

# Identificación de la existencia de un tiempo óptimo de las sesiones educativas para maximizar el rendimiento académico

Mauricio Valenzuela Corvalán<sup>1</sup>

<sup>1</sup> Universidad de Chile. Facultad de Ciencias Físicas y Matemáticas

---

## Resumen

En la actualidad, la formación online ofrece una vasta gama de opciones en diversas disciplinas y niveles educativos. Actualmente se tienen almacenados una gran cantidad de datos valiosos sobre los procesos educativos de cada estudiante en plataformas de educación online. Este estudio se centra en analizar los datos de la plataforma EOL de la Universidad de Chile para predecir el rendimiento académico de los estudiantes del curso Economía IN2201 del segundo semestre del año 2022. Para explorar si es valioso, se crea un índice de similitud con lista ideal de eventos (ILE) donde evento de la ILE tiene un ponderador que refleja su importancia relativa en el proceso de aprendizaje y también se crea un identificador T haciendo alusión al tipo de estudiante que es según la hora donde acostumbra a estudiar. Este enfoque permite segmentar a los estudiantes en grupos según su adherencia a la ILE o su tipo y controlar por estas variables en modelos de regresión. Se consideran variables como el tiempo promedio por sesión y la cantidad de interacciones registradas como controles importantes en la regresión. Con esto se logró obtener evidencia empírica sobre la existencia de un tiempo de sesión óptimo para la mejora en el aprendizaje en formato virtual con el fin de proponer el diseño de una reestructuración modular de las sesiones de clases presenciales, modificando la duración de estas de 90 a 80 minutos. Este estudio también incluye una matriz de impacto ético para asegurar un uso responsable de los datos y maximizar el beneficio social del proyecto. Este trabajo busca proporcionar insights valiosos para tratar de impulsar una mejora en el desempeño estudiantil en entornos de educación virtual y presencial, tanto para la Universidad de Chile como para la implementación en cualquier institución educativa.

**Palabras Claves:** *Learning Analytics, Tiempo de sesión, Desempeño académico, Concentración*

---

## 1. Introducción

Así como el humano evoluciona, la forma de educarse también. Hoy en día, existe un infinidad de alternativas para acceder a formación online y a distancia para cualquier nivel de estudios, en cualquier disciplina, desde lo más generalista hasta lo más especializado, desde un curso de inglés hasta un máster en dirección y administración de empresas. La formación online es aquella que utiliza las Tecnologías de la Información y la Comunicación (TIC) para todo el proceso de aprendizaje y que se realiza a través de Internet. Asociado a este proceso se almacena todo tipo de datos e información valiosa sobre los procesos educativos de cada estudiante. Indagar en esta área nos da claves importantes para ayudar a mejorar el sistema educativo, potenciar las maneras de enseñar de los instructores y entregarle las mejores metodologías de estudio a los alumnos.

Para llevar a cabo este proceso se utilizan diversas plataformas, ya sea únicas por cada ofertante de cursos online (universidades con plataformas propias) o públicas (Udemy, Khan Academy, etc). Estas plataformas de aprendizaje en línea permiten una amplia gama de interacciones, tales como ver videos, participar en foros y realizar quizzes.

En este caso, se dispone a hacer un estudio de los da-

tos proporcionados por la herramienta de estudio on-line de la Universidad de Chile, EOL, con el objetivo de predecir el rendimiento de los estudiantes que si aprobaron el curso. Para explorar si es valioso, se crea un índice de similitud con lista ideal de eventos (ILE) donde evento de la ILE tiene un ponderador que refleja su importancia relativa en el proceso de aprendizaje y también se crea un identificador T haciendo alusión al tipo de estudiante que es según la hora donde acostumbra a estudiar. Este enfoque permite segmentar a los estudiantes en grupos según su adherencia a la ILE o su tipo y controlar por estas variables en modelos de regresión. Se consideran variables como el tiempo promedio por sesión y la cantidad de interacciones registradas como controles importantes en la regresión. Con esto se busca obtener evidencia empírica sobre la existencia de un tiempo de sesión óptimo para la mejora en el aprendizaje en formato virtual con el fin de proponer el diseño de una reestructuración modular de las sesiones de clases presenciales, modificando la duración de estas de 90 a 80 minutos. Es importante destacar que hay problemas de heterogeneidad y posibles variables omitidas en los modelos de regresión, pero dados los datos son muy difíciles de evitar, por lo que se tomará como un supuesto fuerte la definición de una lista ideal, en el desarrollo de este informe se profundizará en este problema y como se pretende

solucionarlo.

La realización de este proyecto conlleva una responsabilidad por parte de los desarrolladores con respecto al tratamiento de los datos disponibles y a los posibles usos que se le dará a la herramienta. Por ello, se realizará una matriz de impacto ético del proyecto para asegurar un buen uso de los recursos y que la herramienta resultante aporte un beneficio a la sociedad.

Este documento busca obtener claves importantes para el desempeño estudiantil basado en la educación virtual, en la siguiente sección (sección 2) se indagan los diversos estudios relacionados a LA y en específicos aquellos relacionados con registros significativos en la educación, en la sección 3 se entrega la metodología de preprocesamiento de datos para este estudio, en la sección 4 se observan los resultados más importantes obtenidos, en la sección 5 se presentan las principales implicancias de los hallazgos, en la sección 6 se entregan las conclusiones, en la sección 7 se presentan recomendaciones para trabajo a futuro y en la sección 8 se entrega la evaluación del impacto del estudio.

## 2. Antecedentes

Learning Analytics (LA) se define como la recopilación, medición y análisis de datos sobre estudiantes y sus contextos con el objetivo de comprender y optimizar el aprendizaje y los entornos en los que ocurre (Siemens, 2013). Las aplicaciones de LA en educación son variadas e incluyen la identificación de patrones de comportamiento estudiantil que se correlacionan con el rendimiento académico, así como el desarrollo de modelos predictivos para informar intervenciones educativas (Ferguson, 2012).

Dentro de las teorías del aprendizaje, el constructivismo sostiene que los estudiantes construyen activamente su conocimiento a través de interacciones significativas con el contenido y con otros (Piaget, 1952). Por otro lado, la teoría del aprendizaje autorregulado enfatiza la importancia de la autorregulación en el aprendizaje, que incluye el control sobre el tiempo de estudio y las estrategias de aprendizaje utilizadas por los estudiantes (Zimmerman, 2002). Estos dos pilares son los principales para este estudio pues son el fundamento detrás de utilizar ciertas variables específicas para crear los modelos o no utilizarlas.

El primero para segmentar lo que estudios han demostrado, esto es que las interacciones activas, como participar en foros y realizar quizzes, están más fuertemente correlacionadas con el éxito académico que las actividades pasivas, como ver videos (Kizilcec, Papadopoulos, & Sritanyaratana, 2014). Esto sugiere que no todas las interacciones tienen el mismo impacto en el aprendizaje. El segundo para considerar el factor de la duración del tiempo en pantalla como un aspecto crítico en el aprendizaje en línea pues investigaciones indican que existe una relación no lineal entre el tiempo que los estudiantes pasan en la plataforma y su rendimiento académico, sugiriendo la existencia de un "punto óptimo" de tiempo en pantalla que maximiza el aprendizaje (Anderson et al., 2014). Inclusive, la calidad del tiempo en pantalla es igualmente importante, con estudios mostrando que el uso efectivo del tiempo en pantalla, más que la cantidad, es lo que verdaderamente impacta el rendimiento académico (Pardo, Han, & Ellis, 2016).

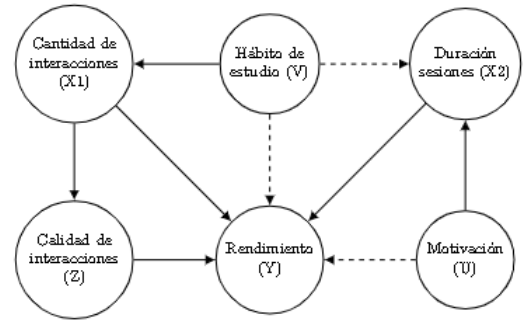


Figura 1: DAG de los efectos causales en el estudio

Aunque los datos están disponibles para estudiar estos dos factores, la causalidad entre estas actividades específicas y el rendimiento estudiantil no es directa, por la variabilidad del comportamiento de los estudiantes y otras variables no medibles. En la figura 1 se observa que no solo la cantidad de interacciones (Gil-Vera & Quintero, 2023) y la duración de las sesiones afectan al rendimiento del estudiante sino que también afecta la calidad de estas interacciones, una variable muy difícil de medir pero también afectan los hábitos de estudio que cada estudiante tiene, por otro lado la motivación (García et al., 2023) de estos afecta a la duración de las sesiones que estos tengan y también podría afectar tanto a la calidad de las interacciones como al rendimiento. De esta forma el problema que teníamos con la endogeneidad se hace notable al identificar las variables causales que no podemos medir que luego podrían sesgar los resultados.

Para intentar disminuir esto se pretende usar un factor de similitud de sus interacciones con la plataforma, esto buscando que la calidad de las interacciones venga representada por el factor, siendo un 100 % la realización de todas aquellas interacciones significativas "que el estudiante debió haber realizado para haber aprendido y luego reflejarlo en una buena calificación (Khalil & Ebner, 2016). Estos serán los controles diseñados para regresionar. Por otro lado, se harán modelos de clasificación, como Naive Bayes o SVM, para predecir el rendimiento académico (Baker & Inventado, 2014). La evaluación de estos modelos se realiza mediante las métricas de precisión y la matriz de confusión. La solución a este problema no solo mejorará la comprensión de las dinámicas de aprendizaje en línea, sino que también permitirá la implementación de estrategias pedagógicas más personalizadas y efectivas. La evidencia es consistente en demostrar que, independiente de la duración de la clase, la atención del estudiantado se desviará de forma frecuente durante una clase pasiva siendo necesarias estrategias interactivas para mantener su atención (Bunce et al., 2010; Christensen et al., 2018; Dvorkin et al., 2016) por lo que es de vital importancia lograr determinar un tiempo óptimo de las sesiones educativas para potenciar el aprendizaje y causar una mejora en el rendimiento académico.

## 3. Preprocesamiento de datos

Los datos utilizados en este estudio provienen de la plataforma EOL, que incluye registros detallados de las interacciones de los estudiantes con los recursos del curso. El preprocesamiento de estos implicó una fase de limpieza y una de transformación.

Inicialmente se leyó la base de datos desde un archivo CSV y se filtraron las columnas no relevantes para mantener solo

la información necesaria. Luego, se convirtió la columna de tiempo al formato adecuado para facilitar su manejo. Posteriormente, se contaron la cantidad de registros por alumno y se mantuvieron sus notas, ordenando estas interacciones de mayor a menor. Se calculó el tiempo promedio de sesión por usuario y el tiempo total de todas las sesiones. Se definió el identificador T que clasifica a los estudiantes en función de la franja horaria en la que más sesiones registran sus actividades en la plataforma educativa. Este identificador se define de la siguiente manera:

- **Diurno (0):** El estudiante tiene la mayoría de sus sesiones registradas entre las 08:01 y las 13:00. Estos estudiantes tienden a interactuar más con la plataforma en las horas de la mañana.
- **Tardío (1):** El estudiante tiene la mayoría de sus sesiones registradas entre las 13:01 y las 21:00. Estos estudiantes tienden a interactuar más con la plataforma en las horas de la tarde.
- **Nocturno (2):** El estudiante tiene la mayoría de sus sesiones registradas entre las 21:01 y las 08:00 del día siguiente. Estos estudiantes tienden a interactuar más con la plataforma en las horas de la noche.

Este identificador T proporciona una manera de categorizar a los estudiantes según sus patrones de uso de la plataforma educativa, lo cual puede ser útil para adaptar estrategias de enseñanza y soporte a sus necesidades específicas. Se identificó una lista ideal de eventos con coeficientes de impacto la cual se ha sido definida con base en la importancia de cada actividad educativa. Los coeficientes de impacto fueron asignados según la relevancia de cada tipo de actividad:

- **Interacciones directas con el video y resolución de problemas:** Estas actividades se ponderan con un 10 % cada una, ya que son fundamentales para el aprendizaje interactivo y activo.
- **Actividades complementarias a la realización de problemas y progreso en el curso:** Estas actividades se ponderan con un 2 % cada una, ya que aunque son importantes, su impacto es secundario.
- **Registros de navegación relevantes como foros, marcadores, resúmenes y ayudas(hints):** Estas actividades se ponderan con 0.1 % o 0.2 %, dependiendo de su utilidad directa en el proceso educativo.

Se adjunta el cuadro 2 en Anexos que especifica el coeficiente de impacto asociado a cada registro de la base de datos que ya solo le quedan aproximadamente la mitad de los registros iniciales. Todas estas métricas fueron añadidas a la base de datos principal. Finalmente, se realiza un análisis de correlación entre variables clave, lo que nos permitirá avanzar hacia un análisis exploratorio de datos (EDA) y el desarrollo de modelos de regresión. La base de datos incluye las siguientes variables:

- **conteo:** Número de interacciones registradas por cada usuario con la plataforma. Indica la cantidad de eventos significativos realizados por el usuario. ( $x_1$ )
- **sessions:** Número de sesiones diferenciadas por día que cada usuario ha tenido. Una sesión se define por un conjunto de interacciones realizadas en un mismo día. (U)

- **label:** Nota o calificación del usuario. Esta variable se usa para medir el rendimiento académico y es el objetivo principal en este estudio. (y)
- **avg\_session\_time:** Tiempo promedio de cada sesión en minutos para cada usuario. Esta métrica ayuda a entender cuánto tiempo invierten los usuarios en cada sesión de estudio. ( $x_2$ )
- **similitud:** Porcentaje de coincidencia de las interacciones del usuario con una lista de eventos considerada significativa. Esta variable mide qué tan alineadas están las actividades del usuario con las prácticas recomendadas. (Z)
- **total\_time:** Tiempo total en minutos que cada usuario ha pasado en todas sus sesiones combinadas. Esta variable proporciona una medida acumulativa del tiempo de uso de la plataforma. (V)
- **T:** Indicador que clasifica a los estudiantes en función de la franja horaria en la que más sesiones registran sus actividades (V)

Variable	Promedio
conteo	$\mu_{conteo} = 1650$
sessions	$\mu_{sessions} = 27$
label	$\mu_{label} = 5,1$
avg_session_time	$\mu_{ast} = 80 \text{ min}$
similitud	$\mu_{similitud} = 78 \%$
total_time	$\mu_{tt} = 2274 \text{ min}$

Cuadro 1: Promedios de cada variable

En el documento datos.Rmd y la Figura 4 de Anexos se puede observar más detalles sobre la metodología de la creación de la DB y un resumen de esta.

## 4. Resultados

En el archivo resultados.rmd se realizó un análisis exploratorio de datos (EDA) con el objetivo de identificar patrones y tendencias en los datos que indiquen comportamientos de aprendizaje y también modelos de regresión y de clasificación. Primeramente se genera un identificador para clasificar el rendimiento académico (RA) de los estudiantes en cuatro clases distintas: R, S, B y MB. Estas clases corresponden a diferentes rangos de notas obtenidas por los estudiantes y se definen de la siguiente manera:

- **R (Reprobado):** Notas menores a 4.
- **S (Suficiente):** Notas entre 4 y 5.
- **B (Bueno):** Notas entre 5 y 6.
- **MB (Muy Bueno):** Notas entre 6 y 7.

Se realizan gráficos de boxplot para notar diferencias entre clases de los identificadores T y RA con respecto a las notas y al tiempo promedio por sesión. Los hallazgos más relevantes son que para el indicador T, los alumnos tipo Diurno tienen el promedio de notas más bajo entre todos los grupos de estudiantes lo que podría indicar que este tipo de estudiante no se dedica el tiempo suficiente fuera de los horarios típicos de clases presenciales a estudiar de manera autónoma el contenido del curso por lo que identificar e impulsar ayudas para estos

estudiantes de manera temprana en el curso podría causar una mejor en el rendimiento académico. Así también, se identificó que el tiempo promedio por sesión no influye según tipo de estudiante. Por otro lado, para esta misma variable se observa que segmentando por el indicador RA hay una tendencia observable, el tiempo promedio por sesión por segmento es mayor a medida que el segmento es de mejor rendimiento.

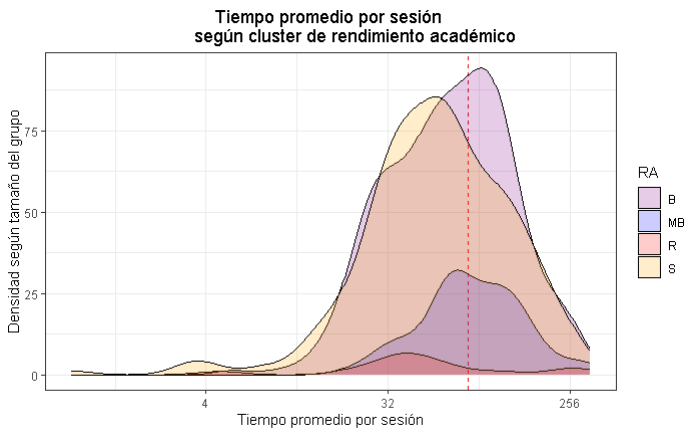


Figura 2: Tiempo promedio de sesión según identificador RA

En la figura 2 se observa la distribución del tiempo promedio por sesión según el tamaño de cada segmento identificando la línea vertical equivalente a un tiempo promedio por sesión de 80 minutos, lo más destacable es que para el segmento MB su mayor densidad está un poco atrás de la línea de corte (coherente con el promedio obtenido) indicando que si puede existir un tiempo promedio por sesión óptimo.

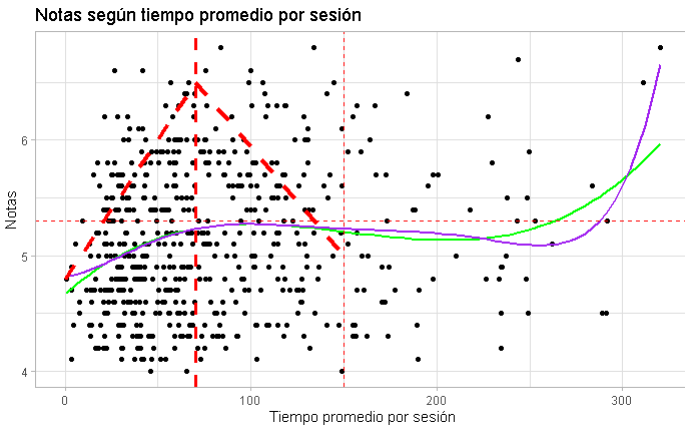


Figura 3: Gráfico de notas según el tiempo promedio por sesión

En la figura 3 se observan las notas respecto al tiempo promedio por sesión con regresiones de distinto grado (lineal, polinomial grado 2, 3 y 4). Se busca encontrar evidencia que refuerce el argumento de la importancia del tiempo promedio por sesión y que este posee un máximo óptimo. Se realizan varios modelos de regresión, tanto lineales como polinomiales de las cuales los resultados más interesantes son:

- Si realizamos los modelos considerando los alumnos que reprobaron se obtienen resultados diferentes a los modelos realizados sin los alumnos que reprobaron.
- El índice de similitud mejora un poco el modelo pero no es significativo por lo que este análisis de existencia de

evento y coeficiente de impacto no es lo suficientemente sofisticado para crear una herramienta de análisis eficaz dentro de este modelo de regresión.

- Se prueban 4 diferentes formas de introducir el conteo de interacciones, en donde introducir esta variable como una aproximación polinomial de grado 10, es mejor que introducirla como una variable lineal, polinomial de grado 2 o logarítmica y lineal y que si es un factor significativo al momento de regresionar.
- La variable del tiempo promedio por sesión es mejor introducirla como un polinomio de grado 3, lo relevante de esto es que en la mayoría de los modelos, el signo del regresor de la variable cuadrática es negativa lo cual indica que para esta variable (tiempo promedio por sesión) existe un máximo dada la concavidad de la función ( $\frac{\delta^2 y}{\delta x^2} < 0$ ) (Ver línea punteada de Figura 3).
- Cuando se controla por la variable de tiempo total se penaliza el tiempo promedio por sesión (cambia sus signo a todos sus regresores), la constante de regresión (se vuelve negativa) y la cantidad de sesiones (pierde significancia) por lo que dado su alto impacto en otras variables de medición y su poco aporte a la significancia no se considera incluir esta variable en el modelo de regresión final.
- Cuando el modelo se controla por el tiempo promedio por sesión y se condiciona según el tipo de estudiante se observa que no hay significancia de esta variable ( $x_2 * T$ ) pero es interesante notar que para el tipo tardío, el signo del regresor es positivo mientras que para el tipo nocturno el signo del regresor es negativo lo que implica en que estudiar sesiones extensas de noche son perjudiciales para las notas (Una nueva adaptación del índice  $T$  podría comprobar o refutar esta hipótesis).
- El número de sesiones es significativo como variable para el modelo lo que si coincide con el análisis previo de causalidad e inclusive el número promedio de sesiones es 27 lo que coincide con dos sesiones por semana que es lo recomendado.

Es por esto que el mejor modelo de regresión que se pudo obtener es el de la Figura 5 en Anexos. Para la construcción de los modelos de clasificación, se dividieron los datos en set de entrenamiento y de prueba y se entrenaron dos tipos de modelo de clasificación supervisada utilizando el primer set. Estos modelos aprenderán los patrones y relaciones entre las variables de entrada y las notas.

Los modelos utilizados fueron Naive Bayes Y Super Vector Machine (SVM), Una vez que el modelo se entrenó, se aplicaron a los datos de testeo de los estudiantes para predecir sus notas y asignarles una clase (R, S, B o MB). Luego de esto, se evaluó el rendimiento de cada modelo mediante las métricas de accuracy y la matriz de confusión. Con la accuracy se pudo identificar que el mejor modelo de clasificación entre estos dos es Naive Bayes, esto se debe a que el modelo de SVM clasifica a demasiados estudiantes de clase S o MB en la clase B empeorando sus predicciones. Como avance en paralelo se realizó una matriz de impacto ético para el proyecto la cual es puede observar en la Figura 6 de Anexos.

## 5. Discusión

Los resultados obtenidos entregan evidencia de la significancia que tienen la cantidad de interacciones de los alumnos con la plataforma y el tiempo promedio por sesión en el aprendizaje. De esta forma, se levanta información de la Universidad católica (1) (Centro de Desarrollo Docente, abril, 2023), de la USACH (2) y de la UDD (3) con respecto al tiempo ideal para una clase presencial. Aquí se destaca que las clases presenciales en las tres casas de estudio cambiaron su duración total de 90 minutos a 70 minutos, apuntando a 'enmarcarse dentro de las buenas prácticas educativas a nivel internacional', mientras que en la Universidad de Chile las clases presenciales duran 90 minutos. Al ver que el tiempo promedio por sesión de los alumnos en la plataforma online es de aproximadamente 80 minutos es interesante comenzar a considerar que esta información recopilada puede servir como marco teórico para realizar una propuesta sobre una nueva modulación horaria de las clases presenciales en la Universidad de Chile buscando en todo momento beneficiar la salud mental de los alumnos (y con esto su motivación) y con ello esperar obtener un mejor rendimiento.

## 6. Conclusiones

Este estudio valida la existencia de un tiempo de sesión óptimo en el contexto de experiencias educacionales. Las técnicas utilizadas ofrecen caminos prometedores para mejorar la intervención educativa en plataformas de aprendizaje online y presencial, destacando la necesidad de enfoques personalizados basados en el comportamiento de los estudiantes por ejemplo, buscar identificar de manera temprano el identificador T de cada alumno para reforzar a los alumnos del tipo 0 (Diurnos). Así también, el no obtener resultados favorables con respecto a una mejora en la predicción del rendimiento de los estudiantes gracias a la definición de una lista de eventos no implica que una variable más sofisticada (por ejemplo una creada basada en sequence mining) podría llegar a guiar a una mejor estructuración de los cursos para asegurar un aprendizaje óptimo y una mejor recepción por parte del estudiantado. Por último, el poder obtener un valor óptimo del tiempo de las sesiones en una plataforma online entrega interesantes señales para considerar la elaboración de nuevas políticas universitarias para la duración de las clases presenciales en todas las carreras de esta casa de estudio.

## 7. Trabajo a Futuro y Recomendaciones

- Mejorar la base teórica de la demostración de la existencia de un máximo local en el modelo de regresión controlado por el tiempo promedio por sesión.
- Mejorar el identificador T creando una nueva segmentación definiéndose ahora como Diurno, Tardío, Nocturno y Madrugador y analizar el modelo que controla por la variable  $x_2 * T$  para evidenciar si es significativo que la duración promedio de las sesiones durante la noche o durante la madrugada pueden afectar negativamente el rendimiento.
- Analizar otros modelos de clasificación para observar si alguno es mejor que los que ya se probaron.

- Propuesta de nueva modulación horario para la Universidad de Chile basado en la evidencia encontrada.

## 8. Evaluación de impacto

Este documento podría ser el puntapié inicial a la realización de una propuesta para un ajuste a la modulación horario en la FCFM de la Universidad de Chile basado en la evidencia encontrada.

## Bibliografía

Siemens, G. (2013). Massive Open Online Courses: Innovation in Education? In R. McGreal, R. Kinuthia, W. Marshall, S., & McNamara, T. (Eds.), *Open Educational Resources: Innovation, Research and Practice* (pp. 5-15). Athabasca, Canada: Athabasca University Press. <http://oasis.col.org/handle/11599/486>

Ferguson, R. (2012). The State of Learning Analytics in 2012: A Review and Future Challenges. Technical Report, Knowledge Media Institute, The Open University, UK. <http://kmi.open.ac.uk/publications/pdf/kmi-12-01.pdf>

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York, NY: W.W. Norton & Co. <https://doi.org/10.1037/11494-000>

Zimmerman, Barry. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*. 41. 64-70. 10.1207/s15430421tip4102\_2.

Kizilcec, René & Papadopoulos, Kathryn & Sritanyaratana, Lalida. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. *Conference on Human Factors in Computing Systems - Proceedings*. 10.1145/2556288.2557207.

Anderson, Ashton & Huttenlocher, Daniel & Kleinberg, Jon & Leskovec, Jure. (2014). Engaging with Massive Online Courses. *WWW 2014 - Proceedings of the 23rd International Conference on World Wide Web*. 10.1145/2566486.2568042.

Pardo, Abelardo & Han, Feifei & Ellis, Robert. (2016). Pardo, A., Han, F., & Ellis, R. (2016). Exploring the relation between self-regulation, online activities, and academic performance: A case study. In *Proceedings of the 6th International Learning Analytics and Knowledge Conference* (pp. 422-429). Edinburgh, UK.. 422-429. 10.1145/2883851.2883883.

Gil-Vera, Víctor & Quintero López, Catalina. (2023). Análisis de variables asociadas al rendimiento académico en cursos universitarios virtuales. *Formación universitaria*. 16. 33-42. 10.4067/s0718-50062023000400033.

García Vélez, Karen & Mendoza, Lilibeth & Cedeño, Dolores. (2023). Motivación estudiantil en los Entornos Virtuales de Aprendizaje. *Educación*. 29. 1-19. 10.33539/educacion.2023.v29n2.2970.

Khalil, Mohammad & Ebner, Martin. (2016). Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories. *Journal of Computing in Higher Education*. 29. 1-19. 10.1007/s12528-016-9126-9.

Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In *Learning Analytics* (pp. 61-75). Springer. [https://doi.org/10.1007/978-1-4614-3305-7\\_4](https://doi.org/10.1007/978-1-4614-3305-7_4)

Centro de Desarrollo Docente (abril, 2023). Orientaciones para la docencia en la nueva modulación horaria. Pontificia Universidad Católica de Chile. <https://desarrollodocente.uc.cl/recursos/tematicasdocentes/ajuste-de-modulacion-ho>

Bunce, D. M., Flens, E. A., & Neiles, K. Y. (2010). How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education*, 87(12), 1438-1443. <https://doi.org/10.1021/ed100409p>

Chen Zi'an, & Huang Bosheng. (2021). The effect of micro-course on attention and learning retention. *Journal of Educational Psychology*, 52(4), 885-907. [https://doi.org/10.6251/BEP.202106\\_52\(4\).0007](https://doi.org/10.6251/BEP.202106_52(4).0007)

Christensen, D., Villanueva, I., & Benson, S. (2018, November). Understanding First-Year Engineering Students' Perceived Ideal Learning Environments. In *2018 World Engineering Education Forum Global Engineering Deans Council (WEEF-GEDC)* (pp. 1-6). IEEE.

Dvorkin Camiel, L, Mistry, A., Schnee, D., Tataronis, G., Taglieri, C, Zaiken, K., Patel D., Nigro S., Jacobson S., and Goldman, J. Students' Attitudes, Academic Performance and Preferences for Content Delivery in a Very Large Self-Care Course Redesign. *American Journal of Pharmaceutical Education* May 2016, 80 (4) 67. <https://doi.org/10.5688/ajpe80467>

## Referencias

- [1] Universidad de Santiago de Chile, “Ajustes horarios de clases desde 2º semestre,” [En línea]. Disponible en: <https://www.vra.usach.cl/noticias/ajustes-horarios-de-clases-desde-2dosemestre>. [Accedido: 3-jul-2024].
- [2] Universidad del Desarrollo, “UDD reduce en 10 minutos la duración de módulos de clases desde este segundo semestre,” [En línea]. Disponible en: <https://www.udd.cl/noticias/2023/07/28/udd-reduce-en-10-minutos-la-duracion-de-modulos-de-clases-desde-este-segundo-semester/>. [Accedido: 3-jul-2024].
- [3] Pontificia Universidad Católica de Chile, “UC acortará en 10 minutos duración de módulos de clases a partir del segundo semestre de 2023,” [En línea]. Disponible en: <https://www.uc.cl/noticias/uc-acortara-en-10-minutos-duracion-de-modulos-de-clases-a-partir-del-segundo-semester-de-2023/>. [Accedido: 3-jul-2024].

## Anexos

Los coeficientes de importancia fueron ajustados para asegurar que su suma total sea igual a 1. De esta manera, cada evento recibe un peso proporcional a su impacto en el proceso de aprendizaje. Esta lista ideal de eventos permite evaluar y clasificar las actividades de los estudiantes, contribuyendo a un análisis más preciso de su desempeño y adherencia a las prácticas educativas recomendadas.

Evento	Coficiente de Importancia
load_video	0.1
play_video	0.1
pause_video	0.1
stop_video	0.1
seek_video	0.1
speed_change_video	0.1
problem_check_fail	0.1
problem_show	0.1
showanswer	0.02
problem_check	0.02
problem_graded	0.02
seq_next	0.02
seq_prev	0.02
seq_goto	0.02
save_problem_fail	0.02
edx.forum.thread.created	0.002
edx.forum.thread.viewed	0.002
edx.bookmark.added	0.002
edx.course.tool.accessed	0.002
edx.bookmark.listed	0.002
edx.bookmark.removed	0.002
edx.forum.response.created	0.002
edx.problem.hint.feedback_displayed	0.002
edx.problem.hint.demandhint_displayed	0.002
edx.grades.problem.submitted	0.002
edx.forum.thread.voted	0.002
edx.forum.comment.created	0.002
edx.forum.searched	0.002
edx.course.home.resume_course.clicked	0.001
edx.user.settings.changed	0.001

Cuadro 2: Lista ideal de eventos y sus coeficientes de importancia ajustados.



```

      username      label      conteo      nocturna
1      : 1      Min. :1.00      Min. : 42.0      Min. : 0.0
2      : 1      1st Qu.:4.60      1st Qu.: 965.8      1st Qu.: 461.8
3      : 1      Median :5.10      Median :1426.0      Median : 733.0
4      : 1      Mean :5.05      Mean :1650.7      Mean : 911.6
5      : 1      3rd Qu.:5.60      3rd Qu.:2040.8      3rd Qu.:1137.8
6      : 1      Max. :6.80      Max. :8101.0      Max. :5598.0
(Other):578
      tardía      diurna      T      sessions
Min. : 0.0      Min. : 0.00      0: 10      Min. : 1.00
1st Qu.: 272.0      1st Qu.: 6.75      1:196      1st Qu.:20.00
Median : 507.5      Median : 57.50      2:378      Median :26.00
Mean : 634.6      Mean : 104.52      Mean :27.38
3rd Qu.: 870.0      3rd Qu.: 133.25      3rd Qu.:33.00
Max. :3902.0      Max. :1192.00      Max. :73.00

      avg_session_time      total_time      similitud
Min. : 0.8592      Min. : 4.663      Min. :16.40
1st Qu.: 37.3635      1st Qu.: 937.951      1st Qu.:72.50
Median : 62.8384      Median : 1582.537      Median :82.40
Mean : 79.7369      Mean : 2273.599      Mean :78.32
3rd Qu.:106.7809      3rd Qu.: 2967.945      3rd Qu.:82.70
Max. :320.0450      Max. :12407.863      Max. :94.60

      rendimiento_academico
B :257
MB: 68
R : 14
S :245

```

Figura 4: Resumen de la DB creada para el estudio

Dependent variable:		
	label	
	(1)	(2)
poly(avg_session_time, 3)1	0.869 (0.833)	-4.127* (2.325)
poly(avg_session_time, 3)2	-1.384* (0.809)	-1.550* (0.809)
poly(avg_session_time, 3)3	0.456 (0.811)	0.592 (0.811)
poly(conteo, 10)1	1.405 (1.042)	1.265 (1.040)
poly(conteo, 10)2	-3.887*** (0.854)	-4.157*** (0.859)
poly(conteo, 10)3	1.670** (0.806)	1.860** (0.808)
poly(conteo, 10)4	-2.043** (0.794)	-2.123*** (0.792)
poly(conteo, 10)5	3.048*** (0.802)	3.132*** (0.800)
poly(conteo, 10)6	-2.117*** (0.789)	-1.905** (0.792)
poly(conteo, 10)7	2.050*** (0.779)	2.114*** (0.776)
poly(conteo, 10)8	-3.788*** (0.778)	-3.674*** (0.777)
poly(conteo, 10)9	2.414*** (0.778)	2.512*** (0.776)
poly(conteo, 10)10	-1.488* (0.778)	-1.430* (0.776)
sessions	0.017*** (0.004)	0.008 (0.006)
similitud	0.003 (0.005)	0.004 (0.005)
total_time		0.0001** (0.0001)
Constant	4.302*** (0.380)	4.262*** (0.379)
Observations	584	584
R2	0.260	0.267
Adjusted R2	0.240	0.246
Residual Std. Error	0.773 (df = 568)	0.770 (df = 567)
F Statistic	13.288*** (df = 15; 568)	12.883*** (df = 16; 567)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figura 5: Modelo de regresión mejor ajustado según  $R^2$

## Matriz de Dimensiones y Principios Éticos Afectados

Dimensiones / Principios	Beneficencia	No Maleficencia	Justicia	Autonomía
Estudiantes	3	5	5	3
Mandante (Institución Educativa)	5	5	5	5
Equipo de Proyecto (Investigadores, Analistas, Desarrolladores)	3	5	5	5
Sociedad (Comunidad Educativa y Pública en General)	5	3	4	5

Figura 6: Matriz ética de impacto del proyecto