

AI TOOL FOR EFFICIENT EVALUATION AND FEEDBACK

Ramona Fuentes Valdéz¹; María Luisa Mendoza Martínez²; Pedro Nájera García³

^{1,2,3}First, Second and Third Author's Tecnológico de Monterrey, México, rfuentes@tec.mx, ml.mendoza@tec.mx,
pedro.najera@tec.mx

Abstract– This article describes the innovative experience implemented in the TC1003B Engineering Modeling with Computational Mathematics Training Unit at the School of Engineering and Sciences of the Tecnológico de Monterrey, Campus Querétaro, using the Gradescope platform. This tool provides support in the assessment and feedback process for assignments and exams, streamlining the time dedicated to this task. It also allows for the analysis of results by activity and a summary of the main difficulties detected in each of the proposed exercises. This allows, if necessary, a clarification of the topic and an awareness of the most common errors developed in the exercises, enriching student learning. Gradescope uses Artificial Intelligence (AI) to classify students' responses in each assignment and exam according to response type. This makes the grading and feedback process more efficient and also promotes a more objective and impartial evaluation. Additionally, it streamlines the identification of areas of improvement in the proposed questions, in terms of wording, level of difficulty, time required to complete them and a specific rubric for each exercise in the implemented activities. Throughout the study, three variables were measured: engagement, learning gain, and use of technology. The results obtained from the pre-test and post-test comparison reflected high levels of satisfaction in each of these variables. Likewise, 88% of the students achieved a learning gain greater than 24 points, demonstrating a significant improvement in their academic performance and in the attainment of the course competencies.

Keywords– Technological Innovation, Artificial Intelligence, Higher education, Gradescope, Educational Innovation.

HERRAMIENTA DE IA PARA LA EVALUACIÓN Y RETROALIMENTACIÓN EFICIENTE

Ramona Fuentes Valdéz¹; María Luisa Mendoza Martínez²; Pedro Nájera García³
^{1,2,3}First, Second and Third Author's Tecnológico de Monterrey, México, rfuentes@tec.mx, ml.mendoza@tec.mx,
pedro.najera@tec.mx

Resumen– Este artículo describe la experiencia de innovación que se implementó en la Unidad de Formación TC1003B Modelación de la Ingeniería con Matemática Computacional de la Escuela de Ingeniería y Ciencias del Tecnológico de Monterrey en Campus Querétaro, utilizando la plataforma Gradescope. Esta herramienta brinda un apoyo en el proceso de evaluación y retroalimentación de tareas y exámenes, agilizando el tiempo dedicado a ello; permitiendo, además, analizar los resultados por actividad y logrando una síntesis de las principales dificultades detectadas en cada uno de los ejercicios propuestos, lo que permite, en caso necesario, realizar una clarificación del tema y tener presente los errores más comunes que se desarrollan en los ejercicios enriqueciendo el aprendizaje del alumno. Gradescope hace uso de Inteligencia Artificial (IA) para clasificar las respuestas de los estudiantes en cada tarea y exámenes según su tipo de respuesta. Esto permite hacer más eficiente el proceso de calificación y retroalimentación, además de favorecer una evaluación más objetiva e imparcial. Adicionalmente, agiliza la identificación de áreas de oportunidad en los reactivos propuestos, en cuanto a redacción, grado de dificultad, tiempo requerido para resolverse y se puede contar con una rúbrica particular para cada uno de los ejercicios de las actividades implementadas. A lo largo del estudio se midieron tres variables: engagement, ganancia de aprendizaje y uso de la tecnología; los resultados obtenidos en la comparación del pre-test y del post-test reflejaron altos valores de satisfacción en cada una de las dimensiones de estas variables. Asimismo, el 88% de los estudiantes alcanzó una ganancia de aprendizaje superior a 24 puntos, evidenciando una mejora significativa en su desempeño académico y en el logro de las competencias del curso.

Palabras clave– Innovación Tecnológica, Inteligencia Artificial, Educación superior, Gradescope, Innovación Educativa.

I. INTRODUCCIÓN

La unidad de formación de Modelación de la Ingeniería con Matemática Computacional, es una materia de cinco semanas, que se imparte en el segundo semestre de las carreras de Tecnologías Computacionales (ITC), Robótica y Sistemas Digitales (IRS), Transformación Digital de Negocios (ITD) y Ciencia de Datos y Matemáticas (IDM), por lo que los antecedentes esperados en los temas que se abordan son distintos. La innovación permitirá tener identificados esos temas que se requieren retomar y reforzar según las necesidades específicas de cada uno de los grupos en los que se imparte la materia. Además, de que un problema común en el ámbito educativo es el tiempo dedicado a evaluar y retroalimentar las actividades de aprendizaje desarrolladas en los diversos cursos, lo cual puede llegar a ser una actividad repetitiva [10] que con el uso de herramientas actuales, como

las de IA, pudiera ser más eficiente y efectivo sin causar tanto desgaste del tiempo académico, el cual pudiera utilizarse de una forma mucho más productiva en la elección, diseño e implantación de elementos innovadores en los cursos y de mayor impacto en los resultados de evaluación esperados, así como en la adquisición de competencias disciplinares; estas herramientas en equipo con el docente, parecen ser el futuro en las tutorías.

Durante los últimos años se ha estado incorporando el uso de chatbots inteligentes que apoyan en diversas áreas, por ejemplo: ChatGPT que es una herramienta diseñada para generar lenguaje natural que proporciona la sensación de estar interactuando con otra persona [7]. En las diferentes áreas de especialidad, como la de diseño digital que es parte del curso de Matemática Computacional, también han surgido herramientas particulares que permiten apoyar en el desarrollo de diseños auxiliados con IA, como lo es Flux.ai que cuenta con un chatbot que permite diseñar más rápido y controlar los errores que se producen en el desarrollo de prototipos, es decir, se identifican los errores y se señalan a los usuarios de la herramienta [6].

Se puede observar cómo los estudiantes, cada vez en mayor medida, interactúan con herramientas de IA y sería de esperar que para un proceso tan crítico y exhaustivo como el proceso de evaluación y retroalimentación se utilicen herramientas que incorporen IA para facilitarlos. Gradescope es una de ellas y permite agilizar el proceso de evaluación de los grupos y facilitar el diseño y desarrollo de actividades (exámenes, quizzes, tareas, actividades de clase y revisión de código) que se pueden clasificar por tipo de respuestas para proporcionar una retroalimentación dirigida; lo que proporciona una imagen clara de los resultados de los estudiantes [11], [12].

Es por ello que se pretenden agilizar los tiempos dedicados al proceso de evaluación y destinarlos al acompañamiento de los alumnos, además al identificar el tipo de errores más comunes se puede tener un proceso de retroalimentación mucho más efectivo.

II. MATERIALES Y MÉTODOS

El estudio se llevó a cabo en el primer tercio del semestre febrero a junio 2024, en dos grupos de la Unidad de Formación TC1003B Modelación de la Ingeniería con

Matemática Computacional, del segundo semestre de las carreras de ITC, IRS, ITD e IDM, con una población de 51 estudiantes, 37% mujeres y 63% hombres, edad promedio de 19 años. Por otra parte, en la plataforma Gradescope se diseñaron y evaluaron 5 actividades y un examen; la duración de la experiencia fue de 5 semanas y se aplicó un instrumento en dos ocasiones (pre-test y pos-test) para medir la variable de ganancia de aprendizaje, consistente en un quiz de 15 preguntas sobre los contenidos abordados en el curso: Lógica, Teoría de Conjuntos y Álgebra Lineal. Considerando los resultados de ambas aplicaciones, así como la calificación promedio final obtenida por cada uno de los alumnos se realizó un análisis de estadística descriptiva: media, varianza, observaciones, correlación Pearson, P de una cola, así como la prueba t de ganancia de aprendizaje de todos los estudiantes. Adicionalmente, a partir del conjunto de datos obtenidos se realizaron otros cálculos: media, error estándar, mediana, moda, desviación estándar, varianza muestral, asimetría, rango, máximo y mínimo.

En la variable Engagement se realizó un análisis de las dimensiones: vigor (motivación, energía y disposición); dedicación (entusiasmo, inspiración y sentimiento de orgullo); y absorción (felicidad, inmersión e intensidad). Se integraron los resultados de las tres dimensiones que componen al instrumento para medir el Engagement, tanto en pre-test como en post-test. En el análisis se obtuvieron datos atípicos que fueron descartados en la prueba t para dos muestras relacionadas, y solo se consideraron los datos con ambos resultados en las variables.

En la variable de uso de la Tecnología se evaluaron 5 dimensiones: Experiencia de aprendizaje (diseño de la actividad, instrucciones y explicación de contenidos); Enjoyment (comodidad, motivación y tiempo dedicado); Engagement (actitud positiva, espero con entusiasmo la actividad en la plataforma, interés y curiosidad y uso de la herramienta en el futuro); Aprendizaje significativo (relacionar conocimiento, aplicar conocimiento y conectar con conocimiento previo); y Experiencia de Usuario (satisfacción, actividad de aprendizaje y acompañamiento docente). El análisis de datos, estadística descriptiva, gráficos y tablas se realizaron utilizando Excel; en el apartado de hallazgos se explica con mayor detalle el análisis de los resultados obtenidos de los datos, así como tablas y gráficos de las variables analizadas.

El curso se trabaja bajo un enfoque de aprendizaje colaborativo y vivencial, al desarrollar un reto enfocado en la elaboración de un nuevo producto para apoyar en abordar la problemática que presenta la industria vidriera sobre costos, ventas, diseños, entre otros; lo que hace que el análisis de datos se realice con una situación conocida e identificada por los alumnos y se realizan encuestas para validar sus propuestas de solución para los diferentes grupos de edades (jóvenes,

adultos y adultos mayores), según sea el caso. También se tienen sesiones dedicadas a la elaboración de los instrumentos de adquisición de datos, así como para su análisis posterior.

III. HALLAZGOS

El uso de una herramienta como Gradescope de apoyo en el proceso de evaluación y retroalimentación agiliza el tiempo dedicado a estos procesos, además de permitir analizar los resultados por actividad de los grupos generados en el proceso de evaluación y tener un resumen de los principales problemas que se detectan en los ejercicios propuestos como actividad o tarea [8]. Además, permite tener una evaluación auténtica sin sesgos, ya que solo se visualizan las actividades a evaluar independientemente del alumno que haya realizado la entrega.

Cuando la evaluación se diseña para relacionar el avance directo en los objetivos de aprendizaje el lograr tener una retroalimentación puntual, clara y oportuna permite que los estudiantes avancen de mejor manera en el desarrollo de los contenidos del curso y fomenta una mayor participación y conocimiento de los contenidos esperados [2].

En esta experiencia de innovación se realizó un pre-test y un post-test, para medir cada una de las tres variables que se consideraron a lo largo del estudio: ganancia de aprendizaje, engagement y uso de la tecnología. Para la variable de ganancia de aprendizaje se aplicó un quiz que permitiera establecer las diferencias individuales y determinar el punto de partida de los estudiantes [3], obteniéndose los resultados de la estadística descriptiva que se muestra en la tabla 1.

Aprendizaje	Post-test	Pre-test
Mean	86.6470	41.0784
Standard Error	1.6638	2.0077
Median	87	39
Mode	87	49
Standard Deviation	11.8824	14.3385
Sample Variance	141.1929	205.5937
Kurtosis	2.6991	0.0812
Skewness	-1.5124	0.4758
Range	52	68
Minimum	48	9
Maximum	100	77
Sum	4419	2095
Count	51	51

TABLA I.

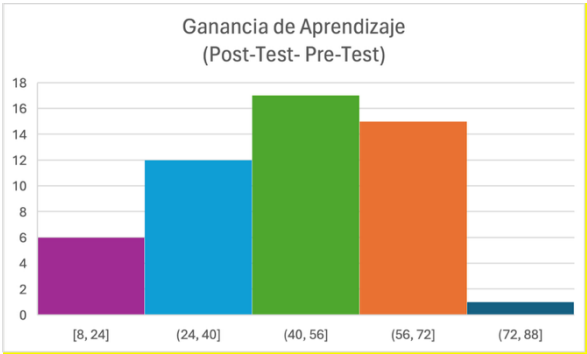
Estadística descriptiva de la variable de ganancia de aprendizaje

(Fuente: Elaboración propia).

En estos resultados se puede observar que los valores de tendencia central del post-test son muy cercanos entre sí lo

que implica una distribución más parecida a la normal, con un rango más pequeño que el pre-test, y teniendo sus valores extremos significativamente mayores; el pre-test muestra una distribución con un sesgo negativo, observado por la diferencia entre sus valores de los parámetros de tendencia central. Atendiendo a los valores de la desviación estándar de los parámetros de tendencia central y al encontrarse muy cercanos se concluye que tienen una dispersión equivalente.

Cuando categorizamos el intervalo de la ganancia de aprendizaje en 5 subintervalos con 16 puntos de magnitud por subintervalo, se observa que 12% obtuvo el menor aprendizaje (menor o igual a 24 puntos) y el complemento 88% obtuvo una ganancia de aprendizaje de más de 24 puntos, haciendo evidente el beneficio de la estrategia didáctica (ver Gráfica 1).

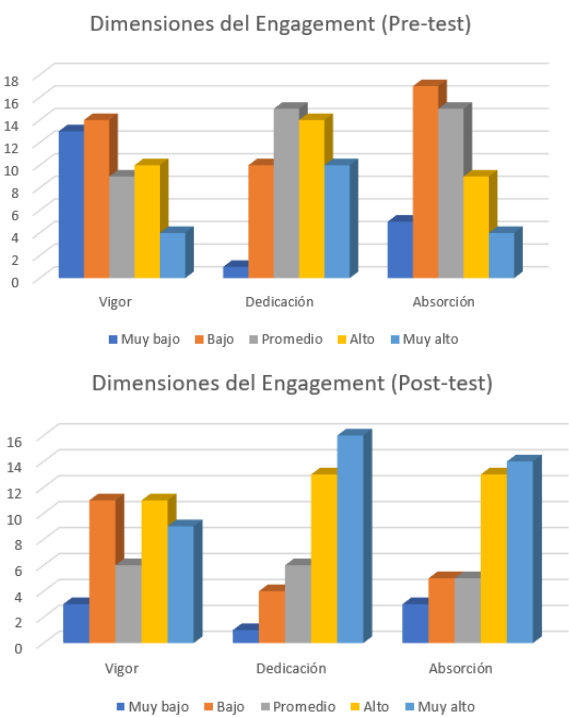


GRÁFICA 1.

Histograma de la variable de ganancia de aprendizaje

(Fuente: Elaboración propia).

Con relación a la variable de Engagement, se incluyeron los elementos de vigor (que está determinado por la motivación, la energía y la disposición), la dedicación (establecida en términos del entusiasmo, la inspiración y el sentimiento de orgullo), así como la absorción (definidos en función de la felicidad, la inmersión y la intensidad), dejándose de lado edad y sexo considerando que no tienen un efecto significativo en este tipo de estudios [4].

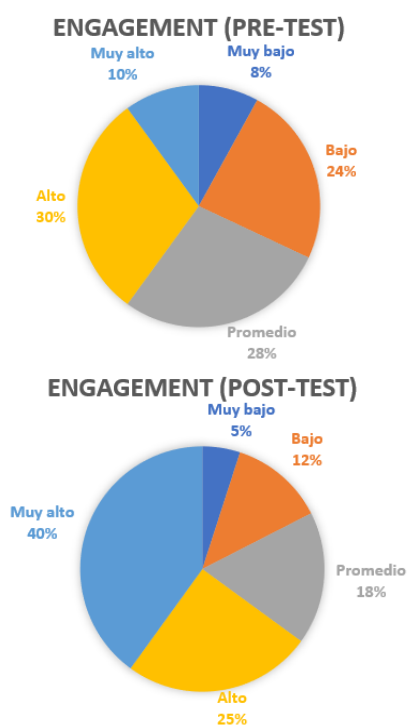


GRÁFICA 2.

Dimensiones de la variable Engagement

(Fuente: Elaboración propia).

Con relación a los elementos del Engagement (ver gráfica 2), el post-test muestra un logro de altos valores de satisfacción en todos los elementos considerados: vigor, dedicación y absorción comparado con la etapa del pre-test. Por otro lado, si se integran los resultados de estos tres elementos para post-test y para pre-test de la variable Engagement, se puede visualizar en el gráfico circular (ver gráfica 3) el incremento en los resultados que se obtuvieron de forma significativa del 40% al 65% integrando las categoría de Muy Alto y Alto, al igual que la reducción de los valores de las categorías de Bajo y Muy Bajo del 32% al 17% mostrado.



GRÁFICA 3.

Integración de las dimensiones de la variable Engagement

(Fuente: Elaboración propia).

La variable de Engagement se ve afectada de forma positiva por el desarrollo de las etapas del reto a lo largo de las cinco semanas (ver Fig. 1), el tener un aprendizaje activo permite que el estudiante se mantenga con una mayor motivación [9] y se involucre en mayor medida en lograr un buen aprendizaje y por lo tanto, que el proceso de evaluación y retroalimentación sea significativo para permitirle corregir en caso necesario y avanzar hacia las siguientes etapas.



Fig. 1 Etapas del desarrollo del reto de la unidad de formación

(Fuente: Elaboración propia).

El instrumento utilizado para medir la variable de uso de la Tecnología que incluye las dimensiones de: Experiencia de aprendizaje, Enjoyment, Engagement, Aprendizaje significativo y Experiencia de Usuario, se aplicó al final del uso de Gradescope como herramientas de IA para apoyar en el proceso de evaluación y retroalimentación. Con este instrumento se pretendían medir aspectos de la retroalimentación personalizada, detallada, evitando sesgos entre los participantes, pero además considerar aspectos como la accesibilidad, sostenibilidad y correcta forma de uso [5]

tanto para actividades individuales como colaborativas, y se obtuvieron los resultados que se muestran en la tabla 2.

Para las dimensiones de Experiencia de Aprendizaje y Experiencia de Usuario se identifica una diferencia significativa entre los valores promedio correspondientes a los reactivos que los evalúa; sin embargo, para las dimensiones de Enjoyment, Engagement y Aprendizaje Significativo los valores promedio de las respuestas significativas los muestran con una alta correlación. Lo cual se corrobora con los valores de las desviaciones estándar correspondientes.

Dimensión	Descripción pregunta	Pregunta	Promedio de Respuesta	Desviación Estándar
Experiencia de aprendizaje	Diseño de la actividad	1	82.3333	14.9888
	Instrucciones	2	98.3333	4.0824
	Explicación de contenidos	3	88	12.8840
Enjoyment	Comodidad	4	84	21.3541
	Motivación	5	92.5	18.3711
	Tiempo dedicado	6	91	17.6068
Engagement	Actitud positiva	7	89.6666	18.9384
	Ansio actividades en la plataforma	8	79.1666	19.1668
	Interés y curiosidad	9	87.5	19.7762
	Uso de la herramienta en el futuro	10	82	18.9631
	Relacionar conocimiento	11	84.6666	15.8324
Aprendizaje Significativo	Aplicar conocimiento	12	87.6666	17.4661
	Conectar con conocimientos previos	13	86.6666	16.1946
Experiencia de Usuario	Satisfacción	14	88.3333	21.3603
	Actividad de Aprendizaje	15	87.3333	14.7873
	Acompañamiento docente	16	100	0

TABLA II

Estadística descriptiva de la variable de ganancia de aprendizaje

(Fuente: Elaboración propia).

IV. IMPLICACIONES

A los alumnos les emociona el empleo de una herramienta de IA que pueda facilitar el proceso de aprendizaje y la herramienta de Gradescope les generó una alta expectativa con relación al proceso en cómo iba a ser empleada por lo que consideramos positivo contar con una herramienta que genere esta disposición positiva al uso de tecnologías de IA por parte del docente.

Fue importante comunicar a los alumnos que la herramienta de Gradescope no realizaba la evaluación de manera automática al entregar sus actividades, proceso que seguía estando a cargo del profesor responsable del módulo correspondiente en la unidad de formación; es decir, “La docencia no se está reemplazando, pero se está redefiniendo” [13].

Lo que si realiza la plataforma de Gradescope es la clasificación por tipo de ejercicios, para poder facilitar la tarea de retroalimentación, no únicamente de cada ejercicio en lo individual sino del desempeño de todo el grupo en cada ejercicio, generando una buena retroalimentación y detectando áreas de oportunidad, tanto de aspectos no desarrollados apropiadamente (como las competencias), así como mejorar

reactivos (en su redacción, grado de dificultad y tiempo necesario para responderlos).

La facilidad de la herramienta para la identificación y clasificación de los ejercicios generando categorías, permite evaluar los de un mismo tipo en una sola ronda de revisión. Sin embargo, el aspecto del anonimato de los ejercicios, que pudiera considerarse una magnífica manera de mantener una evaluación justa, en algunos casos puede llegar a complicar la evaluación, dado que en esta unidad de formación en particular se desarrollaron ejercicios encadenados, necesitando identificar lo que cada alumno había desarrollado previamente, situación que no se pudo realizar de forma adecuada con la herramienta y se tuvo que reevaluar acudiendo al documento original del alumno.

Otro aspecto del uso de la herramienta es considerar que para preguntas abiertas o aquellos reactivos que tienen múltiples maneras de ser resueltos, no son muy apropiados para actividades evaluables en Gradescope porque cada reactivo se clasifica como de un tipo diferente.

Con relación a los docentes, es imprescindible contar con un acercamiento apropiado a este tipo de herramientas. La configuración y manejo del sitio requiere de cierto grado de pericia, alta paciencia y un sentido de exploración para determinar la mejor manera de emplearla de acuerdo al tipo de actividades de cada curso. No hay manera única de hacer uso de la herramienta para cualquier tipo de unidad de formación, por lo que siempre habrá algo que experimentar. Además, es necesario tener claro el objetivo de incluir una herramienta de IA, la justificación, el diseño, los ajustes entre plataformas, el análisis de los datos que se generan sobre los reactivos, así como la organización de los docentes para hacer un uso conjunto de la plataforma [1].

V. CONCLUSIONES

El uso de la herramienta Gradescope en la unidad de formación facilitó la evaluación y retroalimentación, permitiendo analizar resultados por actividad y de este modo tener un resumen de los problemas más comunes para después tener un repaso general sobre ese tema y clarificar todas las dudas que se tuvieran; al final, esto permitió a los estudiantes mejorar la comprensión de los conceptos de los diferentes temas que se abordan en el curso: lógica, conjuntos, álgebra lineal y hacer uso de simuladores; además, se logró un mejor desempeño y un mayor avance en el nivel de logro de las competencias del curso. Por otro lado, el uso de Gradescope permitió tener una evaluación sin sesgos dado que sólo se visualizan las actividades a evaluar sin referencia al alumno que realizó la entrega de la misma. La retroalimentación es personalizada para cada uno de los ejercicios y, se puede tener una rúbrica tan detallada como se requiera para los problemas

siendo útil sobre todo en la evaluación de procedimientos detallados.

Dado que con esta innovación se agilizan los tiempos que se tienen dedicados al proceso de evaluación y retroalimentación, esto permite destinar ese tiempo en el acompañamiento de los alumnos, además poder identificar y tener presente el tipo de errores más comunes y poder reforzar esos puntos del tema en particular implica que la retroalimentación será más efectiva y el éxito en esta experimentación radica en que motiva y mejora el aprendizaje de los estudiantes de primer año de ingeniería sentando unas buenas bases para sus siguientes cursos.

Los resultados obtenidos en el análisis de las tres variables medidas ponen de manifiesto la importancia de utilizar la tecnología de manera efectiva para potenciar la experiencia de aprendizaje y estimular el desarrollo de habilidades clave en la formación de los estudiantes. Asimismo, se observó que el 88% de los estudiantes alcanzó una ganancia de aprendizaje superior a 24 puntos, lo que evidencia una mejora significativa en su desempeño académico y en el logro de las competencias del curso, reforzando el impacto positivo de las estrategias implementadas.

AGRADECIMIENTO

Los autores agradecen al Writing Lab del Instituto para el Futuro de la Educación del Tecnológico de Monterrey, México, por el apoyo financiero brindado para la realización de este trabajo. Los autores agradecen a la iniciativa NOVUS del Tecnológico de Monterrey por el apoyo brindado para la implementación del Proyecto IEAD & Novus “Innovación en la Evaluación y retroalimentación Académica con Inteligencia Artificial: Gradescope” (Propuesta 576787). También agradecemos al Departamento de Computación y al Departamento de Ciencias del Tecnológico de Monterrey Campus Querétaro por su apoyo en el desarrollo de este Proyecto. Agradecemos a Ileana A. Ochoa Arias, Celeste A. Romero Pardo, Sergio H. Sánchez Martínez, Verónica A. Pérez Aguirre y Gilberto J. Bermudez Langarica por el constante acompañamiento durante el proceso de diseño e implementación de esta iniciativa. El segundo autor agradece a SNII-México y SECIHTI por el apoyo recibido.

REFERENCIAS

- [1] Bagur Pons, S., Rosselló Ramon, M. R., Paz Lourido, B., & Verger, S. (2021). El Enfoque integrador de la metodología mixta en la investigación educativa. RELIEVE- Revista Electrónica De Investigación Y Evaluación Educativa, 27(1). <https://doi.org/10.30827/relieve.v27i1.21053>
- [2] Brown, S. (2015). “La evaluación auténtica: el uso de la evaluación para ayudar a los estudiantes a aprender”. RELIEVE - Revista Electrónica De Investigación Y Evaluación Educativa, 21(2). <https://doi.org/10.7203/relieve.21.2.7674>
- [3] Escudero Escorza, T. (2014). Desde los test hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación

- [illegible]