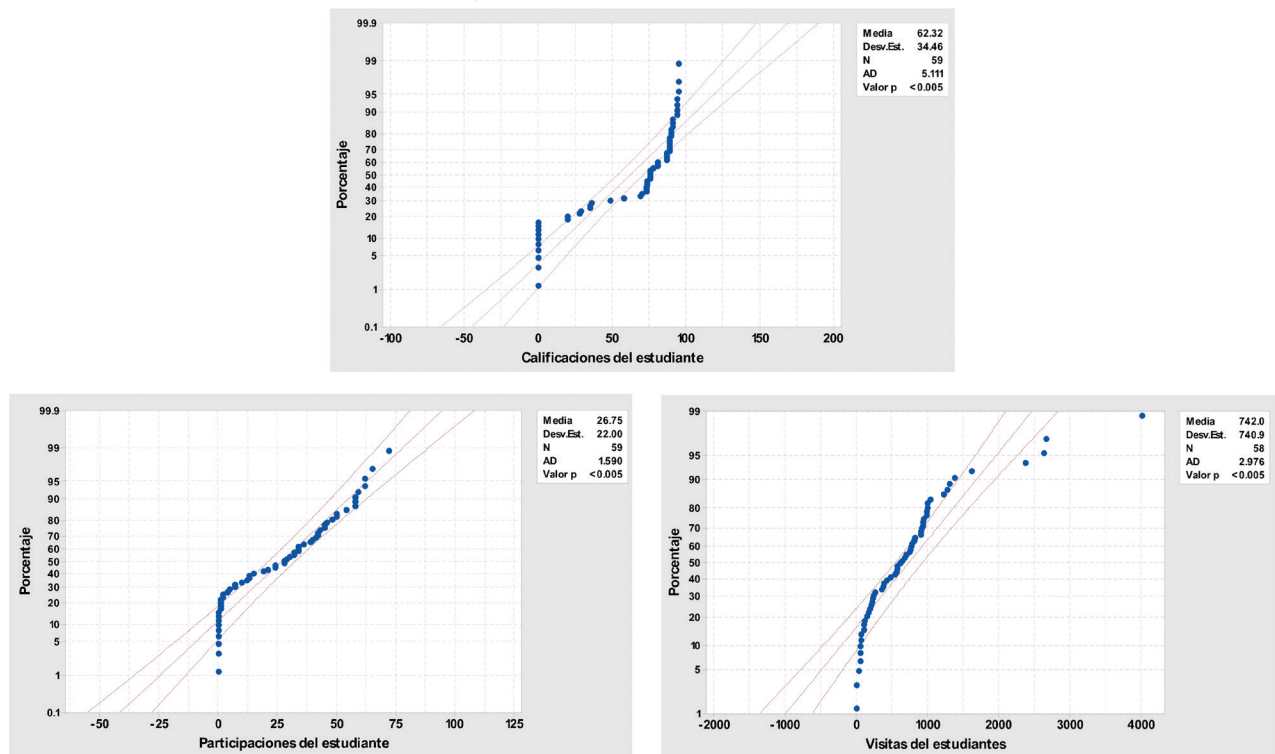
Fuente: Elaboraci3n propia.



Posteriormente, para cumplir con el primer objetivo específico del estudio, se importaron los datos mediante Microsoft Excel 365 (Bharara, Sabitha & Bansal, 2018) y se les asignó un formato de tabla dinámica para una correcta gestión de los insumos. La información recopilada comprendió datos generados por cada estudiante en tres aspectos: visitas a páginas dentro del curso en Canvas, participaciones en actividades del curso mediante el portal web y las calificaciones. Los datos que se descargaron contenían información recopilada por la plataforma web desde el 14 de marzo del 2021 hasta el 9 de junio del 2021, agrupados en 13 semanas. En total, se logró recopilar 2056 filas que cuantificaron 43 038 visitas a las distintas páginas del curso y 1578 participaciones de los estudiantes en actividades del curso tales como: autoevaluaciones, foros, ejercicios, tareas, entre otras. Ya que los datos se procesaron en Microsoft Excel 365, también se hizo uso de las herramientas disponibles en dicha hoja de cálculo para la generación de gráficos de áreas, histogramas y gráficas de dispersión, los cuales a su vez permitieron cumplir con el objetivo específico de implementar las métricas diseñadas y describir la interacción del estudiantado con el curso virtual.

El último objetivo específico planteaba generar un entorno de *learning analytics* a través de la correlación de las métricas de *web analytics* y el rendimiento académico del curso virtual. Es por ello por lo que el análisis inferencial de la investigación se basó en demostrar la relación entre las variables cuantificadas: visitas, participaciones y calificaciones del estudiante. Sin embargo, antes de determinar la prueba estadística a utilizar, se evaluaron los datos de cada variable para constatar su normalidad y de esa manera establecer si el estudio aplicaba una prueba inferencial paramétrica o no paramétrica. En la figura 2 se puede ver que el estadístico de Anderson Darling, utilizado para el análisis de normalidad de distribuciones muestrales (Tobón y Bedoya, 2017), tuvo un nivel de significancia menor a 0.05, por lo cual se estableció que las variables analizadas no seguían una distribución normal.

Figura 2. Prueba de normalidad para los datos obtenidos respecto a participaciones, visitas y calificaciones de los estudiantes



Fuente: Elaboración propia.

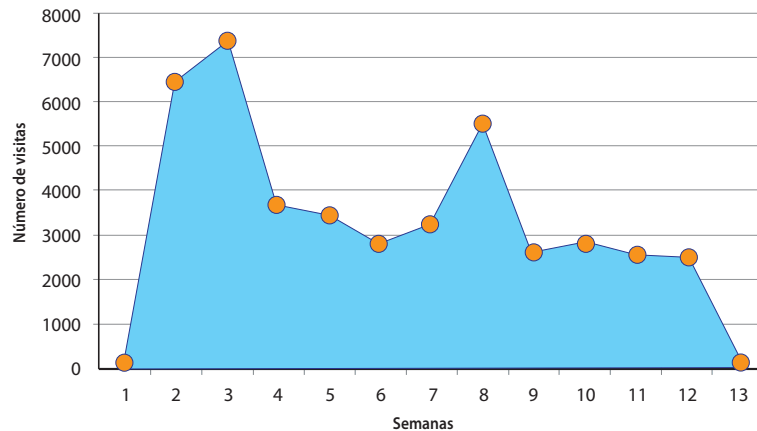
En concordancia con los resultados de la prueba de normalidad, la metodología de la investigación estableció el uso de la rho de Spearman para medir la correlación de las variables, ya que dicha opción representa una alternativa no paramétrica ante el incumplimiento de los supuestos para la aplicación de pruebas paramétricas tales como Pearson (Hernández et al., 2018). Para el coeficiente de Spearman se consideraron valores entre 0 y 1, con un nivel de significancia de 0.05 (Restrepo y González, 2007).

DISCUSIÓN DE LOS RESULTADOS

Los resultados de la investigación lograron sintetizar la información recopilada, a través de gráficos estadísticos que permitieron describir la interacción de los estudiantes con el material disponible para el curso. Dicha interacción se cuantificó a través de las visitas y de las participaciones.

La figura 3 muestra el comportamiento de las visitas que se han registrado en la página del curso. Durante la semana 3 se registró el valor más alto del periodo analizado, equivalente a 7276 visitas. Durante la semana 8 se tuvo un repunte correspondiente a 5485 visitas, para luego nuevamente descender. En el caso de la semana 13 (152 visitas) se puede ver un descenso justificado porque la semana de estudio se estaba iniciando, mientras que en la semana 1 (92 visitas) se debe tener en cuenta que es la primera semana de estudio. En ese sentido, ambos datos se pueden considerar como valores atípicos, por lo cual no se incluyeron al momento de estimar la media semanal y diaria. Al respecto, el promedio semanal fue estimado en 3809 visitas, en tanto que el valor medio diario se calculó en 563 visitas. Asimismo, se pudo establecer que cada estudiante realizaba en promedio 10 visitas diarias.

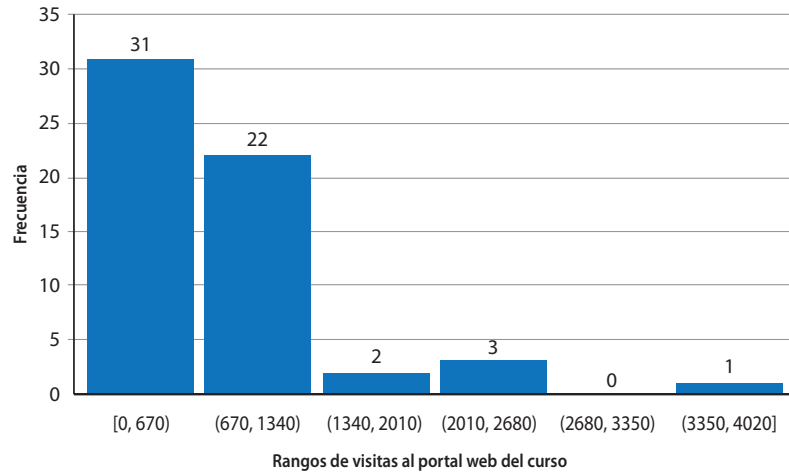
Figura 3. Visitas registradas a la página del curso durante el periodo comprendido desde el 14 de marzo hasta el 6 de junio del 2021



Fuente: Elaboración propia.

La figura 4 permite notar que 31 estudiantes, es decir el 52.54%, realizaron visitas a la página web del curso (foros, prácticas, material de estudio, etc.) entre 0 y 670 veces. De la misma manera, se pudo identificar que el 37.29% (22 estudiantes) frecuentó la página del curso entre 671 y 1340 veces. Solo el 10.17%, equivalente a 6 estudiantes, superó las 1340 visitas durante el periodo analizado. En promedio, un estudiante ha visitado la página del curso unas 729 veces. En el caso de estudiantes aprobados, el valor más bajo de visitas fue de 57, mientras que el más alto se ubicó en 4010.

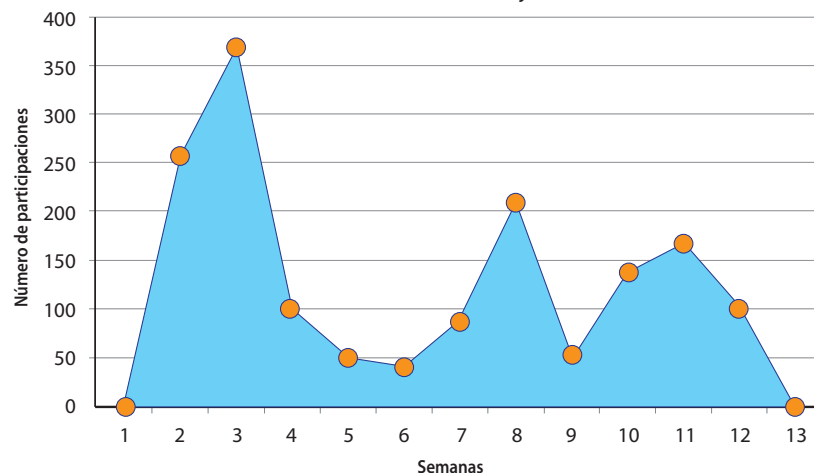
Figura 4. Histograma de frecuencia de visitas registradas por estudiante a la página del curso durante el periodo comprendido desde el 14 de marzo hasta el 6 de junio del 2021



Fuente: Elaboración propia.

La figura 5 muestra el comportamiento de las participaciones contabilizadas en la página del curso. Durante la semana 3 se registró el valor más alto del periodo analizado, equivalente a 368 participaciones. Durante las semanas 8 y 11 se puede apreciar un repunte correspondiente a 209 y 169 participaciones, respectivamente. En el caso de la semana 13, se puede ver un descenso justificado porque la semana de estudio se estaba iniciando, mientras que en la semana 1 se debe tener en cuenta que es la primera semana de estudio. En ese sentido, ambos datos se pueden considerar como valores atípicos, por lo cual no se incluyeron al momento de estimar la media semanal y diaria. Al respecto, el promedio semanal fue estimado en 143 participaciones; por su parte, el valor medio diario se calculó en 21 participaciones.

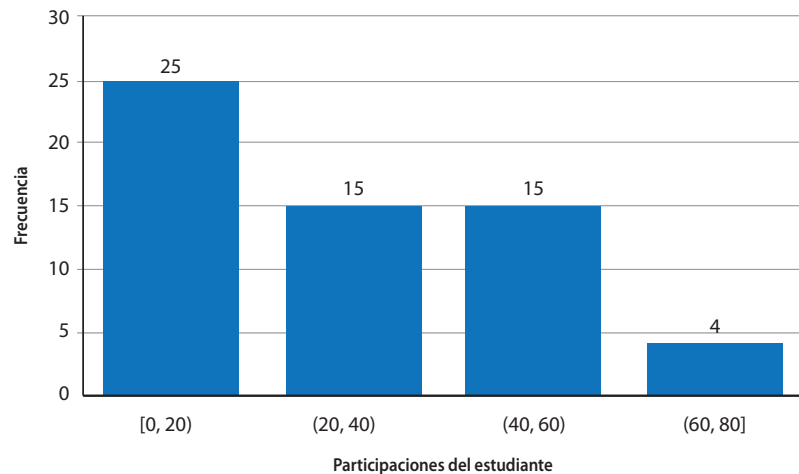
Figura 5. Participaciones contabilizadas en la página del curso durante el periodo comprendido desde el 14 de marzo hasta el 6 de junio del 2021



Fuente: Elaboración propia.

La figura 6 permite notar que 25 estudiantes, es decir el 42.37%, participaron en actividades disponibles en la página web del curso (foros, prácticas, material de estudio), cumpliendo entre 0 y 20 de las mismas. De la misma manera, se pudo identificar que el 25.42% (15 estudiantes) cumplió con su participación de entre 20 y 40 actividades programadas. El mismo porcentaje tuvo un nivel de participación que fluctuó entre 40 y 60 actividades del curso. Solo el 6.79%, equivalente a 4 estudiantes, superó la participación en 60 actividades programadas durante el periodo analizado. En promedio, un estudiante participó en 27 actividades planteadas por el desarrollo del curso. En el caso de estudiantes aprobados, el valor más alto de participación ha sido de 72 actividades.

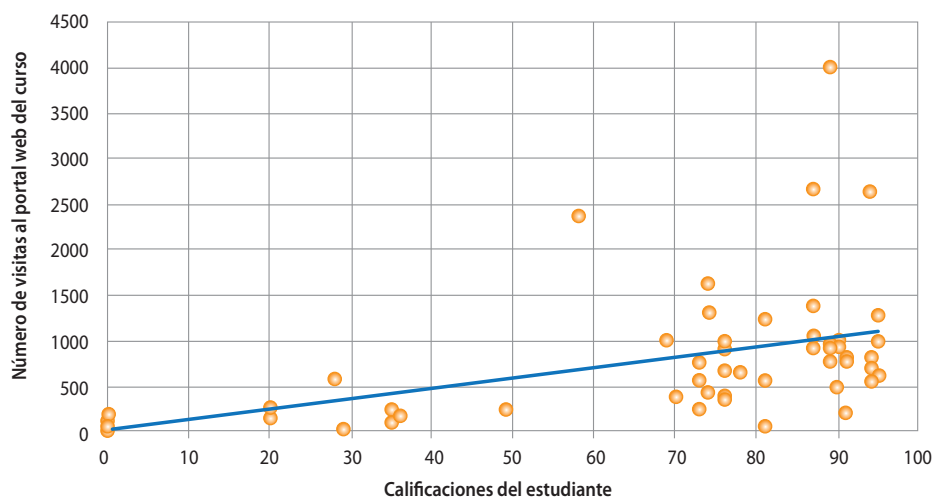
Figura 6. Histograma de frecuencia de las participaciones contabilizadas por estudiante en la página del curso durante el periodo comprendido desde el 14 de marzo hasta el 6 de junio del 2021



Fuente: Elaboración propia.

La figura 7 muestra un análisis gráfico de la relación entre el número de visitas y las calificaciones obtenidas por cada uno de los estudiantes. El análisis estadístico mediante la rho de Spearman mostró una relación positiva media entre dichas variables, ya que el coeficiente obtuvo un valor de 0.628 con un nivel de significancia de 0.00.

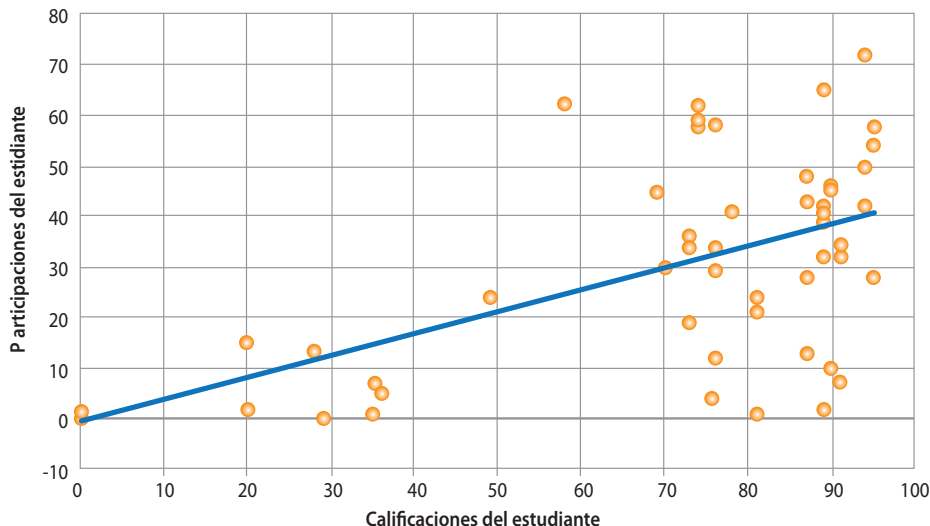
Figura 7. Diagrama de dispersión entre calificaciones del estudiante y el número de visitas al portal web del curso



Fuente: Elaboración propia.

La figura 8 muestra un análisis gráfico de la relación entre el número de participaciones y las calificaciones obtenidas por cada uno de los estudiantes. El análisis estadístico mediante la rho de Spearman mostró una relación positiva media entre dichas variables, ya que el coeficiente obtuvo un valor de 0.638 con un nivel de significancia de 0.00.

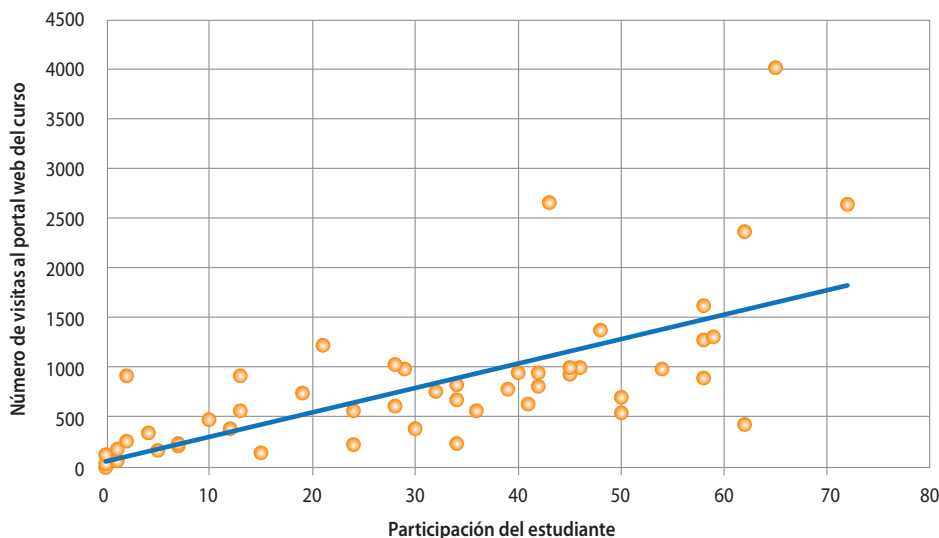
Figura 8. Diagrama de dispersión entre calificaciones del estudiante y el número de participaciones del estudiante



Fuente: Elaboración propia.

La figura 9 muestra un análisis gráfico de la relación entre el número de participaciones y las visitas de cada uno de los estudiantes. El análisis estadístico mediante la rho de Spearman mostró una relación positiva fuerte entre dichas variables, ya que el coeficiente obtuvo un valor de 0.814 con un nivel de significancia de 0.00.

Figura 9. Diagrama de dispersión entre el número de participaciones y las visitas de los estudiantes



Fuente: Elaboración propia.

Los distintos resultados evidencian que a través de la analítica web se puede recopilar información de vital importancia para la autogestión del docente en entornos de aprendizaje virtual. Barráez (2020) resalta que todo docente que desee gestionar adecuadamente la educación a distancia deberá estar en la capacidad de direccionar sus acciones hacia prácticas eficientes de enseñanza. Sin embargo, la eficiencia de las estrategias implementadas se lograría al partir de información válida y relevante. En concordancia con lo mencionado, se puede asegurar que el tráfico web generado en los cursos virtuales puede ser aprovechado para el diseño e implementación de métricas como punto de partida para un entorno de *learning analytics* con miras al entendimiento del comportamiento estudiantil. Tal como afirman Menezes et al. (2021), la aplicación de *learning analytics* facilita el análisis con miras a entender y mejorar los procesos de aprendizaje y el rendimiento de los estudiantes. Además, si se implementa de manera total permitiría comparar y explicar distintos escenarios a partir de métricas. En otras palabras, se puede considerar a *learning analytics* como una técnica que conjuga el potencial de la *web analytics* y la inteligencia de negocios para su implementación en el sector educativo con miras a lograr un aprendizaje adaptativo (Cukierman y Vendrell, 2020).

En el presente estudio, partiendo de métricas de *web analytics*, se pudo describir el comportamiento de los estudiantes, el cual demostró un alto tráfico de visitas y participaciones durante las primeras semanas del curso; sin embargo, dicha tendencia tuvo una pendiente negativa con algunos picos durante las últimas. Al respecto, Bortulé et al. (2020) menciona que las interacciones de los estudiantes con el material del curso virtual suelen incrementarse durante el periodo de evaluaciones. En ese sentido, dicha aseveración podría explicar los valores más altos recopilados durante las semanas 3 y 8, donde se registran actividades calificadas; sin embargo, el número de interacciones contabilizadas es considerablemente más bajo a partir de la semana 4, lo cual podría deberse a que el estudiante pierde interés por actividades no calificadas. De la misma manera, si el estudiante solo se concentra en la presentación de actividades calificadas resta el número de visitas al material de estudio (teoría). Por otro lado, debido a las limitaciones de la información, en el presente estudio no se han podido medir otras métricas relevantes, tales como: duración en minutos por cada visita, lecturas en línea o enlace a fuentes externas (Chan, Botelho & Lam, 2019; Pérez, García y López, 2013).

Posteriormente, en concordancia con los principios de *learning analytics*, se analizó la información recopilada en contraste con el rendimiento académico. En ese aspecto, la evidencia estadística, en el presente estudio, determinó coeficientes de correlación entre 0.62 y 0.64, a través de aplicar la rho de Spearman. En ese sentido, se pudo afirmar la existencia de una relación positiva media al correlacionar las métricas web (visitas y participaciones) con las calificaciones. En la misma línea investigativa, se puede mencionar el trabajo realizado por Álvarez et al. (2020), quienes hallaron coeficientes de correlación más débiles al relacionar calificaciones con métricas web, utilizando rho de Spearman; y entre los valores que obtuvieron se pueden referenciar los siguientes: notas con relación al acceso a URL ($r=0,283$ $p<0,05$), notas con relación a visitas ($r=0,282$ $p<0,01$) y notas con relación al acceso a recursos ($r=0,372$ $p<0,01$). A pesar de ello, otros estudios han demostrado una correlación media positiva ($p < 0.05$ para rho de Spearman) al analizar la utilización de recursos virtuales y el logro del aprendizaje (Benavente, 2018; Laura, Ramos y Suni, 2018). Otro punto que se debe destacar es que la información recopilada a través de *web analytics* y *learning analytics* podría sustentar distintos análisis de regresión lineal (Pérez, García y López, 2013) o, por otro lado, podrían permitir el diseño de modelos de predicción sobre el rendimiento académico. En el presente estudio no se pudieron emplear técnicas estadísticas más robustas, pero con una muestra más grande se podría seguir la metodología planteada en la investigación de Reyes et al. (2019), donde se utilizó una muestra de 300 estudiantes y se aplicaron redes neuronales y *learning analytics* para predecir el rendimiento académico.

SÍNTESIS Y REFLEXIONES FINALES

El estudio realizado ha permitido evidenciar que los entornos de aprendizaje virtual han tenido un crecimiento importante durante el 2020-2021 como consecuencia de la pandemia del covid-19. Dicho crecimiento también ha venido generando una cantidad enorme de datos e información que, en muchos casos, no están siendo aprovechados por las instituciones educativas ni por la gestión docente. Las entidades educativas deben preocuparse por adquirir soluciones informáticas que permitan sintetizar toda la información generada por los estudiantes, ya que son una fuente de vital importancia para descubrir, describir y predecir determinados comportamientos y resultados. De la misma manera, los docentes deben poseer las competencias digitales para recopilar, procesar o interpretar la información generada en entornos de aprendizaje virtuales.

De la misma manera, se ha podido concluir que los conceptos de *web analytics* y *learning analytics* aún no se dominan plenamente en el campo de la educación, lo que condiciona a seguir utilizando herramientas convencionales para la gestión del aprendizaje. La investigación ha permitido determinar que la *web analytics* resultaría una herramienta fundamental para que el docente pueda implementar un seguimiento efectivo sobre los estudiantes en cursos virtuales o a distancia.

Por otro lado, los resultados han permitido demostrar que los datos no solo serían útiles para una descripción en el tiempo respecto a la interacción del estudiante con el material de estudio y actividades del curso; sino que, al mismo tiempo, podrían explicar el rendimiento académico expresado en las calificaciones. Entonces, no solo se estarían diseñando métricas de seguimiento al tráfico web del estudiante, sino que se estaría desarrollando un entorno de *learning analytics* donde los datos recopilados y procesados permitirían comprender variables significativas para mejorar la gestión del aprendizaje a través de estrategias específicas y basando la toma de decisiones en la información. A partir de lo mencionado, el estudio ha podido concluir que la *web analytics* conlleva a desarrollar un entorno de *learning analytics* y, de la misma manera, permitiría medir variables con un grado de correlación importante respecto al rendimiento académico de los estudiantes en cursos virtuales.

REFERENCIAS

- Álvarez, P., Álvarez, A., Bravo, C., Cristóbal, J. y Angulo, T. (2020). Tipologías de estudiantes de Fisioterapia según el uso que hacen del campus virtual. *Revista d'Innovació Docent Universitària*, 1(12), 74-81. <https://doi.org/10.1344/ridu2020.12.8>
- Barráez, D. (2020). La educación a distancia en los procesos educativos: Contribuye significativamente al aprendizaje. *Revista Tecnológica-Educativa Docentes 2.0*, 8(1), 41-49. Recuperado de <https://ojs.docentes20.com/index.php/revista-docentes20/article/view/91>
- Benavente, E. (2018). *El uso de los recursos virtuales y su relación con el aprendizaje del idioma inglés en los estudiantes del VIII ciclo de la especialidad de Inglés-Español como Lengua Extranjera de la Universidad Nacional de Educación Enrique Guzmán y Valle, La Cantuta* (Tesis de segunda titulación). Universidad Nacional de Educación Enrique Guzmán y Valle, Perú. <http://repositorio.une.edu.pe/handle/UNE/1731>
- Bharara, S., Sabitha, S. & Bansal, A. (2018). Application of learning analytics using clustering data mining for students' disposition analysis. *Education and Information Technologies*, 23(2), 957-984. <https://doi.org/10.1007/s10639-017-9645-7>
- Bortulé, M., Scagliotti, A., Frisco, A., Corvalán, D., Cuch, D. y Vigh, C. (2020). Enseñanza virtual durante la pandemia, un curso de Física elemental. *Latin-American Journal of Physics Education*, 14(4), 1-9. <http://www.lajpe.org>

- Chan, A. K. M., Botelho, M. G. & Lam, O. L. T. (2019). Use of learning analytics data in health care-related educational disciplines: Systematic review. *Journal of Medical Internet Research*, 21(2), e11241. <https://doi.org/10.2196/11241>
- Cukierman, U. y Vendrell, E. (2020). Aprendizajes reales en ambientes virtuales. El rol de la tecnología en la era de la inteligencia artificial y el *big data*. *Cuaderno de Pedagogía Universitaria*, 17(34), 59-67. <https://doi.org/10.29197/cpu.v17i34.396>
- Hernández, J., Espinosa, J., Penaloza, M., Díaz, E., Bautista, M., Riaño, M., ... Bermúdez, V. (2018). Sobre el uso adecuado del coeficiente de correlación de Pearson: verificación de supuestos mediante un ejemplo aplicado a las ciencias de la salud. *Archivos Venezolanos de Farmacología y Terapéutica*, 37(5), 552-561. <https://www.redalyc.org/jatsRepo/559/55963207020/55963207020.pdf>
- Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación*. México D.F., México: McGraw-Hill Interamericana.
- Laura, J., Ramos, M. y Suni, L. (2018). *Los servicios del Moodle y Facebook y su relación en el proceso enseñanza-aprendizaje de la unidad didáctica de materiales industriales del I semestre del área académica de Mecánica de Producción, del I.E.S.T.P. José Carlos Mariátegui, Moquegua 2016*. (Tesis de maestría). Universidad Católica de Santa María, Perú. <http://tesis.ucsm.edu.pe/repositorio/handle/UCSM/7445>
- Menezes, A., de Souza, G., da Silva, R. & Teixeira, E. (2021). Uso de *learning analytics* sobre dados de simulados para apoio à avaliação da aprendizagem por professores e gestores. *Revista RENOTE Novas Tecnologias Na Educacao*, 18(2), 388-397. <https://doi.org/10.22456/1679-1916.110259>
- Nguyen, A., Gardner, L. & Sheridan, D. (2020). Data Analytics in Higher Education: An Integrated View. *Journal of Information Systems Education*, 31(1), 61-71. <https://aisel.aisnet.org/jise/vol31/iss1/5>
- Pérez, A. L., García, M. J. y López, M. L. (2013). Evaluación del proceso de un programa realizado a través de internet y de la telefonía móvil para promover conductas saludables en estudiantes de educación secundaria de España y México. *Revista Española de Salud Pública*, 87(4), 393-407. www.alertagrumete.com/backend.php
- Pinkwart, N. & Liu, S. (2020). *Artificial Intelligence Supported Educational Technologies* (Springer, Ed.).
- Ranjeeth, S., Latchoumi, T. P. & Paul, P. V. (2020). A Survey on Predictive Models of Learning Analytics. *Procedia Computer Science*, 167, 37-46. Elsevier B.V. <https://doi.org/10.1016/j.procs.2020.03.180>
- Restrepo, L. y González, J. (2007). De Pearson a Spearman. *Revista Colombiana de Ciencias Pecuarias*, 20(2), 183-192. <https://revistas.udea.edu.co/index.php/rccp/article/view/324135>
- Reyes, N. S., Morales, J. B., Moya, J. G., Terán, C. E., Rodríguez, D. N. & Altamirano, G. C. (2019). Model to predict academic performance based on neural networks and learning analytics. *RISTI - Revista Iberica de Sistemas e Tecnologías de Informacao*, (E17), 258-266. https://www.researchgate.net/publication/331178569_Model_to_predict_academic_performance_based_on_neural_networks_and_learning_analytics
- Rohloff, T., Oldag, S., Renz, J. & Meinel, C. (2019). Utilizing web analytics in the context of learning analytics for large-scale online learning. *IEEE Global Engineering Education Conference, EDUCON, April-2019*, 296-305. <https://doi.org/10.1109/EDUCON.2019.8725118>
- Salinas, J. (2020). Educación en tiempos de pandemia: tecnologías digitales en la mejora de los procesos educativos. *Innovaciones Educativas*, 22(Especial), 17-21. <https://doi.org/10.22458/ie.v22iEspecial.3173>
- Sciarrone, F. (2018). Machine learning and learning analytics: Integrating data with learning. *2018 17th International Conference on Information Technology Based Higher Education and Training, ITHET 2018*, 1-5. <https://doi.org/10.1109/ITHET.2018.8424780>
- Stewart, C. (2017). Learning Analytics: Shifting from theory to practice. *Journal on Empowering Teaching Excellence*, 1(1), 10. <https://doi.org/10.15142/T3G63W>
- Tobón, C. y Bedoya, J. (2017). Influencia de la asimetría en el tamaño de la muestra para el cumplimiento del teorema central del límite en distribuciones continuas. *Scientia et Technica*, 22(4), 398-402. <https://doi.org/10.22517/23447214.9325>