

Rendimiento académico y pruebas departamentales estandarizadas: un estudio en educación superior

Academic performance and standardized departmental tests: A study in higher education

Jorge Gustavo Gutiérrez Benítez • José Manuel Casillas Domínguez • Karina Olguin Jiménez

RESUMEN

Actualmente las instituciones educativas son cada vez más exigidas para brindar resultados que evidencien la calidad de sus procesos, razón por la que optan por aplicar procedimientos de evaluación estandarizados que contribuyan al logro de tales exigencias. Este trabajo analiza la calidad psicométrica y los resultados obtenidos sobre rendimiento académico en pruebas departamentales estandarizadas para Morfología y Morfosintaxis de la Segunda Lengua, asignaturas de mayor relevancia curricular del tronco común de Lenguas en la Facultad de Idiomas de la Universidad Autónoma de Baja California, analizando principalmente tres factores de influencia en las mismas: el género del sustentante, el docente que imparte la clase y el formato de la prueba. Empleando análisis de varianza, correlacionales y pruebas de muestras independientes a una muestra de 1,025 estudiantes que aplicaron la prueba desde el año 2017 al 2023, se observó un nivel de rendimiento académico promedio satisfactorio para una prueba mientras que uno deficiente para la otra, así mismo se identificó al docente como principal factor influyente en el rendimiento académico, seguido por el formato de aplicación de la prueba. Los resultados evidencian una calidad psicométrica en las pruebas superior al estándar, señalan la importancia de establecer un programa de capacitación docente, así como de estructurar un proceso enseñanza-aprendizaje homologado pero lo suficientemente flexible que resulte en mejorar el rendimiento en la prueba departamental para todos los estudiantes.

Palabras clave: Segunda lengua, estandarización, evaluación, psicometría.

ABSTRACT

Educational institutions are increasingly required to pursue results that demonstrate the quality of their processes, which is why they choose to apply standardized assessment methods and procedures to contribute to achieve such requirements. The present work analyzes the psychometric quality and the results obtained in terms of academic performance of standardized departmental tests for the two subjects of greatest curricular relevance of the common core of languages in the School of Languages at the Universidad Autónoma de Baja California, analyzing mainly three factors of influence in them. Variance analyses, correlational analyses, and independent sample tests were used on a sample of 1,025 students who answered the test from 2017 to 2023. A satisfactory average level of academic performance was observed for one test while a deficient one for the other, and the teacher was identified as the main influencing factor in academic performance, followed by the format of application of the test. The results show a higher-than-standard psychometric quality in the tests, pointing out the importance of establishing a teacher training program as well as structuring an approved teaching-learning process but flexible enough to result in improving performance in the departmental test for all students.

Keywords: Second language, standardization, assessment, psychometry.

INTRODUCCIÓN

En la actualidad cada vez se exige más que las instituciones educativas persigan mejoras con las cuales evidenciar la calidad de sus procesos. En el caso particular de las universidades tales exigencias se traducen en la implementación de nuevos métodos o estrategias que permitan concretar una mejor calidad educativa, iniciativas que no se limitan solo a los aspectos de la enseñanza o el aprendizaje, sino que también buscan ser implementadas en los procesos evaluativos.

En razón de lo anterior es que algunas instituciones optan por aplicar métodos y procesos de evaluación estandarizados para determinar el grado de competencia que los estudiantes alcanzan en las asignaturas que cursan en los diferentes programas educativos. Una de las ventajas que ofrecen este tipo de pruebas es alcanzar valoraciones más objetivas y el señalamiento concreto de parámetros en el rendimiento que estén vinculados a ciertas metas educativas (Hernández et al., 2018), y dados los procedimientos metodológicos con los que se construyen este tipo instrumentos estandarizados es posible determinar tendencias en el rendimiento de los estudiantes, obtener indicadores del rendimiento del profesor, así como coadyuvar a una toma de decisiones más efectiva al contar con información válida y confiable (Fernández et al., 2017).

La evaluación estandarizada ha recibido múltiples opiniones sobre su incidencia dentro del proceso de aprendizaje de una lengua, siendo estas un medio para el control de la calidad y la eficacia del propio sistema educativo (Benavides, 2015). En este tenor, la Universidad Autónoma de Baja California –UABC– desde el año 2016 implementó un programa de capacitación en la elaboración y diseño de pruebas estandarizadas. Esta estrategia se basó en la participación de las diferentes unidades académicas a nivel estatal que pertenecen a la UABC en un diplomado impartido por el Instituto de Investigación y Desarrollo Educativo –IIDE–.

Jorge Gustavo Gutiérrez Benítez. Profesor-Investigador de la Facultad de Idiomas Tijuana, extensión Tecate, Universidad Autónoma de Baja California, México. Es Ingeniero en Computación por la Facultad de Ciencias Químicas e Ingeniería de la Universidad Autónoma de Baja California, Maestro en Tecnologías de la Información y Comunicación por la UABC y Doctor en Innovación en Tecnología Educativa por la Universidad Autónoma de Querétaro. Tiene una especialidad en evaluación estandarizada del aprendizaje. Miembro del Sistema Nacional de Investigadoras e Investigadores y del Centro de Estudios Avanzados e Investigación en Ciencias Sociales y Humanidades. Correo electrónico: gutierrez.jorge@uabc.edu.mx. ID: <https://orcid.org/0000-0003-3392-6398>.

José Manuel Casillas Domínguez. Profesor-Investigador de tiempo completo de la Facultad de Idiomas Mexicali de la Universidad Autónoma de Baja California, México. Doctor en Educación. Tiene el reconocimiento al perfil PRODEP. Entre sus publicaciones recientes se encuentra el libro *Análisis de un examen estandarizado: el caso del examen departamental de Morfosintaxis de la Segunda Lengua de la Facultad de Idiomas*. Correo electrónico: jcasillas@uabc.edu.mx. ID: <https://orcid.org/0000-0002-3627-990X>.

Karina Olguin Jiménez. Técnico Académico de Tiempo Completo adscrita a la Facultad de Idiomas Tijuana, extensión Tecate, de la Universidad Autónoma de Baja California, México. Es Licenciada en Docencia de Idiomas y Maestra en Lenguas Modernas, así como especialista en traducción e interpretación, estudios realizados en la Universidad Autónoma de Baja California. Es Doctora en Educación y trabaja las líneas de investigación de la traducción literaria, calidad educativa, formación docente, políticas lingüísticas y traducción inclusiva. Correo electrónico: karina.olguin@uabc.edu.mx. ID: <https://orcid.org/0009-0008-8651-5545>.

La Facultad de Idiomas –FI– fue una de las participantes en este programa, teniendo como resultado la elaboración de dos pruebas departamentales estandarizadas, específicamente de las asignaturas de mayor relevancia curricular del tronco común (primeros dos semestres del programa) de idiomas según se encuentra establecido en el plan de estudios y mapa curricular para los programas de Licenciatura en Enseñanza de Lenguas y Licenciatura en Traducción. Estas materias fueron Morfología de la Segunda Lengua (inglés) de primer semestre y Morfosintaxis de la Segunda Lengua (inglés) de segundo semestre, mismas que se encuentran seriadas entre sí.

Ambas pruebas se han aplicado desde el año 2017 en las cuatro unidades académicas que integran a la FI en el estado de Baja California, por ello el propósito de este estudio es analizar los principales resultados obtenidos respecto al rendimiento académico de una muestra representativa de estudiantes en dichas pruebas departamentales, durante el periodo comprendido del 2017 al 2023, con el fin de exponer los beneficios que la evaluación estandarizada provee en los contextos de educación superior, caso particular de la disciplina de las lenguas, así como también poner de manifiesto las posibles limitaciones y adecuaciones necesarias para mejorar la efectividad de estas pruebas en el logro de las competencias esperadas por los estudiantes, establecidas en los programas de cada una de las asignaturas en cuestión.

Existe una perspectiva estratégica con la que suele utilizarse la evaluación, en el caso de las evaluaciones estandarizadas, “también conocidas como objetivas o de gran escala, ya que rebasan el ámbito del aula para proporcionar resultados que sean confiables, válidos y comparables entre distintas poblaciones de estudiantes” (Backhoff, 2018, p. 3), proveyendo así de información con la cual tomar decisiones que giren en torno a la mejora de los procesos educativos.

Puesto que esto último, respecto a la toma de decisiones, es de especial interés para las instituciones educativas, distintos organismos evaluadores de Latinoamérica (como lo son el INEE en México, el SIMCE en Chile, el INEED en Uruguay, el INEP en Brasil, entre otros) reconocen la importancia de este tipo de evaluación, pues identifican que “tiene como propósito monitorear el desarrollo de aprendizajes esenciales de los estudiantes para la mejora continua en el proceso de enseñanza y aprendizaje” (Ministerio de Educación Pública, 2023, p. 5).

REFERENTES TEÓRICOS

Evaluación y pruebas estandarizadas

Las pruebas estandarizadas, entendidas en el marco internacional como “evaluaciones comparativas que realizan las agencias y organizaciones internacionales; usadas como instrumentos estandarizados de medición, aplicadas en varios países de manera conjunta, seleccionando grupos o grados de estudiantes similares” (Demarchi, 2020, p. 111), pueden ser categorizadas en dos: pruebas de norma referenciada y de criterio referenciado

(Jiménez et al., 2017, Burkett, 2018). En lo que concierne a las primeras, el rendimiento de los aplicantes evaluados es comparado con otros aplicantes seleccionados con base en características como lo son la edad y el género (Yang, 2023).

Dentro de las principales características de la evaluación normativa se destaca la de tener como objetivo determinar la posición relativa de cada alumno en relación a los demás, que cumple su función cuando se logra identificar quién tiene mejor dominio y quién no. De esta manera se pueden establecer comparaciones para posteriormente realizar una selección entre los alumnos con base en el criterio de quién es mejor (Heredia, 2009). Las evaluaciones basadas en pautas de normalidad son instrumentos para lograr una discriminación entre los examinados en una situación de relación cerrada en la que solo existe posibilidad de admisión para un número limitado y definido de individuos.

Por otra parte, en las pruebas basadas en criterios, introducidas por Robert Glaser en 1968, los participantes son evaluados con un criterio fijo y estándar. Este tipo de evaluación decide si el participante ha alcanzado un objetivo determinado o domina una habilidad. (Yang, 2023). La característica fundamental de este tipo de evaluación radica en determinar el logro de los objetivos de forma individual en cada alumno sin necesidad alguna de compararlo con el resto de los estudiantes. La evaluación criterial permite operativizar la evaluación del aprendizaje individual del alumno respecto a objetivos educativos o metas claramente definidas que a su vez funcionan como un criterio comparativo (Heredia, 2009).

Expuestos ambos tipos de evaluaciones, sus características y principales usos, en este proyecto se determinó utilizar evaluaciones referidas a criterio ya que las pruebas estandarizadas están elaboradas con base en el currículo de la asignatura, mismo con el cual se puede observar el nivel de dominio individual alcanzado por cada estudiante como prueba de su rendimiento académico.

Teorías para la elaboración de pruebas

Además de clasificar las evaluaciones estandarizadas en los tipos antes mencionados, para la elaboración de este tipo de pruebas existen diferentes métodos, los cuales usualmente se resumen en la aplicación de dos teorías principalmente: por un lado, la teoría clásica de los tests –TCT–, y por el otro, la teoría de respuesta al ítem –TRI–.

Abordando a la TCT, sus principios datan del siglo XX con la construcción del primer test estandarizado de inteligencia por Binet, empleando para ello los trabajos de Pearson y Spearman, esta teoría después tuvo un mayor impulso y difusión en los años 50 (Muñiz, 2010). El avance en el campo de la medición y elaboración de pruebas gracias a la TCT derivó en el surgimiento de otras teorías que tenían como principal fin mejorar aquellos aspectos susceptibles de error y ausencia de medición en la TCT, tales como la teoría de la generalización y la teoría de respuesta al ítem.

Específicamente la TRI pretende un fundamento probabilístico al problema de la medición de constructos latentes, en especial los no observables, lo anterior representa

una de las ventajas de la TRI en relación a la TCT, misma que tiene que ver con el sesgo, ya que con la TRI se proporciona la posibilidad de conceptualizar los sesgos a nivel del ítem (Cortada, 2004).

Actualmente se han desarrollado diversos trabajos basados en la TRI (Baladrón et al., 2018; Ferreyra y Backhoff, 2016; Gutiérrez y Acuña, 2023), los cuales muestran que esta teoría es aplicable en diferentes áreas del conocimiento.

En términos generales, la calidad técnica de un examen estandarizado puede observarse a través de atributos de validez y confiabilidad. Por un lado, la validez para una prueba basada en la TRI se observa empleando el cálculo del índice de dificultad (valor p medido de 0 a 1) de cada reactivo como de la prueba en general; índice de discriminación (valor D medido de 0 a 1), definido como la propiedad que tiene un ítem para poder separar a los sujetos sustentantes que tienen una mejor puntuación final en la prueba de quienes tienen una menor puntuación (Medina et al., 2019), y el coeficiente de discriminación (valor r_{pbis} medido de -1 a 1) que permite observar la correlación existente entre los puntajes obtenidos por los sujetos en un ítem específico con el puntaje obtenido en su totalidad en la prueba (Molina et al., 2015, en Aguilar-Salinas y de las Fuentes, 2023). Estos tres atributos permiten observar la validez de la prueba.

En el caso de los atributos que permiten demostrar la confiabilidad se emplean cálculos de consistencia interna (Viladrich et al., 2017) como el índice de confiabilidad de Kuder-Richardson (KR20), también se encuentra el índice del alfa de Cronbach, aunque este es mayormente utilizado para pruebas con ítems que evalúan un único dominio o unidimensionales (para ítems que examinan más de una dimensión se emplea el cálculo del alfa de Cronbach estratificado).

Rendimiento académico

Como se mencionó anteriormente, emplear evaluaciones estandarizadas tiene un aporte sustancial para la toma de acciones por parte de los responsables en las instituciones educativas, mismas que se norman en razón de atributos o indicadores observables que den pauta a la calidad educativa. En este sentido, en una escuela comúnmente se evalúan aspectos de la producción (disciplina, puntualidad, limpieza, infraestructura, excelencia en logros académicos, excelencia en logros no académicos, clima organizacional y satisfacción de los clientes) y aspectos de la calidad (estudiantes, profesores, recursos instruccionales, calidad del liderazgo directivo y sistema de evaluación del desarrollo), siendo estos últimos a los cuales se pretende contribuir mediante el uso de las pruebas estandarizadas (Dhal, 2021).

Al observar la calidad del docente pueden mencionarse varios elementos, como lo son la formación inicial (la carrera estudiada por el profesorado y sus líneas de investigación o rendimiento), los años de experiencia en la docencia, la metodología de enseñanza seleccionada para la materia, el dominio de las temáticas impartidas durante los cursos, entre otros (Fandiño, 2017).

Comúnmente se asocia el uso de pruebas o las pruebas en sí a la valoración de lo que se conoce como *rendimiento académico*, concepto un tanto complejo que se ve afectado por múltiples factores y atributos, entre los cuales se destacan las características del propio estudiante, las características del docente, la naturaleza del programa educativo, la institución educativa, entre muchos otros. Debido a esto, efectuar una medición o valoración del rendimiento académico en un estudiante es un procedimiento que implica la recolección de múltiples variables que están directamente relacionadas con la ejecución de una o múltiples tareas, sobre las cuales se puede observar a un grado razonable esta movilización de saberes o competencias dentro de todo el universo de conocimientos que el estudiante posee, movilización necesaria para el logro de dichas tareas, en cuyo caso la intención de estos procesos de recolección es contar con información que provea un pronóstico de un posible rendimiento académico con el cual a su vez sea posible tomar decisiones en aras de mejorar el aprendizaje en los estudiantes (Gutiérrez-Monsalve et al., 2021; Rico y Gaytán, 2022).

En términos generales, el rendimiento académico se entiende como la relación existente entre aquello que el estudiante aprende y lo logrado en el proceso de enseñanza-aprendizaje resultado de la ejecución de las diferentes tareas académicas en las que se desempeñó (Masrek y Zainol, 2015; Rodríguez et al., 2004). En el contexto de educación superior el poder determinar el rendimiento académico de los estudiantes retoma considerable importancia, puesto que en esta etapa se pretende que se alcance una formación profesional sobre la cual el estudiante conducirá su vida laboral, misma que aspira a satisfacer una demanda social de profesionistas.

En razón de lo anterior, se han realizado diversas investigaciones que giran en torno a metodologías o procedimientos que permitan pronosticar el rendimiento académico de los estudiantes, por ejemplo, utilizando algoritmos de aprendizaje automático, técnicas de árbol de decisión, regresiones lineales, entre otras (Contreras et al., 2020; Castrillón et al., 2020; Guambuete et al., 2023; Gutiérrez-Monsalve et al., 2021). Algunos autores clasifican los predictores del rendimiento en dos grupos principales, factores internos y externos, sobre los cuales se distinguen aquellos atributos que inciden directamente en el rendimiento académico (Grimaldo y Manzanares-Medina, 2023; Neto y Losada, 2021). Entre aquellos asociados a factores internos se encuentran el horario de la clase, tipos de estrategias implementadas, tecnología utilizada, tipo de exámenes, cantidad de alumnos en el grupo, por mencionar algunos; mientras que, por otro lado, los factores externos se refieren a cuestiones de índole social o familiar, laborales, económicos, entre otros.

Para las instituciones educativas es importante identificar oportunamente estos factores internos y externos que influyen en el rendimiento académico de los estudiantes, principalmente aquellos que se encuentran dentro de la clasificación de los predictores internos, ya que, al contar con ellos, estos sirven como un insumo base en la toma de decisiones y acciones encaminadas al beneficio de los estudiantes, es decir, mejoras directas a la calidad educativa que reciben los estudiantes en el día a día del quehacer académico (Mendoza y Herrera, 2013; Torres-Zapata et al., 2022).

Siguiendo esta misma línea, varios estudios (Borja et al., 2021; Garbanzo, 2007; Hernández et al., 2018; Soza, 2021) han demostrado la influencia que existe en el rendimiento académico asociado a factores internos como los antes mencionados, incluyendo desde luego los tres que se abordan en este trabajo, dando así en parte el sustento que da razón de lo pertinente que es analizar en esta investigación estos factores de influencia.

MÉTODO

Unidad de análisis

La Facultad de Idiomas de la UABC tiene presencia en diferentes municipios del estado, por lo que para esta prueba se trabajó con distintas poblaciones (correspondientes a cada uno de los distintos campus presentes en los cuatro municipios) a lo largo del periodo comprendido entre los años 2017 al 2023, considerando que a raíz de la pandemia de la COVID-19 en los años 2020 y 2021 no se aplicaron estas pruebas. Se contó con la participación de 1,025 alumnos provenientes de las cuatro sedes (Mexicali, Ensenada, Tijuana y Tecate), quienes cursaron el tronco común de la Licenciatura en Idiomas. El género de la población fue de un 32.29% masculino y un 67.71% femenino. El rango de edad de los participantes osciló entre 18 y 24 años, considerando que estos exámenes se aplican en el primer y segundo semestres. Los 1,025 alumnos representaron el 93.7% de la totalidad de alumnos que realizaron estas pruebas durante el lapso que comprende esta investigación, mismos que presentaron tanto el examen de Morfología como el de Morfosintaxis. Es importante señalar que de los 1,025 alumnos solo 851 de este mismo grupo de participantes aplicaron el examen de Morfosintaxis de manera regular en segundo semestre, la diferencia de los 174 que no fueron contabilizados se dio por las siguientes razones: repitieron la asignatura de Morfología, se dieron de baja al concluir el primer semestre, exentaron la prueba o deliberadamente no presentaron el examen. Cabe mencionar que en algunos ciclos escolares en ciertas sedes no se aplicaron las pruebas, aunado al periodo de la COVID-19 que ya se mencionó. Lo anterior permitió contar con una muestra más que representativa del universo de alumnos de la FI de la UABC, a la luz de analizar los resultados de la evaluación en diferentes cohortes de rendimiento académico.

Así mismo es importante mencionar que esta iniciativa de institucionalizar la aplicación de pruebas departamentales surgió como una respuesta a una de las políticas educativas de la UABC con las que se pretendía mejorar la calidad educativa de los estudiantes universitarios particularmente en torno a las evaluaciones.

Estos exámenes fueron elaborados por el comité colegiado de evaluación del aprendizaje constituido por docentes de la FI, comité que recibió formación especializada en la elaboración de pruebas estandarizadas, empleando una metodología específica para la elaboración de exámenes referidos a criterio. Actualmente solo se aplican estos dos exámenes estandarizados.

En relación a la ética con la que se desarrolla esta investigación, se señala que todos los datos recolectados para este estudio no involucraron información confidencial o privada de los estudiantes, los datos se manejaron bajo la nomenclatura de participante 1, 2 ... n , mismo escenario aplicado para los docentes participantes en el estudio.

Instrumentos

En relación con el instrumento que se empleó, se precisa que los exámenes departamentales de Morfología y Morfosintaxis están diseñados para evaluar el nivel de dominio que el estudiantado posee en relación con todo el universo de conocimientos curriculares de dichas asignaturas. En el caso de la asignatura de Morfología, al ser una asignatura seriada con Morfosintaxis, implica la adquisición de conocimientos clave que permiten concretar otros conocimientos, por ello la importancia de determinar el nivel de aprendizaje alcanzado por el alumnado y con ello hacer predicciones sobre el rendimiento futuro.

La metodología utilizada en la elaboración de los exámenes departamentales fue formulada en el Instituto de Investigación y Desarrollo Educativo –IIDE– (Contreras, 2000; Contreras y Backhoff, 2004) empleando como fundamento el paradigma psicométrico expuesto por Nitko (1994) para la confección de pruebas de gran escala bajo el enfoque criterial. Dicha metodología contempla seis etapas en el desarrollo del examen: 1) definición del dominio de resultados que pretende el currículo, 2) análisis del currículo, 3) desarrollo de un plan de evaluación, 4) producción y validación de ítems, 5) análisis primario de resultados y 6) análisis secundario de resultados. Estos exámenes están orientados por el currículo, que implica que todas aquellas decisiones sobre lo que se va a evaluar y la forma de evaluarlo están directamente orientadas por lo que se establece en el currículo, este mismo está expresado en los programas de las unidades de aprendizaje –PUA– o cartas descriptivas de dichas asignaturas. En dichos PUA están definidas las competencias que se desea que los estudiantes alcancen a nivel de unidad temática, así como del curso en general.

Ambas pruebas emplean reactivos de opción múltiple, con una duración de 120 minutos; el examen de Morfología está compuesto por un total de 63 reactivos y el de Morfosintaxis 61. Ambos exámenes se aplican en la semana inmediata en la que se termina el calendario oficial de clases, comúnmente llamada como periodo de exámenes finales, de forma tal que los estudiantes ya hayan cursado todo el contenido temático de la asignatura.

Respecto a la calidad técnica de ambos instrumentos, para dar pauta de la confiabilidad y validez de los mismos se aplicaron los cálculos del índice de dificultad (p), índice de discriminación (D) y el coeficiente de discriminación ($rpbis$), así como el cálculo de la consistencia interna ($KR20$).

En relación a los instrumentos considerados para la evaluación formativa-sumativa que hace el docente, se consideraron solamente los promedios o calificaciones finales de la asignatura, datos que fueron obtenidos a través de los sistemas institucionales o de consulta directa con los docentes. Cabe mencionar que en ambas asignaturas los docentes refirieron utilizar exámenes escritos para las evaluaciones formativas y sumativas durante la impartición de dichas asignaturas.

Procedimiento

Para los análisis entre variables, como por ejemplo las relaciones entre los resultados obtenidos en la prueba departamental estandarizada y los obtenidos en la evaluación formativa-sumativa que realiza el docente en la asignatura, se aplicaron procedimientos estadísticos con los cuales determinar las posibles correlaciones existentes, siendo así el uso del coeficiente de correlación de Pearson como medida de la dependencia lineal entre las dos variables cuantitativas mencionadas anteriormente (Fiallos, 2021), el análisis de varianza (Anova) y la prueba de muestras independientes.

Para esta investigación se consideraron tres factores internos como variables independientes con las cuales analizar el posible comportamiento del rendimiento académico de los estudiantes en las pruebas departamentales, a saber, el género, el formato de la prueba y el docente que impartió el curso. La elección de estos criterios surgió como parte de un proceso de juicio en el que participaron docentes que impartieron la materia, personal directivo y el comité de evaluación colegiada del aprendizaje, haciendo hincapié en que estos se definieron como factores internos para una primera aproximación a la descripción del problema de estudio, además de considerar la información que se tenía disponible hasta el momento en que se realizó el estudio y la que era viable de obtener de los periodos comprendidos abarcados por el mismo. A su vez, el rendimiento académico como variable dependiente se observó con la calificación obtenida en las pruebas departamentales estandarizadas en conjunto con la calificación sumativa emitida por el docente al finalizar cada uno de los cursos en cuestión.

Para realizar los cálculos estadísticos se empleó el *software* estadístico Statistical Package for the Social Sciences (SPSS) en su versión 28 y el *software* especializado SIEXAES para generar algunos de los reportes de resultados de las aplicaciones de las pruebas departamentales estandarizadas, otros se construyeron de forma manual con base en la información recabada por cada una de las sedes de la FI. Así mismo se utilizó un compendio de hojas de cálculo diseñado exprofeso para integrar todos los resultados del rendimiento de los estudiantes que presentaron la prueba departamental estandarizada en el periodo abarcado por el estudio. Las calificaciones generadas tanto en los exámenes departamentales estandarizados como las de los docentes se dan en una escala que va de 0 a 100 puntos.

Para los resultados de las evaluaciones formativas realizadas por el docente se recurrió a consultar los promedios finales obtenidos en cada una de las asignaturas organizados por ciclo escolar y grupo. Posteriormente se integraron en un nuevo compendio de hojas de cálculo y se hizo un cruce de datos para asociar el resultado de la evaluación sumativa de la prueba departamental con el resultado de la evaluación formativa del docente. Este último compendio de datos es el que se utilizó en su mayoría para el análisis con el *software* SPSS.

RESULTADOS

A continuación se presentan los principales hallazgos después de la ejecución de los procedimientos de análisis de los datos recabados. Primero se presentan los resultados de las evaluaciones sumativas correspondientes a las pruebas departamentales a nivel estado en la Tabla 1. En dicha tabla se observa que los estudiantes logran un nivel de rendimiento mínimo satisfactorio (promedio de 70.03) sobre el dominio de todo el universo de conocimientos comprendido por la prueba de Morfología. Por su parte, en la asignatura de Morfosintaxis el promedio general de rendimiento se ubicó en 53.81, siendo significativamente menor al promedio de la asignatura anterior, es decir que aproximadamente cinco de cada diez estudiantes obtienen una calificación menor al mínimo necesario de dominio con el que se acredita la prueba.

Ahora bien, según el análisis psicométrico realizado a la calidad técnica de cada una de las pruebas, para el caso de Morfología el índice de dificultad de la misma se ubicó en .64, lo que significa que la prueba tiene una tendencia a ser más fácil que difícil, el índice de discriminación se ubicó en .344, mientras que el coeficiente de discriminación (r_{pbis}) se ubicó en .337, logrando casi un valor excelente (.35) para este índice (Carlos et al., 2011). Para la asignatura de Morfosintaxis el análisis psicométrico de la prueba arrojó un índice de dificultad de .458, esto significa que el examen tiene una tendencia a ser más difícil, se alcanzó un valor del índice de discriminación de .283, y por su parte el coeficiente de discriminación fue de .271, alcanzando un valor bueno para este índice. Los valores antes expresados reflejan una calidad psicométrica en las pruebas superior a los estándares mínimos deseados.

Tabla 1

Promedio general del rendimiento académico en el examen departamental de Morfología de la Segunda Lengua y Morfosintaxis de la Segunda Lengua

	N	Mínimo	Máximo	Media	Desviación
Morfología	1025	25.40	95.24	70.03	9.65
Morfosintaxis	851	24.24	91.80	53.81	8.62

Fuente: Construcción personal.

Continuando con los resultados, el análisis Anova permitió comparar entre profesores los resultados de rendimiento obtenidos por sus estudiantes, lo anterior sin aplicar ningún criterio particular asociado a los docentes que hiciera distinción entre los mismos. Parte de los principios que giran en torno a las pruebas estandarizadas establece el hecho de que los contenidos temáticos incluidos en la prueba deben ser iguales independientemente del profesor que imparta la clase, ya que la prueba tiene como eje rector la carta descriptiva que a su vez rige los contenidos que el docente trabaja durante el semestre. Sin embargo, los análisis estadísticos evidenciaron diferencias significativas ($p = 0.000$) en los resultados obtenidos por los estudiantes en el examen en razón del docente que impartió la asignatura (ver Tabla 2), estas variaciones van desde una media de 65.04 en la calificación del rendimiento en la prueba hasta un 74.27, este comportamiento se observa en la Tabla 3.

Tabla 2

Relación entre rendimiento académico y el docente

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	14501.48	12	1208.45	10.44	.000
Dentro de grupos	117127.62	1012	115.73		

Fuente: Construcción personal.

Tabla 3

Promedios generales por docente

Docente	Calificación			
	Media	Desviación estándar	Máximo	Mínimo
DA	72.70	11.16	88.89	38.10
DB	68.02	10.80	85.71	44.44
DC	68.60	9.56	80.95	42.86
DE	72.89	9.42	85.71	25.40
DF	68.68	9.26	85.71	44.44
DG	71.93	9.26	90.48	44.44
DH	65.04	7.91	82.54	47.62
DI	69.25	11.78	92.06	44.44
DJ	65.66	9.29	82.54	41.27
DK	68.37	11.66	94.00	36.51
DL	71.47	9.72	95.24	47.62
DM	74.27	7.65	85.71	58.73
DN	73.62	7.53	85.71	58.73

Fuente: Construcción personal.

Con los datos anteriores se evidencia que la variable docente influye en los resultados que se obtienen en la prueba. En las cuatro sedes se observó esta variación en relación al docente, para ejemplificar esto, se analizan a continuación los resultados en la sede Mexicali para el reciente ciclo escolar 2023-2. Se aplicó el examen departamental de Morfología de la Segunda Lengua a todos los grupos de primer semestre (cuatro), cada grupo tenía asignado a un docente diferente. El promedio general que se obtuvo en el examen fue de 65 puntos en una escala de 0 a 100, el mínimo aprobatorio para todas las asignaturas del plan de estudio es 60, por lo cual este resultado promedio es una calificación aprobatoria baja, apenas por encima del mínimo necesario.

Realizando un análisis comparativo entre el rendimiento obtenido en la prueba estandarizada departamental y el rendimiento en la asignatura como resultado final del curso (actas de evaluación final) se observó un índice de homogeneidad con un promedio de 16 puntos de diferencia, con excepción de un grupo cuya diferencia entre un rendimiento y otro fue solo de 7 puntos. En la mayoría de los grupos no todos los alumnos presentaron el examen, lo anterior se puede observar en la Tabla 4.

Tabla 4
Resultados del examen departamental de Morfología. Mexicali

Grupo	Número de alumnos	Número de alumnos que presentaron el examen	Media del examen %	Media de evaluación final	Diferencia entre calificación del examen y calificación final	
1	100	30	24	58	75	17
2	101	31	26	69	76	7
3	102	33	31	64	85	21
4	103	31	27	68	87	19

Fuente: Construcción personal.

En la Tabla 4 también se observa que hay una variación entre la cantidad de alumnos listados en el grupo y la cantidad de alumnos que realizaron el examen, esto se debe en parte a que algunos estudiantes pueden exentar dicho examen o bien no lo presentan de manera deliberada, ya que el examen departamental tiene un peso porcentual específico para la calificación final en la asignatura, siendo así posible acreditar la materia sin necesidad de presentar dicho examen.

Hay que considerar que la evaluación final del curso está comprendida por una variedad de actividades y tareas, que no se limitan a una sola evaluación o prueba, por lo que, en parte, esto explica la variación entre los resultados obtenidos. Sin embargo, como forma de analizar y comprobar este comportamiento, se realizó un ejercicio de la correlación de Pearson entre los resultados del examen departamental

y las calificaciones de los estudiantes en la asignatura, lo anterior dio como resultado una correlación positiva moderada (Díaz, 2021) de .61, lo que significa que existe una relación considerable entre el rendimiento obtenido en la prueba departamental y el obtenido en la asignatura a lo largo del curso.

Por otra parte, analizando el caso de la asignatura de Morfosintaxis de la Segunda Lengua, en el mismo ciclo escolar 2023-2 se aplicó el examen departamental a todos los grupos de segundo semestre (tres), y de igual forma cada grupo tenía un docente diferente. El promedio general que se obtuvo en el examen fue de 48 puntos en una escala de 0 a 100, lo cual indica una calificación insuficiente, no llegando ni al mínimo aprobatorio. Los promedios generales obtenidos por los grupos tanto en la prueba departamental como en la evaluación final del curso fueron homogéneos (ver Tabla 5), sin embargo, existió una diferencia más amplia entre ambos resultados, ya que el promedio de la diferencia de un rendimiento y otro fue de 36 puntos, casi el doble de la presente en el examen de Morfología de la Segunda Lengua.

Tabla 5

Resultados del examen departamental de Morfosintaxis. Mexicali

	Grupo	Número de alumnos	Número de alumnos que presentaron el examen	Media del examen %	Media de evaluación final	Diferencia entre calificación del examen y calificación final
1	104	14	13	49	93	44
2	105	13	12	49	80	31
3	106	17	16	46	80	34

Fuente: Construcción personal.

Nuevamente se realizó un análisis de la correlación de Pearson entre los resultados del examen departamental de Morfosintaxis y las calificaciones finales de los estudiantes en la asignatura. Lo anterior dio como resultado que dos grupos obtuvieron una correlación positiva moderada (.59 y .50 respectivamente), y un tercero una correlación positiva débil (.31).

Los resultados anteriores reflejan una diferencia significativa entre el nivel de dominio de conocimiento que el estudiante presenta objetivamente en la prueba departamental y el que el docente juzga como resultado de los diferentes procesos evaluativos que implementa a lo largo del curso, si bien no necesariamente deben ser iguales, pero se esperaría que las diferencias entre un rendimiento y otro fueran cercanas, ya que se parte del hecho de que el contenido que evalúa el docente durante el semestre es el mismo que se evalúa en la prueba departamental para las dos asignaturas en cuestión.

Continuando con los estudios realizados, se observó que el género no guardó relación con el rendimiento que los estudiantes obtuvieron en la prueba. La prueba de muestras independientes confirma este hecho, es decir, que el género no influye en el rendimiento obtenido en ninguna de las pruebas departamentales, ya que se obtuvieron valores de significancia $p = .416$ y $p = .532$ para Morfología y Morfosintaxis respectivamente. Los valores medios para ambos géneros pueden verse en las tablas 6 y 7.

Tabla 6

Prueba de muestras independientes: género y rendimiento en la prueba de Morfología

	N	Media	Desviación
Femenino	694	70.182	9.344
Masculino	331	69.895	9.963

Fuente: Construcción personal.

Tabla 7

Prueba de muestras independientes: género y rendimiento en la prueba de Morfosintaxis

	N	Media	Desviación
Femenino	588	54.596	8.437
Masculino	263	53.029	8.816

Fuente: Construcción personal.

Otro factor que se estudió respecto a su influencia en el rendimiento trataba sobre el formato de la prueba; en sus inicios y casi en su mayoría, el examen se aplicó en formato físico, sin embargo, en el último ciclo escolar dos sedes aplicaron el examen empleado un *software* diseño exprofeso para los exámenes departamentales ($n = 355$), por lo cual se hizo un análisis para comparar el rendimiento en la prueba teniendo como referencia el formato en que se realizó; el análisis Anova mostró que en efecto existió una influencia significativa ($p = .022$) en los estudiantes. Tomando en cuenta que la totalidad de la población fue de 1,025 estudiantes, una muestra representativa de esta población con un nivel de confianza del 95% sería de 280, por lo que contar con 355 alumnos que aplicaron la prueba con el *software* puede dar un margen de probabilidad alto sobre la influencia del formato de la prueba en el rendimiento de la misma.

DISCUSIÓN Y CONCLUSIONES

La evaluación estandarizada del aprendizaje brinda una posibilidad lo suficientemente confiable y válida para determinar el dominio de conocimientos que un estudiante posee al momento de ejecutar una prueba, dado que, como parte de la política edu-

cativa de la UABC, el interés de implementar este tipo de pruebas radica en promover el uso de métodos evaluativos que mejoren la calidad educativa y ofrezcan un insumo informativo sobre el cual tomar decisiones más oportunas.

En este trabajo se presentó un análisis del rendimiento académico logrado en dos pruebas estandarizadas que se aplican desde el año 2017, y en ese tenor se analizaron los factores internos que se considera que influyen en el rendimiento de las mismas. A partir del análisis realizado a las diferentes variables de estudio se observó que dos de los tres atributos o factores posibles de influencia fueron significativos en el rendimiento de las pruebas departamentales. Primero, la variable de formato de la prueba, se observó que esta tuvo un impacto significativo en el rendimiento alcanzado por los estudiantes en la prueba, aunque este análisis se realizó considerando a solo dos de las cuatro sedes de la FI que aplican la prueba departamental, sin embargo la cantidad de alumnos que participaron en la prueba utilizando el *software* especializado son una muestra estadísticamente representativa, por lo que este resultado no puede ser ignorado o despreciado directamente.

En este mismo tenor, en relación a este factor de influencia en el rendimiento académico, institucionalmente en las próximas aplicaciones de esta prueba las otras dos sedes la aplicarán también con el *software* especializado, de tal forma que se vuelva a hacer el análisis con la representatividad total de la comunidad estudiantil, y poder estimar con mayor certeza la influencia el rendimiento de la prueba con base en el formato de la misma.

Referente a este mismo factor que intervino positivamente en el rendimiento, una posible explicación puede ser atribuida a los beneficios que tiene el uso de tecnología, por ejemplo, las interfaces gráficas que permiten representar los textos en un tamaño y formato más agradables, una interacción con el usuario más natural al navegar entre las páginas que componen la prueba, la facilidad con la que se puede cambiar una respuesta, entre algunos otros aspectos.

Ahora bien, en relación a la variable género del sustentante, después de aplicar el análisis correspondiente se encontró que no existe influencia significativa en el rendimiento de la prueba departamental, este hecho de cierta manera aporta validez a la prueba, ya que las pruebas estandarizadas tienen entre otros propósitos el de garantizar una igualdad de probabilidades de acreditar o no la prueba, poniendo énfasis en los procedimientos metodológicos que rigen la construcción de la prueba con el fin de que la mayor variación posible de rendimiento en la prueba se debe a factores externos y no a aquellos asociados a la prueba en sí, de forma que la ejecución y el rendimiento radiquen en el dominio de conocimiento que el sustentante posee al momento de realizar la prueba.

Por último, abordando la tercera variable de estudio, fue la influencia del docente en el rendimiento académico, misma que resultó ser la variable con mayor influencia

en el resultado del rendimiento; los resultados demostraron que los estudiantes obtuvieron diferencias significativas en razón al docente que les impartió la asignatura.

El hecho anterior puede significar que las diferencias existentes entre las estrategias de enseñanza y aprendizaje, las competencias profesionales y demás factores asociados a la práctica docente entre un profesor y otro inciden directamente en el rendimiento académico de los estudiantes. Considerando lo anterior desde una perspectiva educativa, los resultados obtenidos también ponen de manifiesto que la elección del docente que imparta la asignatura debe ser un proceso de seria consideración, mismo que debe tener como objetivo asignar a los docentes más indicados, con mejor grado de habilitación docente o competencia para dar el curso.

Claro está, la conclusión anterior no es absoluta ni es la única explicación, es por ello que como trabajo futuro será necesario realizar estudios más detallados, de enfoque mixto, empíricos y longitudinales, que permitan dar pauta de manera más específica de las diferencias existentes a nivel pedagógico entre los docentes, mismas que pudieran dar el fundamento sobre el cual estriba la variación en el rendimiento estudiantil.

Sin embargo, a pesar de ser necesario aplicar otro tipo de estudios, esta investigación sí encontró la relación existente de influencia del docente sobre el rendimiento académico, por lo que esto es un insumo suficiente para que la institución educativa considere importante implementar procedimientos más objetivos al momento de asignar docentes para que impartan estas asignaturas, así como prestar atención a la puesta en práctica de programas de capacitación para mejorar la práctica docente.

Continuando con los hallazgos, en relación a los promedios generales obtenidos en la prueba se observó un rendimiento más favorable a nivel general en la prueba departamental de Morfología, logrando un promedio de 70 sobre 100; considerando que las pruebas departamentales comprenden todo el universo de conocimientos de la asignatura, se puede asumir que este promedio en la ponderación de la prueba se clasifica como un rendimiento mínimo satisfactorio.

Por su parte, el rendimiento general de la prueba departamental de Morfosintaxis se ubicó en 53.4 sobre 100, lo que se clasifica como un nivel de rendimiento insuficiente. Esto podría deberse en parte a la dificultad de la prueba, la cual se ubicó en .45, sin embargo, también habría que considerar que esta materia tiene una serie previa con la asignatura de Morfología. Estos valores promedio de rendimiento podrían ser el resultado de las deficiencias obtenidas en la anterior materia, mismas que se agregan a las generadas dentro del proceso educativo que el alumno desarrolla durante el curso de la asignatura de Morfosintaxis, dando como resultado un rendimiento significativamente menor.

Un aspecto importante a resaltar y que da sustento considerable a las interpretaciones que hasta este punto se han efectuado es que ambas pruebas demostraron

poseer una calidad psicométrica que se ubicó en los rangos de valores buenos hasta casi llegar a los excelentes. Este es un detalle importante, ya que se tiene el sustento matemático-estadístico que permite cumplir con la premisa de las evaluaciones estandarizadas, en las que se puedan atribuir que en su mayoría las variaciones en los resultados obtenidos en dichas pruebas están en razón del sujeto evaluado o factores concretos de intervención (como los tres analizados en esta investigación), pero no a la calidad técnica del instrumento o a los procesos con que esta fue construida.

Aún existen limitaciones que deben tomarse en cuenta al interpretar estos resultados y las conclusiones a las que se ha llegado, ya que para esta investigación solo se han tomado en cuenta tres variables de influencia del rendimiento académico, no obstante, existen algunos otros factores o predictores del rendimiento académico que deberán ser estudiados en un trabajo futuro, por ejemplo la formación inicial, actualización en el campo de estudio, socioeconómicos, emocionales, hábitos de estudio, entre algunos otros.

Así mismo, en este estudio se tiene como marco referencial de observación del rendimiento académico a la prueba departamental estandarizada, sin embargo, dicha medida no es la única que puede explicar el rendimiento escolar de los estudiantes. Por otra parte, haría falta realizar un estudio que permita dar razón de la posible validez predictiva que tenga el rendimiento obtenido en la asignatura de Morfología y el obtenido en Morfosintaxis, ya que son asignaturas seriadas, y con ello poder identificar los factores que pueden afectar tal rendimiento. También sería necesario contar con una mayor cantidad de información de aplicaciones de la prueba utilizando el *software* especializado para estos exámenes estandarizados con los cuales realizar estudios y corroborar el comportamiento en el rendimiento.

Siguiendo la línea anterior, para un trabajo futuro sería importante señalar la realización de un análisis curricular más específico del dominio que los alumnos presentan en la prueba, saber realmente sobre qué contenidos específicos de todo el universo de conocimientos que comprende la prueba se tiene un dominio, ya que algunos temas son más relevantes que otros, o bien son temas que inciden en mayor proporción a la concreción de las competencias esperadas a nivel de unidad temática o del curso en general. Identificar cómo están dosificados estos temas a nivel curricular permite interpretar de mejor manera la posición real que el alumno presenta en cuanto a dominio de conocimientos en toda la asignatura, y no solo limitarse a una ponderación, como la calificación.

Esta investigación ha contribuido a demostrar el grado de validez y confiabilidad que poseen las pruebas departamentales, señalando cómo con estas se puede obtener información que a luz de ciertos tratamientos estadístico-matemáticos evidencia áreas de oportunidad. Como se ha mencionado, una de las bondades de utilizar evaluaciones estandarizadas radica en contribuir a que las posibles variaciones en el rendimiento

de la prueba estén en razón de factores ajenos a la calidad técnica y psicométrica de la prueba, es decir, que un mal rendimiento en la prueba no sea motivo de una mala construcción de la misma, de una falta de validez de contenido o factores propios de la composición de la prueba.

La anterior conclusión fue demostrada con dos de los tres factores de influencia analizados respecto al rendimiento de los estudiantes en la prueba, mismos que demostraron diferencias estadísticas significativas, a saber, el docente que impartió la clase y aquellos alumnos que aplicaron la prueba en un formato tradicional y aquellos que lo aplicaron con un *software* diseñado expreso para la prueba.

Por otra parte, la investigación aportó información que hace un señalamiento al diseño de nuevos esquemas o marcos metodológicos que permitan homologar ciertas prácticas asociadas a los procesos de enseñanza-aprendizaje de los distintos docentes que imparten las asignaturas correspondientes a estas pruebas departamentales, ya que al ser el docente el principal factor de influencia en el rendimiento de la prueba se condiciona en gran manera el nivel de dominio de conocimiento que los estudiantes logran alcanzar. Fue posible apreciar que el rendimiento docente presenta deficiencias al contrastarlo con los resultados de la prueba, mismos que requieren de un análisis más profundo para determinar con exactitud qué factores inciden en el rendimiento de este profesorado en particular.

Así mismo sería importante hacer estudios futuros con cortes de datos asociados a aquellos alumnos que cursaron la materia de Morfología en modalidad virtual a raíz de la pandemia de la COVID-19 y que continuaron sus estudios de segundo semestre en modalidad presencial cursando ahora la materia de Morfosintaxis en otra modalidad. Esto permitiría observar si esta transición generó deficiencias en la adquisición de conocimientos necesarios para la concreción de otros conocimientos, tales como los que se dan en esta relación de seriación entre las asignaturas de Morfología y Morfosintaxis. Así mismo podrían hacerse valoraciones asociadas al tipo de estrategias pedagógicas implementadas por los docentes en ambientes virtuales en comparación con ambientes tradicionales o de tipo presencial, así como nuevos factores de influencia como podrían ser el tipo de interacción, los recursos tecnológicos de los que se dispone, entre otros.

Como se mencionó al inicio de este trabajo, profundizar y determinar de manera más exacta los factores que inciden en el rendimiento académico de los estudiantes sigue siendo un objeto de estudio que se encuentra a la orden del día en las instituciones de educación superior, son estos mismos factores los que en muchos casos guían o determinan las estrategias pedagógicas al interior de las aulas junto a las herramientas tecnológicas con las cuales se aplican. Lo anterior inherentemente conlleva la necesidad de contar con instrumentos de evaluación con validez y confiabilidad que supere los estándares mínimos de calidad.

En este sentido, las evaluaciones estandarizadas cobran relevancia ya que para las instituciones de educación superior contar con este tipo de instrumentos válidos y confiables, contribuye de manera significativa a la toma de decisiones en aras de mejorar la calidad educativa que ofrecen. Todo esto se traduce en mejores condiciones para poder competir como institución en satisfacer la demanda social cada vez más exigente en la formación de futuros profesionistas.

REFERENCIAS

- Aguilar-Salinas, W., y De las Fuentes, M. (2023). Examen colegiado y predictores de éxito en los estudiantes de álgebra lineal. *Bolema: Boletim de Educação Matemática*, 37(6), 797-822. <http://dx.doi.org/10.1590/1980-4415v37n76a20>
- Backhoff Escudero, E. (2018). Evaluación estandarizada del logro educativo: contribuciones y retos. *Revista Digital Universitaria*, 19(6). <https://doi.org/10.22201/codeic.16076079e.2018.v19n6.a3>
- Baladrón, J., Sánchez, F., Romeo, J., Curbelo, J., Villacampa, P., y Jiménez, P. (2018). Evolución de los parámetros dificultad y discriminación en el ejercicio de examen MIR. Análisis de las convocatorias de 2009 a 2017. *FEM: Revista de la Fundación Educación Médica*, 21(4), 181-193. <https://dx.doi.org/10.33588/fem.214.955>
- Benavides, J. (2015). Las pruebas estandarizadas como forma de medición del nivel de inglés en la educación colombiana. En J. Bastidas y G. Muñoz (eds.), *Fundamentos para el desarrollo profesional de los profesores de inglés* (pp. 29-56). Editorial Universitaria-Universidad de Nariño.
- Borja, G., Martínez, J., Barreno, S., y Haro, O. (2021). Factores asociados al rendimiento académico: un estudio de caso. *Revista Educare. Segunda Nueva Etapa 2.0*, 25(3), 54-77. <https://www.revistas.investigacion-upelipb.com/index.php/educare/article/view/1509/1516>
- Burkett, T. (2018). Norm-referenced testing and criterion-referenced testing. En *The TESOL Encyclopedia of English Language Teaching* (pp. 1-5). Wiley. <https://doi.org/10.1002/9781118784235.eelt0351>
- Carlos, E., Galván, L., y Ruiz, R. (2011). *Análisis de las propiedades psicométricas de un examen de admisión para aspirantes a ingeniería* [Ponencia]. XI Congreso Nacional de Investigación Educativa, Monterrey, Nuevo León, México. https://www.comie.org.mx/congreso/memoriaelectronica/v11/docs/area_01/1553.pdf
- Castrillón, O., Sarache, W., y Ruiz-Herrera, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación Universitaria*, 13(1), 93-102. <https://doi.org/10.4067/S0718-50062020000100093>
- Contreras, L. (2000). *Desarrollo y pilotaje de un examen de español para la educación primaria en Baja California* [Tesis de maestría, Universidad Autónoma de Baja California]. http://iide.ens.uabc.mx/documentos/divulgacion/tesis/MCE/1998/Luis_Angel_Contreras_Nino.pdf
- Contreras, L., y Backhoff, E. (2004). Metodología para elaborar exámenes criterios alineados al currículo. En S. Castañeda Figueiras (ed.), *Educación, aprendizaje y cognición. Teoría en la práctica* (pp. 298-323). Manual Moderno.
- Contreras, L. E., Fuentes, H. J., y Rodríguez, J. I. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación Universitaria*, 13(5), 233-246. <https://doi.org/10.4067/S0718-50062020000500233>
- Cortada, N. (2004). Teoría de respuesta al ítem: supuestos básicos. *Revista Evaluar*, 4(1), 95-110. <https://revistas.unc.edu.ar/index.php/revaluar/article/view/600>
- Demarchi, G. (2020). La evaluación desde las pruebas estandarizadas en la educación en Latinoamérica. *Contexto Revista de Investigación en Administración, Contabilidad, Economía y Sociedad*, 8(13), 107-133. <https://ojs.tdea.edu.co/index.php/encontexto/article/view/716/887>
- Dhal, P. (2021). Evaluation of school standards. *Research Tracks*, 3(2), 243-246. <https://www.ycjournal.net/ResearchTracks/ResearchDocuments/Evaluation636280255608553750.pdf>

- Díaz, J. (2021). Correlación entre las pruebas PCR y antígeno y el contagio por COVID-19 en Colombia. *Revista Repertorio de Medicina y Cirugía*, 30, 35-40. <https://revistas.fucsalud.edu.co/index.php/repertorio/article/view/1207/1580>
- Fandiño, Y. (2017). Formación y desarrollo docente en lenguas extranjeras: revisión documental de modelos, perspectivas y políticas. *Revista Iberoamericana de Educación Superior*, 8(22), 122-143. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-28722017000200122
- Fernández, M., Alcaraz, N., y Sola, M. (2017). Evaluación y pruebas estandarizadas: una reflexión sobre el sentido, utilidad y efectos de estas pruebas en el campo educativo. *Revista Iberoamericana de Evaluación Educativa*, 10(1), 51-67. <https://doi.org/10.15366/riee2017.10.1.003>
- Ferreira, M., y Backhoff, E. (2016). Validez del generador automático de ítems del Examen de Competencias Básicas (Excoba). *Relieve-Revista Electrónica de Investigación y Evaluación Educativa*, 22(1), 1-16. <https://www.redalyc.org/pdf/916/91649056016.pdf>
- Fiallos, G. (2021). La correlación de Pearson y el proceso de regresión por el método de mínimos cuadrados. *Ciencia Latina Revista Científica Multidisciplinar*, 5(3), 2491-2509. https://doi.org/10.37811/cl_rcm.v5i3.466
- Garbanzo, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista Educación*, 31(1), 43-63. <https://doi.org/10.15517/revedu.v31i1.1252>
- Grimaldo, M., y Manzanares-Medina, E. (2022). Variables intervinientes en el rendimiento académico en ingresantes de una universidad privada de Lima. *Revista Electrónica Educare*, 27(1), 1-14. <https://doi.org/10.15359/ree.27-1.14283>
- Guambuquete, C., Sánchez, C., Castro, M., Espinosa, C., y Jordan, C. (2023). Factores asociados al rendimiento académico: un análisis mediante regresión logística multivariante en estudiantes del primer nivel del Instituto Superior Tecnológico Tres de Marzo de la Provincia Bolívar, Ecuador. *Dominio de las Ciencias*, 9(3), 570-589. <https://dominiodelasciencias.com/ojs/index.php/es/article/view/3460>
- Gutiérrez-Monsalve, J. A., Garzón, J., y Segura-Cardona, A. M. (2021). Factores asociados al rendimiento académico en estudiantes universitarios. *Formación Universitaria*, 14(1), 13-24. <https://doi.org/10.4067/S0718-50062021000100013>
- Gutiérrez, J., y Acuña, L. (2023). Evaluación estandarizada del aprendizaje en la educación superior: un estudio de caso en México. *Revista Electrónica Educare*, 27(3), 1-19. <https://doi.org/10.15359/ree.27-3.17218>
- Heredia, A. (2009). Dos formas diferenciadas de evaluación didáctica: evaluación normativa para seleccionar a los alumnos y evaluación criterial para el dominio del conocimiento básico. *Bordón: Revista de Pedagogía*, 61(4), 39-48. <https://dialnet.unirioja.es/descarga/articulo/3109891.pdf>
- Hernández, M., Ramírez, É., y Gamboa, S. (2018). La implementación de una evaluación estandarizada en una institución de educación superior. *Innovación Educativa*, 18(76), 149-170. <https://dialnet.unirioja.es/servlet/articulo?codigo=6791099>
- Jiménez, M., Rodríguez, C., y Rey Paba, L. (2017). Standardized test results: An opportunity for English program improvement. *HOW*, 24(2), 121-140. <https://doi.org/10.19183/how.24.2.335>
- Masrek, M. N., y Zainol, N. Z. M. (2015). The relationship between knowledge conversion abilities and academic performance. *Procedia - Social and Behavioral Sciences*, 174, 3603-3610. <https://doi.org/10.1016/j.sbspro.2015.01.1078>
- Medina, J., Ramírez Díaz, M. H., y Miranda, I. (2019). Validez y confiabilidad de un test en línea sobre los fenómenos de reflexión y refracción del sonido. *Apertura*, 11(2), 104-121. <http://dx.doi.org/10.32870/Ap.v11n2.1622>
- Mendoza, A., y Herrera, R. (2013). *Propuesta para la predicción del rendimiento académico de los estudiantes de la Universidad del Atlántico, basado en la aplicación del análisis discriminante* [Ponencia]. Encuentro Internacional de Educación en Ingeniería, Cartagena de Indias, Colombia. <https://acofipapers.org/index.php/eici/article/view/1442>
- Ministerio de Educación Pública (2023). *Marco de referencia prueba estandarizada 2023*. https://dgec.mep.go.cr/wp-content/uploads/2024/02/marco_de_referencia_marzo_2023.pdf

- Muñiz, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo: Revista del Colegio Oficial de Psicólogos*, 31(1), 57-66. <https://www.redalyc.org/articulo.oa?id=77812441006>
- Neto, M., y Losada, A. (2021). Las capacidades y esfuerzos de los estudiantes como uno de los principales predictores del rendimiento académico. Caso del Instituto Superior de Ciencias de la Educación de Luanda, Angola. *Revista Angolana de Ciencias*, 3(1), 5-28. <https://doi.org/10.54580/R0301.02>
- Nitko, A. (1994). *A model for curriculum-driven criterion-referenced and norm referenced national examination for certification and selection of students* [Ponencia]. Second International Conference of Educational Evaluation and Assessment, Association for the Studies of Educational Evaluation in Southern Africa (ASEESA), Pretoria, Sudáfrica. <https://eric.ed.gov/?id=ED377200>
- Rico, A., y Gaytán, N. (2022). Modelos predictivos del rendimiento académico a partir de características de estudiantes de ingeniería. *IE Revista de Investigación Educativa de la REDIECH*, 13, e1426. https://doi.org/10.33010/ie_rie_rediech.v13i0.1426
- Rodríguez, S., Fita, E., y Torrado, M. (2004). El rendimiento académico en la transición secundaria-universidad. *Revista Educación*, 33(4), 391-414. <https://www.educacionyfp.gob.es/revista-de-educacion/numeros-revista-educacion/numeros-anteriores/2004/re334/re334-22.html>
- Soza, S. (2021). Factores asociados a la calidad del rendimiento académico de estudiantes en la educación superior. *Revista Ciencias de la Salud y Educación Médica*, 3(3), 36-43. <https://revistasnicaragua.cnu.edu.ni/index.php/rcsem/article/view/7374>
- Torres-Zapata, Á. E., Pérez-Jaimes, A. K., Lara-Gamboa, C. C., y Estrada-Reyes, C. U. (2022). Caracterización de los factores docentes en torno al índice de reprobación en universitarios. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(24), e310. <https://doi.org/10.23913/ride.v12i24.1123>
- Viladrich, C., Angulo-Brunet, A., y Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología*, 33(3), 755-782. <http://dx.doi.org/10.6018/analesps.33.3.268401>
- Yang, Q. (2023). Should standardization tests be used to assess student ability? *Journal of Education, Humanities and Social Sciences*, 8, 319-326. <https://doi.org/10.54097/ehss.v8i.4268>

Cómo citar este artículo:

Gutiérrez Benítez, J. G., Casillas Domínguez, J. M., y Olguin Jiménez, K. (2024). Rendimiento académico y pruebas departamentales estandarizadas: un estudio en educación superior. *IE Revista de Investigación Educativa de la REDIECH*, 15, e2161. https://doi.org/10.33010/ie_rie_rediech.v15i0.2161



Todos los contenidos de *IE Revista de Investigación Educativa de la REDIECH* se publican bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional, y pueden ser usados gratuitamente para fines no comerciales, dando los créditos a los autores y a la revista, como lo establece la licencia.