

Evaluación de la Eficacia del Examen de Admisión: Un Estudio Empírico sobre el Área de Pensamiento Lógico

Evaluating the Effectiveness of the Admission Test: An Empirical Study in the Area of Logical Thinking

 **Jair Arody del Valle López**

Universidad de Montemorelos, México

jdelvalle@um.edu.mx

 **Sem Darío Barba Amaya**

Universidad de Montemorelos, México

sem@um.edu.mx

Cómo citar/How to cite

del Valle López, jair A., & Barba Amaya, S. D. Evaluación de la eficacia del examen de admisión: Un estudio empírico sobre el área de pensamiento lógico. *Unaciencia Revista De Estudios E Investigaciones*, 17(33), 97-123. <https://doi.org/10.35997/unaciencia.v17i33.806>

Resumen

Las instituciones educativas enfrentan una creciente demanda, lo que ha llevado a la implementación de exámenes de admisión como filtro principal para seleccionar a los aspirantes con mayor potencial de éxito académico. Sin embargo, existe debate sobre si estos exámenes son adecuados para evaluar el potencial de los aspirantes en áreas específicas del conocimiento. Por ello, es necesario investigar si este tipo de exámenes están bien diseñados y si discriminan adecuadamente a los aspirantes. El propósito de este estudio fue analizar la efectividad del examen de admisión en una universidad del noreste



de México, específicamente en el área de pensamiento lógico. La investigación se desarrolló en tres fases: desarrollo de un banco de preguntas, creación de un entorno controlado para la prueba piloto, y evaluación del desempeño y validez del examen utilizando diversas técnicas. El análisis mostró que la mayoría de los ítems del examen tienen una alta capacidad discriminativa. Sin embargo, se identificaron ítems que necesitan revisión. El coeficiente de discriminación biserial indicó un buen nivel de discriminación, y el grado de dificultad general del examen se inclinó hacia el nivel medianamente difícil. El análisis de distracción validó que las opciones de respuesta cumplen su función. El examen de pensamiento lógico demuestra un enfoque metodológico sólido y una implementación cuidadosa. No obstante, se recomienda una revisión continua de los ítems con baja discriminación y opciones de respuesta ineficaces para mejorar la fiabilidad y validez del examen, consolidando una evaluación precisa y justa de las habilidades en el pensamiento lógico de los aspirantes.

Palabras clave: condiciones de admisión, construcción de pruebas, evaluación de los estudiantes, exámenes de ingreso.

Abstract

Educational institutions face a growing demand, which has led to the implementation of entrance exams as the main filter to select applicants with the greatest potential for academic success. However, there is debate as to whether these exams are adequate to evaluate the potential of applicants in specific areas of knowledge. Therefore, it is necessary to investigate whether such tests are well designed and whether they adequately discriminate between applicants. The purpose of this study was to analyze the effectiveness of the admission test at a university in northeastern Mexico, specifically in logical thinking. The research was developed in three phases: development of a question bank, creation of a controlled environment for the pilot test, and evaluation of the performance and validity of the test using various techniques. The analysis showed that most of the test items have a high discriminative ability. However, items were identified as needing revision. The biserial discrimination coefficient indicated a good level of discrimination, and the overall degree of difficulty of the test leaned toward the moderately difficult level. The distractor analysis validated that the response options fulfill their function. The logical thinking test demonstrates a sound methodological approach and careful implementation. However, a continuous revision of items with low discrimination and ineffective response options is recommended to improve the reliability and validity of the test, consolidating an accurate and fair assessment of the applicants' logical thinking abilities.

Key Words: admission conditions, entrance exams, student evaluation, test construction.



1. INTRODUCCIÓN

Actualmente las instituciones educativas de nivel superior no son suficientes en cantidad para cubrir la demanda. Derivado de esta escasez, surge la necesidad de filtrar a los aspirantes, eligiendo, mediante diversas técnicas, a aquellos con mayor potencial de éxito académico.

¿Cuál es el filtro favorito de las instituciones educativas hoy en día? Indiscutiblemente, los exámenes de admisión. Estos se han convertido en el filtro predilecto por su facilidad de implementación y la sencillez en su proceso. Por ello, Busche et al. (2020) subrayan la importancia de los exámenes de admisión para evaluar la preparación académica y prever el éxito de los estudiantes.

Debido a la trascendencia del examen para el futuro de vida de un aspirante y la facilidad de implementación de un examen de admisión, surge el siguiente cuestionamiento general: ¿el examen de admisión es adecuado para evaluar a un aspirante en una determinada área del saber?

Marco referencial

La Organización para la Cooperación y el Desarrollo Económico (OCDE) ha estado trabajando para lograr lo que podría ser la panacea de los evaluadores: crear una prueba que no se limite simplemente a una dimensión técnica, sino que desemboque en la construcción de un instrumento de evaluación realmente eficaz. Con este propósito en mente y añadiendo la necesidad de identificar a los aspirantes que mejor se desempeñarían en sus estudios (capacidad innata), las instituciones educativas, en 1926, recibieron con entusiasmo la prueba Scholastic's Aptitude Test (SAT), una prueba estandarizada que puede ser aplicada en cualquier lugar y momento.

En 1959, como alternativa al SAT, se desarrolló otra prueba llamada American Collage Testing (ACT). A diferencia del SAT, que mide la capacidad innata del aspirante, el ACT fue creado principalmente para medir el conocimiento adquirido por los aspirantes en distintos niveles escolares.

Es importante mencionar que los resultados de las pruebas estandarizadas como son el SAT y el ACT en ninguna manera deben ser tomadas como infalibles (Zwick, 2012); una razón de gran peso que apoya esto es la asociación que existe entre los puntajes de las pruebas de admisión y el nivel socioeconómico de los aspirantes; esto es debido a que el contenido con el que se hacen estas pruebas está basado, principalmente, en el material impartido en escuelas con estudiantes de un nivel socioeconómico alto. Evidentemente, esto deja fuera el material al cual los alumnos de bajos recursos tienen acceso.



Por otro lado, según Geiser (2016), lo anterior no tiene un buen fundamento y es muy necesaria una reevaluación, ya que una investigación realizada por la Universidad de California ha demostrado que el sistema de admisión que está basado en las calificaciones obtenidas en niveles escolares previos y pruebas basadas en estándares pueden predecir el éxito de un aspirante de igual manera como lo haría el SAT o el ACT. Esta investigación arroja datos muy valiosos y de suma importancia, pues brinda la oportunidad para que aquellas personas que son de bajos recursos y pertenecientes a las minorías no sean descartadas para ingresar en una universidad. También recalca de forma contundente el invaluable concepto de que el estudio constante y el esfuerzo valen más que la capacidad innata.

En México, antes de 1994, no existía una prueba oficial a la cual se pudiera recurrir y utilizar para aplicar pruebas de admisión en universidades públicas. Pero a partir de 1994, con la creación del Centro Nacional de Evaluación (CENEVAL), se ha intentado proveer de una prueba de admisión estándar que pueda ser usada por las universidades públicas; pero, aunque es una prueba con validez oficial, no se ha hecho obligatoria. Por lo tanto, al no existir una regulación impuesta para todas las instituciones, todas las entidades han tenido hasta estos momentos completa libertad para, según sean sus necesidades y posibilidades económicas, diseñar e implementar procesos y métodos que creen convenientes para cumplir con la demanda estudiantil, tratando de lograr que los procesos de selección sean lo más equitativos posibles y sean aplicados de manera eficiente y eficaz.

Lo anterior ha provocado la creación de una gran variedad de mecanismos no regulados de selección de aspirantes de los que no se tiene ningún registro y, al no haber una agencia reguladora oficial, existe una gran variedad de calidad en los distintos servicios para la recolección y aplicación de las pruebas de admisión (Benítez Lima et al., 2015).

Debido a esta variedad de situaciones de los exámenes de admisión, se encuentra que Díaz Barriga (2006) enfatiza que es indispensable analizar las pruebas actuales que son usadas para reclutar a los aspirantes y asegurar que efectivamente se logra el propósito de evaluar el nivel de conocimiento requerido. Por su parte, Woo et al. (2023) resaltan la importancia de estudiar la validez de los exámenes de admisión para asegurar que no se conviertan en barreras injustas para la diversidad en la educación. Greatrix, et al. (2021) mencionan que, para establecer la confianza en el uso de herramientas de selección en procesos educativos, son requeridos y esenciales estudios de validez; con esto, los candidatos deben sentir la legitimidad de las medidas, que de otro modo podrían considerarse como un obstáculo adicional.



Planteamiento del problema

Los exámenes de admisión, con el paso del tiempo, se han convertido en una tradición dentro de las instituciones educativas, sean éstas de renombre o no. Muchas veces los aspirantes, padres y maestros que ejercen o desean ingresar a una universidad comparten la creencia acerca de que la calidad de una escuela se puede medir simplemente con un examen de admisión.

Sin embargo, erróneamente se han adjudicado el derecho a catalogar, ya sea por ganar renombre o incrementar la matrícula, como una buena escuela a la que tiene un examen complicado para su acreditación y, por otro lado, las escuelas que tienen un examen catalogado de nivel medio o fácil entran en el círculo de las escuelas denominadas como de segunda clase. Definitivamente, basadas en un punto de vista que apoya las creencias antes mencionadas, las escuelas que no presentan ninguna restricción o examen para admitir a sus aspirantes son tildadas como escuelas que poseen una calidad inferior.

Utilizando la empatía, se debe ser franco con lo que podría ser la perspectiva de una institución educativa en cuanto a los exámenes de admisión; la necesidad de proveer un título académico que avale los conocimientos necesarios para un sinnúmero de puestos laborales o, simplemente, por ser un requisito laboral. Con esta última perspectiva en el ambiente educativo, se ha incrementado en gran medida la demanda por la educación superior hasta llegar al punto en que el espacio físico de las universidades es insuficiente para recibir a tantos estudiantes.

Este fenómeno provocó la necesidad de seleccionar a los aspirantes en lugar de permitir el libre ingreso a todos. Se hizo necesario y se requirió la elaboración de un proceso, no arbitrario, de selección para cualquier aspirante con deseos de ingresar a una universidad. Algunas instituciones optaron por aceptar a los aspirantes que tuvieron los mejores promedios en niveles académicos previos; otras, consideradas más exclusivas, se reservan el derecho de admisión, permitiendo solo a los aspirantes con mejores promedios (aquellos que han obtenido la puntuación mínima requerida por la institución) en sus exámenes de admisión.

El propósito principal que se le ha dado a un examen de admisión es seleccionar a los aspirantes que desean ingresar a una institución, pero para los departamentos de efectividad institucional tienen otro propósito; la información que se puede extraer de los resultados registrados en los exámenes permite analizar y evaluar el nivel de conocimiento adquirido previamente por los aspirantes y, basados en dichas evidencias, crear planes estratégicos que permitan a los mismos alcanzar, en el menor tiempo posible, el nivel de conocimiento requerido para el plan académico elegido.

Por otro lado, se observa el problema de que las calificaciones obtenidas por los



alumnos pueden variar considerablemente. Un examen mal formulado podría generar confusión, lo que inevitablemente alteraría los resultados finales. Otro aspecto es el efecto de repetición (tomar nuevamente la prueba), ya que el mismo tiene el potencial de impedir mediciones válidas (Scharfen et al., 2018).

Es así como las decisiones que se toman en la admisión de estudiantes son de gran relevancia para los estudiantes, las escuelas y la sociedad en general. Por ello es fundamental que los métodos de selección en la educación sean válidos (Lievens, et al., 2022).

Con estos pensamientos y acciones que se han presentado hasta este momento, es necesario llevar a cabo una investigación, en un ambiente controlado, que permita observar más a profundidad el contexto de los exámenes de admisión para que se pueda asegurar que: (a) un examen ha sido elaborado correctamente, (b) garantizar que un alto porcentaje de aquellos aspirantes que lo aprueban realmente son los que tienen el mayor conocimiento adquirido hasta ese punto de su educación y (c) garantizar una correcta valoración para aquellos aspirantes que no aprueban y poder generar un plan óptimo de capacitación para lograr su nivelación en el menor tiempo posible.

Esta investigación propone analizar el examen en el área del pensamiento lógico. Entonces, ¿el examen de pensamiento lógico en una universidad del noreste de México está correctamente diseñado?

Los exámenes de admisión, con el paso del tiempo, se han convertido en una tradición dentro de las instituciones educativas, sean éstas de renombre o no. Muchas veces los aspirantes, padres y maestros que ejercen o desean ingresar a una universidad comparten la creencia acerca de que la calidad de una escuela se puede medir simplemente con un examen de admisión.

Sin embargo, erróneamente se han adjudicado el derecho a catalogar, ya sea por ganar renombre o incrementar la matrícula, como una buena escuela a la que tiene un examen complicado para su acreditación y, por otro lado, las escuelas que tienen un examen catalogado de nivel medio o fácil entran en el círculo de las escuelas denominadas como de segunda clase. Definitivamente, basadas en un punto de vista que apoya las creencias antes mencionadas, las escuelas que no presentan ninguna restricción o examen para admitir a sus aspirantes son tildadas como escuelas que poseen una calidad inferior.

Utilizando la empatía, se debe ser franco con lo que podría ser la perspectiva de una institución educativa en cuanto a los exámenes de admisión; la necesidad de proveer un título académico que avale los conocimientos necesarios para un sinnúmero de puestos laborales o, simplemente, por ser un requisito laboral. Con esta última perspectiva en el ambiente educativo, se ha incrementado en gran medida la demanda por la educación superior hasta llegar al punto en que el espacio físico de las universidades es insuficiente para recibir a



tantos estudiantes.

Este fenómeno provocó la necesidad de seleccionar a los aspirantes en lugar de permitir el libre ingreso a todos. Se hizo necesario y se requirió la elaboración de un proceso, no arbitrario, de selección para cualquier aspirante con deseos de ingresar a una universidad. Algunas instituciones optaron por aceptar a los aspirantes que tuvieron los mejores promedios en niveles académicos previos; otras, consideradas más exclusivas, se reservan el derecho de admisión, permitiendo solo a los aspirantes con mejores promedios (aquellos que han obtenido la puntuación mínima requerida por la institución) en sus exámenes de admisión.

El propósito principal que se le ha dado a un examen de admisión es seleccionar a los aspirantes que desean ingresar a una institución, pero para los departamentos de efectividad institucional tienen otro propósito; la información que se puede extraer de los resultados registrados en los exámenes permite analizar y evaluar el nivel de conocimiento adquirido previamente por los aspirantes y, basados en dichas evidencias, crear planes estratégicos que permitan a los mismos alcanzar, en el menor tiempo posible, el nivel de conocimiento requerido para el plan académico elegido.

Por otro lado, se observa el problema de que las calificaciones obtenidas por los alumnos pueden variar considerablemente. Un examen mal formulado podría generar confusión, lo que inevitablemente alteraría los resultados finales. Otro aspecto es el efecto de repetición (tomar nuevamente la prueba), ya que el mismo tiene el potencial de impedir mediciones válidas (Scharfen et al., 2018).

Es así como las decisiones que se toman en la admisión de estudiantes son de gran relevancia para los estudiantes, las escuelas y la sociedad en general. Por ello es fundamental que los métodos de selección en la educación sean válidos (Lievens, et al., 2022).

Con estos pensamientos y acciones que se han presentado hasta este momento, es necesario llevar a cabo una investigación, en un ambiente controlado, que permita observar más a profundidad el contexto de los exámenes de admisión para que se pueda asegurar que: (a) un examen ha sido elaborado correctamente, (b) garantizar que un alto porcentaje de aquellos aspirantes que lo aprueban realmente son los que tienen el mayor conocimiento adquirido hasta ese punto de su educación y (c) garantizar una correcta valoración para aquellos aspirantes que no aprueban y poder generar un plan óptimo de capacitación para lograr su nivelación en el menor tiempo posible.

Esta investigación propone analizar el examen en el área del pensamiento lógico. Entonces, ¿el examen de pensamiento lógico en una universidad del noreste de México está correctamente diseñado?



Estado del arte

El deseo de contribuir con el desarrollo y mejora de la sociedad actual es la fuente de energía que ha impulsado muchas investigaciones, las cuales buscan encontrar de manera precisa qué es lo que realmente marca el rumbo en el éxito o fracaso de los estudiantes. Están tratando de descubrir cómo es que influyen distintos factores en el desempeño de los estudiantes, tales como: la demografía, el estilo de aprendizaje y las notas de enseñanza media. Los resultados obtenidos han confirmado que un factor aislado, por sí solo, es insuficiente para una predicción precisa sobre el éxito académico y las investigaciones respaldan que una combinación de variables, tales como una medición estandarizada, una evaluación escolar, una entrevista y una evaluación vocacional unidas en un solo instrumento, son mucho más precisas para encontrar a los estudiantes con las más altas probabilidades de concluir sus estudios exitosamente (Grob et al., 2015).

Con este contexto en mente, para Chaves et al. (2008, citados en Tapasco Alzate et al., 2016), el método de selección debe ser lo más transparente posible, teniendo una doble finalidad: (a) seleccionar a los aspirantes con el mejor elemento de juicio y (b) conocer de forma individual y general el perfil de ingreso de cada aspirante.

Por esto, Aguirre Román (2008) resalta con mucho fervor la falta de medidas en algunos de los métodos de admisión existentes, las cuales deberían ayudar a garantizar la equidad. Para evitar problemas de esta índole, el Sistema Único de Admisión (SUA) busca mejorar la equidad en la selección de los mejores aspirantes para la educación universitaria. Su objetivo principal es establecer un proceso óptimo de selección para encontrar a los aspirantes que tendrán éxito en los estudios de nivel superior, que sea al mismo tiempo más justo y equitativo, eliminando la capacidad socioeconómica de cada persona y su género como factores determinantes para impedir el acceso a la educación superior (Larroucau et al., 2015).

2. METODOLOGÍA

La metodología general de esta investigación tiene un diseño cuantitativo descriptivo, prospectivo, transversal exploratorio no experimental. La misma se desarrolló desde tres perspectivas: (a) la propuesta para el desarrollo de un banco de preguntas para el examen de pensamiento lógico, (b) abordar los requisitos necesarios para crear un ambiente o entorno controlado para garantizar la equidad y concentración de los participantes en la prueba piloto y, (c) implementar criterios para evaluar el desempeño y la validez del examen para los aspirantes universitarios de primer ingreso a una universidad del noreste de México ($N = 800$).



Propuesta del banco de preguntas para el examen de pensamiento lógico

Se realizó una búsqueda exhaustiva a través de diferentes medios, tales como libros, artículos y medios digitales donde se provee de ejercicios y problemas relacionados al área de pensamiento lógico. Todo ello con la finalidad de encontrar y recopilar estructuras, dimensiones, ejercicios e ideas que permitieran elaborar ítems de autoría propia.

Para la creación del banco de preguntas que abastecería al examen de admisión, se convocó la participación de cuatro expertos (dos del área de ingeniería en sistemas computacionales y dos del área de educación con especialidad en físico matemático). Se acordó entre los expertos analizar, categorizar y estructurar los ítems para que respondieran a tres áreas: secuencias lógicas, relaciones lógicas y transformaciones lógicas.

En relación con el área de secuencias lógicas, la intención fue elaborar y redactar ítems para evaluar la capacidad del aspirante para percibir patrones de relación/series entre números y letras; midiendo la habilidad del aspirante para organizar información de forma inductiva.

El área de relaciones lógicas tuvo como intención evaluar la capacidad del aspirante para extraer relaciones y hacer comparaciones basadas en reglas de similitud.

Para el área de transformaciones lógicas, la intención fue evaluar la capacidad del aspirante para usar correctamente las reglas de inferencia lógica; analizando una situación en particular y seleccionando la inferencia más apropiada.

Ambiente controlado

Previo al diseño de la eficacia del examen de pensamiento lógico, se llevó a cabo una aplicación piloto en papel con 20 individuos de primer semestre de nivel académico superior. El objetivo de esta prueba piloto fue obtener datos sobre el tiempo necesario para completar el examen, incluyendo: tiempo mínimo (t_{min}), tiempo máximo (t_{max}), media aritmética (\bar{t}), varianza (s^2), desviación estándar (s), asimetría (g_1) y curtosis (g_2).

Los 20 participantes fueron seleccionados por conveniencia y se les ofrecieron puntos extras en algunas materias, previo acuerdo con diversos docentes, en función del puntaje obtenido en el examen. Para la aplicación del examen, se aseguró la disponibilidad de un salón con energía eléctrica ininterrumpida y se designó un supervisor para la prueba.

Los estadísticos descriptivos obtenidos ($t_{min} = 25$, $t_{max} = 50$, $\bar{t} = 36.20$, $s^2 = 7.112$, $s = 50.589$, $g_1 = 0.306$, $g_2 = -0.263$) permitieron calcular el tiempo promedio por ítem (1.20 minutos, $\bar{t} / 30$) y una distribución normal de los datos (Shapiro-Wilk $_{(20)} = 0.963$, $p = .599$).



Con base en estos resultados descriptivos, para la fase de diseño y efectividad del examen, se determinó que el número máximo de ítems sería de 21 con un tiempo máximo de 25 minutos para responderlos.

Diseño de eficacia del examen en pensamiento lógico

Para la aplicación del examen de admisión, en el área de pensamiento lógico, se usó el banco de preguntas resultante de la propuesta inicial y se contó con la participación de ($n = 529$) aspirantes que hicieron admisión a diversos programas académicos de nivel universitario.

Para determinar el grado de dificultad de los ítems y la calidad de los distractores (opción múltiple de los ítems) se emplearon cuatro técnicas. A continuación, se describe cada una de ellas:

1. Índice de discriminación. Propuesto por Ebel y Frisbie (1986). Este índice mide el rendimiento de los aspirantes en el examen. Para hacerlo, se utiliza el índice de Pemberton. Este método propone dividir el grupo en tres: el primer grupo son el 27 % de aspirantes con el puntaje más alto; el segundo grupo es el 27 % de aspirantes con el puntaje más bajo y, el tercer grupo son el 46 % de aspirantes con el puntaje medio. El tercer grupo es descartado, solo se trabaja con el primero y el segundo grupo. Para calcular el índice es necesario sumar el total de respuestas correctas del primer grupo, lo mismo se hace con el segundo grupo. A continuación, se resta el total de la sumatoria del primer grupo al total de la sumatoria del segundo grupo. Posteriormente, según Del Valle López (2005), esa suma se divide entre el número de alumnos que hay en el grupo que tuvo más respuestas correctas. El resultado es el índice de discriminación (d_r) el cual es un número entre -1 y 1. La fórmula para su cálculo puede verse en la Ecuación 1, donde d_r es el índice de discriminación del ítem r ; $GE_{Aciertos}$ es el número de aciertos en el ítem r del 27 % de aquellos aspirantes con las puntuaciones más altas en el examen; $GNE_{Aciertos}$ es el número de aciertos en el ítem r del 27 % de aspirantes con las puntuaciones más bajas en el examen y, $n_{grupomasgrande}$ es el número de aspirantes en el grupo más numeroso de entre $GE_{Aciertos}$ o $GNE_{Aciertos}$.

$$d_r = \frac{GE_{Aciertos} - GNE_{Aciertos}}{n_{grupomasgrande}} \quad (1)$$

2. Coeficiente de discriminación biserial, propuesto por Glass y Stanley (1986). Este coeficiente mide la correlación existente entre dos variables cuando una de éstas es dicotómica (en este caso, aciertos y errores en las preguntas). Para calcular el coeficiente de discriminación biserial es necesario usar la Ecuación 2, donde es la media aritmética de las



puntuaciones totales de aquellos aspirantes que respondieron correctamente la pregunta; es la media aritmética de las puntuaciones totales de aquellos aspirantes que respondieron incorrectamente la pregunta; s_x es la desviación estándar de las puntuaciones totales; n_1 es el número de casos que respondieron correctamente la pregunta; n_0 es el número de casos que respondieron incorrectamente la pregunta y, n es igual a la suma entre n_1 y n_0 .

$$r_{pbis} = \frac{\bar{x}_1 - \bar{x}_0}{s_x} * \sqrt{\frac{n_1 n_0}{n(n-1)}} \quad (2)$$

3. Grado de dificultad, propuesto por Ebel y Frisbie (1986). Esta técnica mide la dificultad para contestar de manera correcta un ítem. La dificultad de un ítem se calcula dividiendo la cantidad de aspirantes que contestaron acertadamente el ítem entre la cantidad total de aspirantes que contestaron el ítem.

4. Análisis de distracción. Este análisis mide que las diversas respuestas en un examen efectivamente cumplen con su función de distracción. Para usarlo, se toman todas las respuestas que hay para un ítem y se cuenta la cantidad de aspirantes que eligieron cada respuesta. Lo que interesa observar para cada ítem es que cada una de sus opciones haya sido seleccionada por una cantidad considerable, de tal forma que pueda estar balanceada la frecuencia de cada una de las opciones de respuesta para la pregunta o ítem. En caso de que nadie elija una, o incluso algunas de las opciones de respuesta a la pregunta, eso es un indicador de que el poder de distracción de esa opción de respuesta debería ser seriamente considerado para su modificación o eliminación, dado que es nulo su poder de distracción.

3. RESULTADOS

El análisis de los resultados comienza por la elaboración y evaluación de la eficacia de un banco de preguntas diseñado para medir el pensamiento lógico en exámenes de admisión. Adicionalmente, se incluyen ejemplos ilustrativos de ítems representativos para proporcionar una visión clara de su formato y contenido. La eficacia del examen se evaluó mediante índices de discriminación y de dificultad, así como un análisis de distracción, asegurando así la validez y confiabilidad de la prueba.

Propuesta del banco de preguntas para el examen de pensamiento lógico

La recopilación total de ítems para evaluar el área de pensamiento lógico fue aproximadamente de 700 reactivos. Después del análisis de los expertos y de la selección



de propiedades que harían de las preguntas reactivos más prometedores para obtener buenos resultados en los exámenes de admisión, se redactaron 269 ítems, los cuales se estructuraron en tres áreas (secuencias lógicas, relaciones lógicas y transformaciones lógicas) y dimensionados tal como se puede observar en la Tabla 1 a la Tabla 3.

Tabla 1.

Área de secuencias lógicas: sus dimensiones y cantidad de ítems.

Nombre de la dimensión	Número de ítems
Series alfabéticas (independientes)	7
Series alfabéticas (parejas)	6
Series alfabéticas (par-impar)	6
Series numéricas (independientes)	8
Series numéricas (parejas)	6
Series numéricas (par-impar)	6
Series alfanuméricas (independientes)	6
Series alfanuméricas (parejas)	6
Matrices (posiciones)	9
Matrices (patrones-fila)	7
Matrices (cuadros mágicos)	5
Matrices (intercaladas-filas)	7
Matrices (aritméticas)	14

Tabla 2.

Área de relaciones lógicas: sus dimensiones y cantidad de ítems.

Nombre de la dimensión	Número de ítems
Analogías cuantitativas (bipartita)	10
Analogías cuantitativas (proporciones)	10
Analogías cuantitativas (consistencia interna)	13
Analogías cuantitativas (bipartita-fraccionarias)	12
Analogías cuantitativas (tripartita)	23

Tabla 3

Área de transformaciones lógicas: sus dimensiones y cantidad de ítems

Nombre de la dimensión	Número de ítems
Conmutación (aplicación)	22
Conmutación (general)	12
Conmutación (diagramas)	74



Por otro lado, para proveer una idea más específica de cómo son los ítems del examen, en la Figura 1 y Figura 2 se observan dos ítems que representan secuencias lógicas para que el aspirante identifique patrones de números o letras y así demostrar el grado de habilidad para organizar información inductivamente.

Figura 1.

Ejemplo de ítem para secuencias lógicas: dimensión series alfanuméricas (parejas).

Instrucción: Organiza la siguiente serie alfanumérica para identificar el elemento que complementa la secuencia lógica.

____ - _____, X17-S16, E16-A16, L16-I17, S17-Q19, Z19-Y22

- a) Q19-J18
- b) J19-X17
- c) J19-L16
- d) Q19-K17

Figura 2.

Ejemplo de ítem para secuencias lógicas: dimensión matrices (intercaladas-filas).

Instrucción: Organiza la siguiente serie alfanumérica para identificar el elemento que complementa la matriz de secuencias lógica.

F	9	G	H	5
7	8	E	6	I
5	F	9	G	H
-	-	-	-	-

- a) H 5 F 9 G
- b) I 6 8 E 7
- c) I 8 E 6 7
- d) I 7 8 E 6

En la Figura 3 y Figura 4, se observan dos ítems del examen que representan relaciones lógicas para que el aspirante pueda extraer relaciones y hacer comparaciones basadas en reglas de similitud.



Figura 3.

Ejemplo de ítem para relaciones lógicas: dimensión analogías cuantitativas (proporciones).

Instrucción: Extrae la relación entre los elementos indicados y selecciona la respuesta que cumpla las mismas reglas de similitud.

35 : 28 ::

- a) 120 : 102
- b) 90 : 81
- c) 60 : 48
- d) 150 : 117

Figura 4

Ejemplo de ítem para relaciones lógicas: dimensión analogías cuantitativas (tripartita)

Instrucción: Extrae la relación entre los elementos indicados y selecciona la respuesta que cumpla las mismas reglas de similitud.

98 : 196 : 2 ::

- a) 85 : 255 : 3
- b) 49 : 147 : 2
- c) 64 : 130 : 2
- d) 76 : 230 : 3

En la Figura 5, se observa un ítem del examen que representa relaciones lógicas para que el aspirante pueda implementar correctamente reglas de inferencia lógica, analizando una situación en particular y seleccionando la inferencia más apropiada.



Figura 5.

Ejemplo de ítem para transformaciones lógicas: dimensión conmutación (general).

Instrucción: ¿Cuál es el valor de la letra C?

4 = B
A x B = C x D
A x D = C
B = D x D
B + 1 = A

a) 10
b) 2
c) 4
d) 5

Diseño de eficacia del examen en pensamiento lógico

Como se mencionó en la sección de metodología, el diseño de la eficacia del examen en pensamiento lógico utilizó el índice de discriminación, el coeficiente de discriminación biserial, el grado de dificultad del examen y el análisis de distracción.

Índice de discriminación

Para determinar el índice de discriminación es necesario tomar la puntuación total del examen y evaluar las preguntas una por una. Tomando en cuenta que el examen y las preguntas están evaluando el área de pensamiento lógico, se entiende que un aspirante que obtiene una puntuación alta en el examen tiene las mismas probabilidades de contestar correctamente las preguntas. Se esperaría lo mismo en caso contrario, es decir, cuando un aspirante obtiene una calificación baja en el examen, tiene bajas probabilidades de contestar correctamente las preguntas.

Si el examen contiene reactivos con una media alta en el índice de discriminación, se podría asegurar que los resultados son confiables; existiendo una mayor probabilidad de diferenciar a las personas con altas y bajas calificaciones al obtener un índice de discriminación más grande.

Para poder interpretar taxonómicamente un ítem según el valor calculado por el índice de discriminación, se usó como guía la Tabla 4, propuesta por Ebel y Frisbie (1986).



Tabla 4.*Interpretación del índice de discriminación.*

Discriminación	Clasificación				
	Pésimo	Pobre	Regular	Bueno	Excelente
d_r	< -.01	.0 a .20	.21 a .29	.30 a .38	> .39

La Ecuación 3 es un ejemplo del proceso realizado para calcular el índice de discriminación. Se usó la información del ítem 206 del examen de pensamiento lógico: $GE_{Acertos} = 16$, $GNE_{Acertos} = 7$ y $n_{grupomasgrande} = 16$, resultando un índice de discriminación ($d_r = 0.5625$) con una clasificación de *excelente* ($dr > .39$).

$$d_r = \frac{16 - 7}{16} = 0.5625 \quad (3)$$

La Tabla 5 presenta los índices de discriminación que resultaron del total de los 269 ítems del área de pensamiento lógico. Se agruparon las preguntas que tienen el mismo índice de discriminación.

Tabla 5.*Índice de discriminación por ítem.*

Ítem	d_r
136	-1
108,139	-0.5
112	-0.33
50	-0.25
31	-0.14
22,30,32,56,75,77,81,90,109,113,115,123,164,169,172,183,186,193,256,279	0
27	0.14
36	0.17
43,79	0.2
16,17,35,71,94,383	0.25
29,92,104	0.29
5,73,96,100,134,160,173,203,226,240	0.33
25,39,116,125,167	0.4
37,58,72,80,101,117,120,138,178,209,211	0.5
206	0.56
24,28,65,129	0.57
7,12,15,21,53,63,82,177,185,224,370	0.6



26	0.62
11,34,38,49,57,60,78,89,95,99,122,132, 161,176,179,189,198,201,220,221,234,243,251,274,277,278,367,372,373	0.67
52,103	0.71
45,51,74,110,118,119,170,199,212,239,259,264,285,375	0.75
8,14,46,47,126,180,192,194,207,208,213,229,232,238,267,270,273,368	0.8
42,55,59,168,214,228	0.83
41,128,181,222,366	0.86
64,262	0.88
6,9,10,13,18,20,23,33,40,44,48,54,61,62,76,91,93,97,98,102,105,106, 107,111, 114,121,124,127,130,131,133,135,137,140,141,142,162,163,165,166,171,174, 175,182, 184,187,188,190,191,195,196,197,200,202,204,205,210,219,223,225, 227,230,231,233,235,236,237,241,242,244,245,246,247,248,249,250,252,253, 254,255,257,258,260,261,263,265,266,268,269,271,272,275,276,281,282,283, 284,286,287,364,365, 369,371,374,376,377,378,379,380,381,382	1

Como se puede observar en la Tabla 5, se tiene representación de todas las clasificaciones de la tabla de interpretación del índice de discriminación. Guiándose mediante la Tabla 4 se encuentra que, de las 269 preguntas, 30 entran en la clasificación de *pésimo y pobre*. Por su parte, nueve de las preguntas están entre el valor .21 y .29, por lo que entran en una clasificación de *regular*; 10 preguntas están entre el valor .30 y .38 así que están consideradas con una clasificación de *buen* índice, y se tienen 220 que se clasificarían como *excelentes*. Por lo tanto, se puede inferir que el índice de discriminación para el examen es bastante aceptable.

Coefficiente de discriminación biserial

Una forma de validar la discriminación de un instrumento es usando el coeficiente de discriminación biserial (r_{pbis}). Es recomendable usar el coeficiente de discriminación para corroborar la información obtenida mediante el índice de discriminación, dado que el coeficiente de discriminación toma en cuenta a todos los que presentaron el examen y no solo al 54 % de los aspirantes como lo hace el índice de discriminación.

El fin de usar el coeficiente de discriminación biserial es saber si las personas adecuadas son quienes obtienen las respuestas correctas. Otro fin es probar la eficacia de predicción en el reactivo.

Para poder interpretar taxonómicamente un ítem, según el valor calculado por el coeficiente de discriminación biserial, se usará como guía la Tabla 6, propuesta por Glass y Stanley (1986).



Tabla 6.

Interpretación del coeficiente de discriminación biserial.

Discriminación	Clasificación				
	Pésimo	Pobre	Regular	Bueno	Excelente
r_{pbis}	< .0	.0 a .14	.15 a .25	.26 a .35	>= .36

La Ecuación 4 es un ejemplo del proceso realizado para calcular el coeficiente de discriminación biserial. Se usó la información del ítem 206 del examen de pensamiento lógico: $x_1 = 8.16$, $x_0 = 5.42$, $s_x = 3.06$, $n_1 = 25$, $n_0 = 14$ y $n = 39$ con un coeficiente de discriminación biserial ($r_{pbis} = 0.4272$) clasificado como *excelente* ($r_{pbis} \geq .36$).

$$r_{pbis} = \frac{8.16 - 5.42}{3.06} * \sqrt{\frac{(25)(14)}{39(39-1)}} \quad (4)$$

La Tabla 7 presenta los coeficientes de discriminación biserial que resultaron del examen, agrupados en las cinco clasificaciones referidas en la Tabla 6 y ponderados por porcentaje.

Tabla 7.

Porcentaje del coeficiente de discriminación biserial por clasificación de los ítems.

Rango	Porcentaje	Ítem
Pésimo	5.57	22,31,50,75,77,108,109,112,123,136,139,169,172,256,279
Pobre	5.94	30,32,81,90,94,113,115,130,164,176,183,186,196,252,253,275
Regular	11.15	11,18,27,29,43,44,56,57,73,78,101,116,117,120,121,131,134,141,160,161,171,173,179,193,209,211,254,264,278,383
Bueno	14.49	5,7,9,17,35,36,38,39,55,58,62,63,72,80,92,98,99,100,102,107,111,132,137,138,142,170,178,188,195,198,199,213,221,224,237,240,249,250,251
Excelente	62.82	6,8,10,12,13,14,15,16,20,21,23,24,25,26,28,33,34,37,40,41,42,45,46,47,48,49,51,52,53,54,59,60,61,64,65,71,74,76,79,82,89,91,93,95,96,97,103,104,105,106,110,114,118,119,122,124,125,126,127,128,129,133,135,140,162,163,165,166,167,168,174,175,177,180,181,182,184,185,187,189,190,191,192,194,197,200,201,202,203,204,205,206,207,208,210,212,214,219,220,222,223,225,226,227,228,229,230,231,232,233,234,235,236,238,239,241,242,243,244,245,246,247,248,255,257,258,259,260,261,262,263,265,266,267,268,269,270,271,272,273,274,276,277,281,282,283,284,285,286,287,300,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,382



En la Tabla 7 se analiza que el 62.82 % del total de los ítems están dentro del rango *excelente*. Sumando los porcentajes obtenidos en los rangos *regular*, *bueno* y *excelente* se obtiene el 88.46 % del total de las preguntas, lo que resalta el buen nivel de evaluación de los aspirantes en el examen de admisión de pensamiento lógico.

La media aritmética del coeficiente de discriminación biserial ($r_{pbis} = .37$), indica que la discriminación, basados en la Tabla 6, es *excelente*.

Grado de dificultad del examen

La dificultad de un ítem p_r , es determinada por la cantidad de alumnos que responden correctamente el ítem. Por lo que, un resultado mayor indicaría un nivel menor de dificultad del ítem y, viceversa, un resultado menor indicaría un nivel mayor de dificultad del ítem.

El grado de dificultad de un ítem se calcula dividiendo el número de personas que contestó correctamente el ítem entre el número total de personas que contestó el ítem.

El resultado de esta ecuación debe ser un número que oscila entre 0 y 1. La fórmula para calcularlo se muestra en la Ecuación 5, donde p_r es el índice de dificultad del ítem r ; a_r es el número de aciertos en el ítem r y, n_r es el número de aciertos más el número de errores en el ítem r .

$$p_r = \frac{a_r}{n_r} \quad (5)$$

A la fórmula anterior se le ha realizado una pequeña modificación para poder interpretar el grado de dificultad de un ítem p_r como proporción directa. Para esto se le restará 1 a p_r y así se obtendrá la medida directa.

Para poder interpretar taxonómicamente un ítem según el valor calculado por el grado de dificultad, se usará como guía la Tabla 8, propuesta por Ebel y Frisbie (1986).

Tabla 8.

Interpretación del grado de dificultad.

Coeficiente	Clasificación				
	Fácil	Medianamente fácil	Dificultad media	Medianamente difícil	Difícil
p_r	.80 a 1	.60 a .79	.40 a .59	.20 a .39	.0 a .19



p_i	.0 a .19	.20 a .39	.40 a .59	.60 a .79	.80 a 1
-------	----------	-----------	-----------	-----------	---------

La Tabla 9 muestra los 269 ítem con su cálculo del grado de dificultad.

Tabla 9.

Grado de dificultad por ítem para el examen teórico.

Ítem	p_r	p_i	Ítem	p_r	p_i	Ítem	p_r	p_i	Ítem	p_r	p_i	Ítem	p_r	p_i	Ítem	p_r	p_i	
5	0.59	0.41	51	0.45	0.65	107	0.26	0.74	169	0.12	0.88	214	0.39	0.61	263	0.4	0.6	
6	0.52	0.48	52	0.57	0.43	108	0.22	0.78	170	0.46	0.54	219	0.41	0.59	264	0.43	0.57	
7	0.38	0.62	53	0.48	0.52	109	0.07	0.93	171	0.13	0.87	220	0.32	0.68	265	0.33	0.67	
8	0.37	0.63	54	0.27	0.73	110	0.33	0.67	172	0.16	0.84	221	0.35	0.65	266	0.3	0.7	
9	0.27	0.73	55	0.41	0.59	111	0.1	0.9	173	0.41	0.59	222	0.52	0.48	267	0.5	0.5	
10	0.08	0.92	56	0.75	0.25	112	0.26	0.74	174	0.4	0.6	223	0.46	0.54	268	0.19	0.81	
11	0.35	0.65	57	0.33	0.67	113	0.12	0.88	175	0.41	0.59	224	0.5	0.5	269	0.28	0.72	
12	0.5	0.5	58	0.62	0.38	114	0.23	0.77	176	0.46	0.54	225	0.24	0.76	270	0.38	0.62	
13	0.15	0.85	59	0.48	0.52	115	0.15	0.85	177	0.55	0.45	226	0.43	0.57	271	0.17	0.83	
14	0.4	0.6	60	0.41	0.59	116	0.65	0.35	178	0.31	0.69	227	0.38	0.62	272	0.08	0.92	
15	0.55	0.45	61	0.52	0.48	117	0.11	0.89	179	0.27	0.73	228	0.33	0.67	273	0.24	0.76	
16	0.6	0.4	62	0.41	0.59	118	0.31	0.69	180	0.34	0.66	229	0.52	0.48	274	0.22	0.78	
17	0.31	0.69	63	0.62	0.38	119	0.24	0.76	181	0.34	0.66	230	0.26	0.74	275	0.18	0.82	
18	0.24	0.76	64	0.43	0.57	120	0.24	0.76	182	0.29	0.71	231	0.15	0.85	276	0.35	0.65	
20	0.44	0.56	65	0.68	0.32	121	0.15	0.85	183	0.07	0.93	232	0.37	0.63	277	0.54	0.46	
21	0.6	0.4	71	0.88	0.12	122	0.25	0.75	184	0.32	0.68	233	0.5	0.5	278	0.42	0.58	
22	0.27	0.73	72	0.64	0.36	123	0.26	0.74	185	0.44	0.56	234	0.68	0.32	279	0.04	0.96	
23	0.29	0.71	73	0.7	0.3	124	0.43	0.57	186	0.09	0.91	235	0.44	0.56	281	0.39	0.61	
24	0.57	0.43	74	0.48	0.52	125	0.82	0.18	187	0.46	0.54	236	0.28	0.72	282	0.44	0.56	
25	0.84	0.16	75	0.17	0.83	126	0.36	0.64	188	0.27	0.73	237	0.32	0.68	283	0.46	0.54	
26	0.65	0.35	76	0.38	0.62	127	0.48	0.52	189	0.44	0.56	238	0.44	0.56	284	0.48	0.52	
27	0.7	0.3	77	0.5	0.5	128	0.48	0.52	190	0.3	0.7	239	0.52	0.48	285	0.53	0.47	
28	0.76	0.24	78	0.21	0.79	129	0.57	0.43	191	0.29	0.71	240	0.6	0.4	286	0.57	0.43	
29	0.75	0.25	79	0.95	0.05	130	0.15	0.85	192	0.41	0.59	241	0.34	0.66	287	0.29	0.71	
30	0.22	0.78	80	0.11	0.89	131	0.16	0.84	193	0.31	0.69	242	0.4	0.6	364	0.21	0.79	
31	0.78	0.22	81	0.16	0.84	132	0.13	0.87	194	0.56	0.44	243	0.31	0.69	365	0.44	0.56	
32	0.19	0.81	82	0.62	0.38	133	0.21	0.79	195	0.38	0.62	244	0.48	0.52	366	0.47	0.53	
33	0.32	0.68	89	0.38	0.62	134	0.27	0.73	196	0.21	0.79	245	0.42	0.58	367	0.36	0.64	
34	0.47	0.53	90	0.14	0.86	135	0.12	0.88	197	0.44	0.56	246	0.33	0.67	368	0.48	0.52	
35	0.94	0.06	91	0.6	0.4	136	0.08	0.92	198	0.32	0.68	247	0.57	0.43	369	0.53	0.47	
36	0.88	0.12	92	0.71	0.29	137	0.11	0.89	199	0.27	0.73	248	0.38	0.62	370	0.31	0.69	
37	0.61	0.39	93	0.39	0.61	138	0.21	0.79	200	0.47	0.53	249	0.23	0.77	371	0.42	0.58	
38	0.48	0.52	94	0.63	0.37	139	0.12	0.88	201	0.43	0.57	250	0.07	0.93	372	0.29	0.71	
39	0.6	0.4	95	0.5	0.5	140	0.2	0.8	202	0.5	0.5	251	0.4	0.6	373	0.48	0.52	
40	0.4	0.6	96	0.82	0.18	141	0.08	0.92	203	0.44	0.56	252	0.06	0.94	374	0.45	0.55	
41	0.69	0.31	97	0.17	0.83	142	0.03	0.97	204	0.42	0.58	253	0.11	0.89	375	0.52	0.48	
42	0.54	0.46	98	0.1	0.9	160	0.43	0.57	205	0.44	0.56	254	0.04	0.96	376	0.26	0.74	
43	0.84	0.16	99	0.29	0.71	161	0.39	0.61	206	0.64	0.36	255	0.42	0.58	377	0.38	0.62	
44	0.46	0.54	100	0.45	0.55	162	0.62	0.38	207	0.36	0.64	256	0.13	0.87	378	0.38	0.62	
45	0.72	0.28	101	0.34	0.66	163	0.24	0.76	208	0.32	0.68	257	0.25	0.75	379	0.36	0.64	
46	0.52	0.48	102	0.17	0.83	164	0.21	0.79	209	0.36	0.64	258	0.14	0.86	380	0.43	0.57	
47	0.56	0.44	103	0.53	0.47	165	0.29	0.71	210	0.44	0.56	259	0.3	0.7	381	0.47	0.53	
48	0.42	0.58	104	0.93	0.07	166	0.15	0.85	211	0.19	0.81	260	0.1	0.9	382	0.13	0.87	
49	0.52	0.48	105	0.43	0.57	167	0.5	0.5	212	0.19	0.81	261	0.25	0.75	383	0.57	0.43	
50	0.63	0.37	106	0.25	0.75	168	0.42	0.58	213	0.36	0.64	262	0.55	0.45				
															Media aritmética general		0.38	0.61

En la Tabla 9 se puede observar que la media aritmética para el p_i es de .61; esto, según la información que se tiene en la Tabla 8, indica un grado de dificultad general, ligeramente cargado a *medianamente difícil*. Se encontraron que el 16.4 % ($f = 44$) de los ítems son *difíciles*, el 32.7 % ($f = 88$) de los ítems son *medianamente difíciles*, el 29.7 % ($f = 80$) de los ítems son de *dificultad media*, el 7.4 % ($f = 20$) de los ítems son de dificultad *medianamente fáciles* y el 13.8 % ($f = 37$) de los ítems son de dificultad *fáciles*.



Análisis de distracción

El análisis de distracción tiene como propósito validar que las posibles respuestas que se muestran en un examen efectivamente cumplen con su función de distracción; es decir, lo que interesa observar para cada ítem es que cada una de sus opciones haya sido seleccionada por lo menos una vez. En la Tabla 10, a manera de ejemplo, se presenta la información que se recopiló acerca del comportamiento observado de los ítems con identificador 5 al identificador 25 del examen.

Tabla 10.

Análisis de distracción por ítem del 5 al 25.

ID	Ítem																			
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	21	22	23	24	25
A	4	6	13	4	10	4	11	8	4	9	2	14	6	7	0	6	3	8	5	3
B	2	15	8	14	11	9	11	7	15	2	15	3	7	1	4	17	9	7	7	1
C	2	7	11	12	2	3	9	17	7	10	3	2	7	10	8	3	6	11	2	17
D	14	2	4	1	9	12	1	4	4	4	5	5	9	6	5	4	6	6	21	0
SR	7	8	9	10	7	13	11	9	9	11	14	8	11	10	12	13	10	13	6	10

Como se puede observar en la Tabla 10, para cada ítem o columna existen una cantidad de celdas en fila, entre cuatro y cinco, las cuales se identifican o etiquetan con las letras A, B, C y D, que representan las opciones múltiples a elegir para cada pregunta del examen. Se puede observar un valor remarcado en negrita, lo que indica que esa opción es la respuesta para el ítem en cuestión; en caso de que existieran una o más celdas con un número 0 en cada columna de entre las filas mencionadas, indica que el análisis de distracción de alguna de las opciones no ha cumplido su propósito, y el cual se observa en el ítem 25.

La opción SR es para indicar sin respuesta, lo que significa el número de alumnos que no respondieron al ítem. El resultado de sumar los valores que aparecen en una columna y dividirlos entre el total de alumnos arroja como resultado un número aproximado al grado de dificultad que se muestra en la Tabla 9 para cada ítem.

4. DISCUSIÓN

La evaluación de la eficacia del examen de admisión en pensamiento lógico ha arrojado resultados significativos que merecen una reflexión crítica y detallada. En esta sección se discutirán los hallazgos principales y sus posibles implicaciones.



Índice de discriminación

El índice de discriminación es un indicador clave de la capacidad de un ítem para diferenciar entre aspirantes con diferentes niveles de pensamiento lógico. En este estudio, la mayoría de los ítems mostraron una alta capacidad discriminativa; este hallazgo permite inferir que el examen parece distinguir entre aspirantes con diferentes niveles de pensamiento lógico.

Sin embargo, la presencia de 30 ítems con baja discriminación indica la necesidad de una revisión continua (pudieran estar mal formulados o ser ambiguos) o, en su defecto, la eliminación de estos. Otro aspecto importante para enfatizar es la presencia de 2.23 % de los ítems (6 de 269) con un índice de discriminación negativo, lo que podría sugerir que los mismos podrían estar favoreciendo a aspirantes con menores habilidades en el pensamiento lógico. Es crucial mejorar los 36 ítems mencionados para asegurar que el examen mantenga su fiabilidad y validez.

Coefficiente de discriminación biserial

El coeficiente de discriminación biserial refuerza los hallazgos del índice de discriminación, proporcionando una validación adicional de que los ítems son efectivos en medir las habilidades en el pensamiento lógico de los aspirantes. La alta proporción (62.83 %, 169 de 269) de ítems con un coeficiente de discriminación biserial clasificado como *excelente* respalda la calidad del examen. No obstante, es importante abordar los ítems con coeficientes bajos para evitar cualquier sesgo en la evaluación.

Por otro lado, el hecho de que un 5.58 % de los ítems (15 de 269) tengan un coeficiente de discriminación biserial negativo es de su consideración, ya que estos ítems no solo fallan en discriminar adecuadamente, sino que pueden estar favoreciendo a aspirantes con menores habilidades en el pensamiento lógico, lo que contradice el propósito del examen. Estos ítems deben ser revisados de inmediato y modificados o eliminados según sea necesario con la intención de mantener la validez del examen.

Grado de dificultad

El análisis del grado de dificultad permite inferir que el examen tiene una distribución balanceada de ítems en términos de dificultad, aunque ligeramente inclinado hacia lo *medianamente difícil*, con una media aritmética de 0.61. Esta distribución es adecuada para evaluar una gama amplia de habilidades en el pensamiento lógico. Sin embargo, es esencial monitorear continuamente esta distribución para evitar que el examen se vuelva demasiado



fácil o *difícil*, lo que podría afectar su capacidad para diferenciar entre niveles de habilidad. Además, el hecho de que el 16.36 % de los ítems (44 de 269) fueron clasificados como *difíciles* podría ser problemático. Un número excesivo de ítems *difíciles* puede desmotivar a los aspirantes y no reflejar adecuadamente sus habilidades en el pensamiento lógico. Es crucial reevaluar estos ítems para asegurarse de que son justos y representativos de las habilidades y la calidad que se pretende medir.

Análisis de distracción

El análisis de distracción permitió inferir que las opciones de respuesta están cumpliendo su función de desviar a los aspirantes menos preparados, lo que es fundamental para la validez del examen. Sin embargo, algunos ítems mostraron distracciones ineficaces, lo que sugiere la necesidad de revisar y mejorar estas opciones, ya que ciertos distractores son demasiado obvios o, en su defecto, están mal diseñados. Es importante mencionar que la ineficacia de los distractores puede llevar a una subestimación de las habilidades de los aspirantes y afectar la validez del examen.

5. CONCLUSIONES

El diseño de eficacia del examen en pensamiento lógico es esencial para asegurar que las evaluaciones realmente midan lo que pretenden medir y para discriminar adecuadamente entre diferentes niveles de pensamiento lógico entre los aspirantes. A continuación, se presenta una evaluación crítica de los elementos involucrados en este diseño.

Índice de Discriminación

El índice de discriminación (d_r) es fundamental para identificar qué preguntas son capaces de distinguir entre aspirantes con diferentes niveles de habilidad. En este estudio, se observa un uso robusto del índice de discriminación, respaldado por la clasificación propuesta por Ebel y Frisbie (1986).

Como fortalezas de aplicar este índice de discriminación se encuentra:

1. Una metodología clara para calcular el índice de discriminación; incluyendo un ejemplo práctico, el cálculo del ítem 206, con la ecuación utilizada.
2. La distribución de los ítems según su índice de discriminación en la Tabla 5 muestra que la mayoría de los ítems (220 de 269) son clasificados como *excelentes*, lo que



indica una alta capacidad discriminativa del examen.

Por otro lado, como debilidad al examen de pensamiento lógico, se cita que la presencia de 30 ítems clasificados como *pésimos* y *pobres* sugiere que hay ítems que no están funcionando adecuadamente, por lo que se podría beneficiar de una revisión o eliminación para mejorar la fiabilidad general del examen.

Coefficiente de discriminación biserial

El coeficiente de discriminación biserial (r_{pbis}) complementa al índice de discriminación al considerar a todos los aspirantes y no solo a un subconjunto de ellos. Este coeficiente permite validar si las personas con las habilidades adecuadas son las que responden correctamente.

Como fortaleza de implementar el coeficiente de discriminación en el examen de pensamiento lógico se encuentran:

1. La media aritmética del coeficiente de discriminación biserial es *excelente*, con un valor de 0.37, lo cual indica que el examen tiene un buen nivel de discriminación en general.
2. La alta proporción de ítems clasificados como *excelentes* (62.82 %) respalda la validez discriminativa de los ítems del examen.

Por otro lado, como debilidad se puede decir que un porcentaje pequeño, pero significativo (5.57 %) de ítems, se clasifican como *pésimos*, lo que sugiere que aún hay espacio para mejorar, específicamente, ciertos ítems del examen.

Grado de dificultad

El grado de dificultad de los ítems es crítico para asegurar que la prueba no sea ni demasiado fácil ni demasiado difícil, permitiendo una evaluación precisa de las habilidades en el pensamiento lógico.

Como fortaleza se puede decir que, el examen de pensamiento lógico está relativamente balanceado en cuanto al grado de dificultad, dado que presenta una distribución de los ítems en diferentes categorías de dificultad; aunque ligeramente inclinada hacia la categoría *medianamente difícil* (32.7 %) y *dificultad media* (29.7 %).

Por otro lado, como debilidad se puede decir que la presencia de un 16.4 % de



ítems clasificados como *difíciles* podría indicar que algunas preguntas son excesivamente desafiantes para la mayoría de los aspirantes, lo que podría impactar negativamente la precisión del examen en ciertos casos.

Análisis de distracción

El análisis de distracción es crucial para validar que todas las opciones de respuesta en un ítem son efectivas en desviar a los aspirantes menos preparados.

Las siguiente dos declaraciones se consideran fortalezas, respecto a implementar el análisis de distracción:

1. La inclusión del análisis de distracción proporciona una capa adicional de validación, asegurando que las opciones incorrectas están cumpliendo su función.
2. La Tabla 10 presenta un análisis detallado del comportamiento de los distractores para una muestra de ítems, lo cual es útil para identificar y corregir opciones que no están funcionando adecuadamente.

Por último, como debilidad se considera que la presencia de ítems donde algunas opciones no han sido seleccionadas (ejemplo, ítem 25) indica que ciertos distractores pueden ser demasiado obvios o inefectivos, lo que podría reducir la validez del ítem y cambiar ligeramente los valores de discriminación.

En resumen, el diseño de la eficacia del examen en pensamiento lógico permite inferir un sólido enfoque metodológico y una implementación cuidadosa de índices de discriminación y coeficientes biserials, así como un análisis detallado del grado de dificultad y de los distractores. No obstante, existen áreas de mejora, particularmente en la revisión de ítems con baja discriminación y opciones de respuesta ineficaces. La revisión y mejora continua de estos ítems aumentará la fiabilidad y validez de la prueba, asegurando una evaluación más precisa y justa de las habilidades en el pensamiento lógico de los aspirantes.

CONFLICTO DE INTERESES

Ningún conflicto de relación financiera, laboral o personal puede influir en el trabajo de los autores.



CONTRIBUCIÓN DE AUTORES

Jair del Valle fue el responsable de la concepción de la idea de investigación, el diseño metodológico y del cálculo e interpretación de resultados. Por su parte, Sem Barba fue responsable de la búsqueda y clasificación de la información, del repaso de la literatura y de la presentación de los resultados en diversas audiencias. Ambos autores contribuyeron al manuscrito final.

FINANCIAMIENTO

Esta investigación se ha financiado exclusivamente con recursos de los autores. Esto nos permite excluir a terceros de cualquier responsabilidad

REFERENCIAS BIBLIOGRÁFICAS

- Aguirre Román, J. O. (2008). Justicia e igualdad en los sistemas especiales de admisión a las universidades: una reflexión a partir de Ronald Dworkin. *Estudios Socio-Jurídicos*, 10(1), 201-226.
- Benítez Lima, M. G., Ramos Ávila, A. E. y González López, C. (2015). Influencia de los antecedentes académicos para el ingreso en una universidad pública mexicana: un estudio de caso. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 13(3), 107-119. <https://dialnet.unirioja.es/servlet/articulo?codigo=5124968>
- Busche, K., Elks, M. L., Hanson, J. T., Jackson-Williams, L., Stephen, R. M., Parsons, W. L., Wofsy, D. y Yuan, K. (2020). The validity of scores from the new MCAT exam in predicting student performance: Results from a multisite study. *Academic Medicine*, 95(3), 387-395. <https://doi.org/10.1097/ACM.0000000000002942>
- Del Valle López, J. A. (2005). *Actitud y aptitud del docente de programación y su relación con el aprovechamiento teórico y práctico en alumnos de instituciones adventistas de pregrado* (Tesis de maestría). Universidad de Morelos.
- Díaz Barriga, Á. (2006). Las pruebas masivas: análisis de sus diferencias. *Revista Mexicana de Investigación Educativa*, 11(29), 583-615. https://www.scielo.org.mx/scielo.php?pid=S1405-66662006000200583&script=sci_arttext
- Ebel, R. L. y Frisbie, D. A. (1986). *Essentials of education measurement*. Prentice Hall.
- Geiser, S. (2016). Medición y evaluación para los procesos de admisión de la educación superior: hallazgos desde California. *Pensamiento Educativo*, 53(1), 1-18. <https://>



doi.org/10.7764/PEL.53.1.2016.7

- Glass, G. y Stanley, J. (1986). *Métodos estadísticos aplicados a las ciencias sociales*. Prentice-Hall Hispanoamericana.
- Greatrix, R., Nicholson, S. y Anderson, S. (2021). Does the UKCAT predict performance in medical and dental school? A systematic review. *BMJ Open*, 11, e040128. <https://doi.org/10.1136/bmjopen-2020-040128>
- Grob, M., Becerra, D., Rodríguez, A., Cristiane, J., Ramírez, V. y Sabag, N. (2015). Relación entre puntaje de prueba de selección universitaria y nota enseñanza media, y el rendimiento académico de la asignatura de Morfología en alumnos de primer año de odontología de la Universidad de los Andes. *International Journal of Morphology*, 33(2), 527-531. <http://dx.doi.org/10.4067/S0717-95022015000200019>
- Larroucau, T., Ríos, I. y Mizala, A. (2015). Efecto de la incorporación del ranking de notas en el proceso de admisión a las universidades chilenas. *Pensamiento Educativo*, 52(1), 95-118. <https://cuadernos.info/index.php/pel/article/view/26089/20945>
- Lievens, F., Sackett, P. R. y De Corte, W. (2022). Weighting admission scores to balance predictiveness diversity: The Pareto-optimization approach. *Med Educ*, 56(2), 151-158. <https://doi.org/10.1111/medu.14606>
- Scharfen, J., Peters, J. M. y Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, 67, 44-66. <https://doi.org/10.1016/j.intell.2018.01.003>
- Tapasco Alzate, O. A., Ruiz Ortega, F. J. y Osorio García, D. (2016). Estudio del poder predictivo del puntaje de admisión sobre el desempeño académico universitario. *Revista Latinoamericana de Estudios Educativos*, 12(2), 148-165. <https://revistasoj.s.ucaldas.edu.co/index.php/latinoamericana/article/view/4031/3739>
- Woo, S. E., LeBreton, J. M., Keith, M. G. y Tay, L. (2023). Bias, fairness, and validity in graduate-school admissions: A psychometric perspective. *Perspectives on Psychological Science*, 18(1), 3-31. <https://doi.org/10.1177/17456916211055374>
- Zwick, R. (2012). The role of admissions test scores, socioeconomic status, and high school grades in predicting college achievement. *Pensamiento Educativo*, 49(2), 23-30. <https://onomazein.letras.uc.cl/index.php/pel/article/view/25803/20721>

Esta obra está bajo una Licencia Creative Commons
"Reconocimiento No Comercial Sin Obra Derivada".

