

Validity and reliability of an unbounded item bank for placement and diagnostics in engineering

Carlos García Franchini¹; Martha Alvarado Arellano²; María Adelina Cruz Rodríguez³; José Víctor Flores Flores⁴; Beatriz Pérez Rojas⁵; Patricia Torrijos Muñoz⁶; Ricardo Morales Juárez⁷

^{1,7}Tecnológico Nacional de México campus Instituto Tecnológico de Puebla, México, carlos.garcia@puebla.tecnm.mx, maraaare@yahoo.com, adelina.cruz@puebla.tecnm.mx, josevictor.flores@puebla.tecnm.mx, beatriz.perez@puebla.tecnm.mx, patricia.torrijos@puebla.tecnm.mx, ricardo.morales@puebla.tecnm.mx

Abstract— *In the admission process for engineering faculties, the predominant selection method is evaluation through multiple-choice items. Under the premise that high scores positively correlate with academic success, it is imperative to have robust and reliable instruments. However, given the evolution towards Society 4.0 and 5.0, these processes must integrate advanced technologies and inclusive approaches that mitigate discrimination. This paper presents the initial phase of the MicroCell project, aimed at defining preparatory strategies for Education 2030, focused on reducing academic deficiencies detected through the selection process. The study centers on the design and validation of an assessment model based on an unbounded item bank, analyzing its quality and robustness as a placement tool. The results demonstrate that this format not only maintains the technical integrity of the exam but also effectively neutralizes random response bias. Thus, the proposal serves not only for selection purposes but also as a diagnostic tool to implement personalized academic actions that enhance the future engineer's performance.*

Keywords— *Placement exam, Unbounded item bank, Psychometric reliability, Assessment model, MicroCell.*

Validez y confiabilidad de una batería de reactivos no acotada para el diagnóstico y ubicación en ingeniería

Carlos García Franchini¹; Martha Alvarado Arellano²; María Adelina Cruz Rodríguez³; José Víctor Flores Flores⁴; Beatriz Pérez Rojas⁵; Patricia Torrijos Muñoz⁶; Ricardo Morales Juárez⁷

^{1,7}Tecnológico Nacional de México campus Instituto Tecnológico de Puebla, México, carlos.garcia@puebla.tecnm.mx, maraare@yahoo.com, adelina.cruz@puebla.tecnm.mx, josevictor.flores@puebla.tecnm.mx, beatriz.perez@puebla.tecnm.mx, patricia.torrijos@puebla.tecnm.mx, ricardo.morales@puebla.tecnm.mx

Resumen— En el ingreso a las facultades de ingeniería, el método predominante de selección es la evaluación mediante ítems de opción múltiple. Bajo la premisa de que un puntaje alto correlaciona positivamente con el éxito académico, resulta imperativo contar con instrumentos robustos y confiables. No obstante, ante la evolución hacia las sociedades 4.0 y 5.0, estos procesos deben integrarse con tecnologías avanzadas y enfoques inclusivos que mitiguen la discriminación. Este trabajo presenta la fase inicial del proyecto MicroCell, orientado a definir estrategias propedéuticas para la Educación 2030 centradas en reducir las deficiencias académicas detectadas mediante el proceso de selección. El estudio se centra en el diseño y validación de un modelo de evaluación basado en una batería de reactivos no acotada, analizando su calidad y robustez como instrumento de ubicación. Los resultados demuestran que este formato no solo mantiene la solidez técnica del examen, sino que neutraliza efectivamente el sesgo por respuesta aleatoria. Así, la propuesta no solo funciona para la selección, sino como una herramienta de diagnóstico para implementar acciones académicas personalizadas que mejoren el desempeño del futuro ingeniero.

Palabras clave— Examen de ubicación, Batería de reactivos no acotada, Confiabilidad psicométrica, Modelo de evaluación, MicroCell.

I. INTRODUCCIÓN

La selección de aspirantes es un proceso crítico para las Instituciones de Educación Superior (IES), dado que la demanda suele exceder la capacidad física o virtual necesaria para garantizar la calidad educativa. Si bien la delimitación de la matrícula responde a factores académicos, socioeconómicos y políticos, el proceso actual debe equilibrar la eficiencia administrativa con la justicia social, implementando acciones afirmativas que erradiquen cualquier forma de discriminación.

La selección no solo define el acceso a una institución, sino que refleja la decisión del estudiante por una carrera específica, elección profundamente influenciada por el entorno económico familiar. Como se menciona en [1], estos procesos constituyen evaluaciones sumativas de alto impacto, lo que exige evidencias de validez sólidas para garantizar que las inferencias derivadas de los resultados sean precisas. En consecuencia, la selección trasciende lo administrativo y genera efectos profundos en el aspirante y su familia, impactando incluso la salud física y mental de quienes son rechazados. Por el contrario, un proceso de selección robusto fortalece el prestigio institucional y optimiza indicadores clave como la eficiencia terminal y la reducción de la deserción.

En este sentido se subraya [2], [3], que las características del perfil de ingreso dictan la necesidad de diseñar estrategias de intervención tempranas para mitigar la deserción. El ingreso al sistema universitario no es un evento aislado, sino un **proceso progresivo** de aculturación académica, donde el estudiante debe identificar, asimilar y adoptar las pautas de comportamiento y las prácticas propias de la educación superior.

No obstante, en concordancia, se señala que la evolución tecnológica constante ha reconfigurado el progreso humano, influyendo en la estructura social, las dinámicas económicas y la organización de procesos complejos [4]. El ámbito educativo no es ajeno a esta transformación; han surgido nuevos esquemas exitosos para salvaguardar la calidad académica en todos los niveles [5]. Estas innovaciones son fundamentales para consolidar el papel de la educación en la sociedad del conocimiento, posicionándola como el preámbulo de las eras 4.0 y 5.0.

Como se ha mencionado anteriormente, la selección debe entenderse también como un mecanismo de equidad y no discriminación [6]. Este proceso facilita la asignación de cuotas para diversos grupos étnicos o sociales [7] y permite cumplir con ordenamientos jurídicos y políticos orientados a mitigar desequilibrios históricos, favoreciendo activamente a las minorías [8]. Asimismo, la selección funciona como una acción afirmativa que beneficia a estudiantes con discapacidad, sectores económicamente vulnerables, grupos migrantes [9], o mediante la aplicación de cuotas de género y programas de incentivos para el alto desempeño académico.

Independientemente del contexto, el objetivo de la selección debe orientarse adicionalmente hacia la ubicación diagnóstica. Esto implica que, más allá de generar una lista jerárquica de aceptados, el proceso debe aportar información precisa sobre las deficiencias académicas de cada estudiante. De este modo, es posible diseñar estrategias de intervención educativa que permitan subsanar lagunas de conocimiento provenientes de niveles previos o detectadas durante su formación inicial en las Instituciones de Educación Superior (IES).

A. El Proyecto MicroCell

Este enfoque de diseño de estrategias de intervención educativa constituye la esencia del proyecto MicroCell, cuyo punto de partida es la revisión exhaustiva de los contenidos y la calidad técnica del instrumento de selección. A partir del análisis de sus resultados y otros factores determinantes, se

busca fundamentar propuestas de intervención pedagógica. Por lo tanto, en este documento, el proceso se conceptualiza como un mecanismo de ubicación más que de simple exclusión. El propósito final es sistematizar esta experiencia y ofrecer recomendaciones para la formación integral de ingenieros como trabajo futuro.

B. Objetivos y Justificación de la Investigación

El objetivo central de esta investigación es verificar la validez de contenido, la confiabilidad y la robustez del instrumento de selección institucional. El estudio se justifica en el origen del instrumento, el cual fue diseñado internamente por docentes de la institución y pilotado con la población real antes de la pandemia. No obstante, debido a la interrupción de dicho periodo, los resultados se consolidan en el presente documento para dar respuesta a la pregunta de investigación: ¿posee el instrumento de selección diseñado y aplicado internamente la confiabilidad y robustez necesarias?

C. Factibilidad y Relevancia

La investigación resultó factible al emplear recursos y procesos institucionales internos, trabajando con la población total de aspirantes del año 2020 en el Tecnológico Nacional de México campus Instituto Tecnológico de Puebla. Las conclusiones, obtenidas en 2023, son altamente relevantes, ya que permiten validar el uso futuro de los resultados que vierta el instrumento dentro del proyecto MicroCell. Esto asegura que los procesos de ubicación beneficien directamente a las futuras generaciones de ingenieros mediante la mejora continua de su perfil académico mediante nuevos recursos de intervención educativa que fortalezcan sus competencias de ingreso y permanencia en la ingeniería.

II. METODOLOGÍA

Por su calidad y trascendencia, es importante para las instituciones contar con un instrumento de selección de aspirantes que sea confiable según la teoría clásica de la medición en donde la validez corresponde con la exactitud de la cuantificación y la confiabilidad se refiere a si el instrumento mide lo que dice y es estable en el tiempo, es decir es robusto [10], [11].

Aunque la validez y la confiabilidad justifican técnicamente por qué un estudiante no es admitido —buscando así mitigar el impacto negativo en el aspirante y su familia [1]—, surgen interrogantes críticas sobre el comportamiento de los sustentantes frente a la estructura del examen. Por ejemplo: ¿resuelven realmente los algoritmos?, ¿infieren las respuestas mediante el razonamiento?, o ¿simplemente seleccionan opciones al azar?

Estas dudas pueden resolverse analizando los resultados del proceso de selección. Al integrar estos datos con la base de datos del Departamento de Servicios Escolares, es posible asociar características sociodemográficas [12], factores derivados de la masificación educativa [9] e incluso perfiles

académicos de los aceptados. Asimismo, esto permite realizar inferencias sobre el comportamiento de los sustentantes y otros análisis estadísticos, ya sean descriptivos o inferenciales [13], fundamentados en la Teoría Clásica de los Tests [10], [11].

El instrumento de selección fue diseñado por docentes de la institución, cumpliendo con los doce componentes de validez para exámenes objetivos de la *American Educational Research Association* (AERA) [1]: (1) plan general del examen, (2) definición del dominio y objetivos de los resultados, (3) especificaciones técnicas, (4) desarrollo de ítems, (5) diseño y montaje, (6) producción, (7) aplicación, (8) calificación, (9) establecimiento del puntaje de corte, (10) reporte de resultados, (11) seguridad y banco de reactivos, y (12) elaboración del reporte técnico.

Resalta el hecho de que el reporte técnico incluyó una auditoría integral a la estructura del examen y a su base de datos en el servidor del proveedor. Esto permitió realizar, de manera previa al lanzamiento, los ajustes necesarios para garantizar una visualización óptima de los reactivos y las correcciones de código requeridas para asegurar la notación científica adecuada en cada área. Dichas acciones quedaron documentadas en un reporte que certifica la operatividad de la plataforma.

Asimismo, se llevó a cabo un análisis robusto de validez y confiabilidad. Este estudio detallado demuestra la integridad del instrumento mediante pruebas estadísticas que justifican su eficacia en la selección de los perfiles más aptos.

El sistema se aplicó de forma virtual a los aspirantes de todos los programas académicos. El instrumento evalúa dos competencias fundamentales: Pensamiento Matemático y Dominio de la Lengua. La sesión tiene una duración de tres horas y emplea un formato flexible **sin un límite fijo de reactivos**, permitiendo que el sustentante avance según su propio ritmo sin un mínimo o máximo preestablecido.

A. Estructura del Instrumento

El banco total consta de 375 reactivos, distribuidos estratégicamente según el perfil del aspirante:

- Matemáticas (244 reactivos): De estos, 202 son comunes para todas las áreas, mientras que 42 son exclusivos para los aspirantes a Ingeniería.
- Dominio de la Lengua (111 reactivos): Se aplican de forma universal a todos los sustentantes.
- Enfoque Administrativo (20 reactivos): Dirigidos únicamente a los aspirantes de Ciencias Económico-Administrativas. Para fines del análisis estadístico, estos reactivos se integran dentro de la competencia de Dominio de la Lengua.

TABLA I
ESTRUCTURA DEL INSTRUMENTO

Área de Conocimiento	Total de Reactivos	Distribución por Perfil
Matemáticas	244	202 compartidos (Ingeniería y Económico-Adm.) y 42 exclusivos para Ingeniería.
Dominio de la Lengua	111	Aplicados a la totalidad de los aspirantes.
Enfoque Administrativo	20	Exclusivos para el área Económico-Administrativa.
Total General	375	

A. Población y Corpus de Datos

La población evaluada fue de 2,296 aspirantes, compuesta por 2,136 candidatos a Ingeniería y 160 a Ciencias Económico-Administrativas. Esta participación generó un volumen de 848,109 registros, cada uno integrado por los siguientes campos de información:

- Identificación (Ficha): Código numérico único asignado al aspirante.
- Identificación del Ítem (Reactivo): Código numérico secuencial del reactivo.
- Clasificación Temática (Área): Competencias matemáticas o de la lengua.
- Segmentación (Carrera): Clasificación del uso del reactivo (Ingeniería, Administración o ambas).
- Respuesta del Usuario: Opción seleccionada (A, B, C, D).
- Estatus del Ítem: Categorización del resultado (Correcta, Incorrecta, Omitida o No mostrada).
- Trazabilidad Temporal: Registros de fecha y hora exacta (*timestamps*) de inicio y cierre para cada reactivo, lo que permite medir latencias de respuesta.

III. RESULTADOS

A. Análisis descriptivo de los datos

Debido a que los reactivos se presentan de forma aleatoria, el sistema asigna la categoría de "Visto" e inicia el cronómetro en cuanto el estudiante accede a ellos. Dependiendo de la acción del sustentante, el estatus se registra como Correcto, Incorrecto u Omitido (si se avanza sin seleccionar una opción).

La Fig. 1 presenta la distribución de frecuencias de los reactivos vistos, cuya media es de 112 con una desviación estándar de 176.5. El histograma muestra un marcado sesgo a la derecha, con una clase modal en el intervalo [74, 86] y una frecuencia de 280; esto indica que la mayoría de los alumnos visualizó alrededor de 82 reactivos. Asimismo, se observa una "cola" a partir de la clase [242, 254], que representa casos anómalos vinculados a respuestas al azar, lo cual se corrobora con los tiempos de respuesta registrados.

Por su parte, el número de reactivos resueltos (Fig. 2) presenta una media de 109.4 y una desviación estándar de 176.5. Aunque este estadístico es técnicamente más preciso

que el de reactivos vistos, la diferencia entre ambos (112 frente a 109) no es significativa.

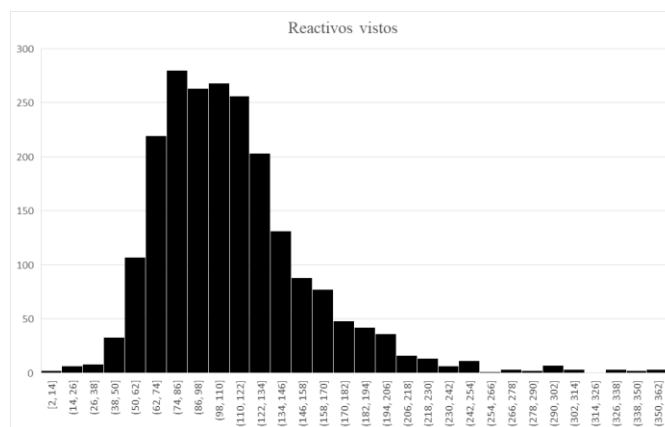


Fig. 1 Histograma de frecuencias de reactivos vistos.

Por su parte, el número de reactivos resueltos (Fig. 2) presenta una media de 109.4 y una desviación estándar de 176.5. Aunque este estadístico es técnicamente más preciso que el de reactivos vistos, la diferencia entre ambos (112 frente a 109) no es significativa.

El histograma de la Fig. 2 mantiene la misma clase modal [74, 86] y el mismo comportamiento de sesgo y anomalías a partir de la clase [242, 254] que la distribución anterior.

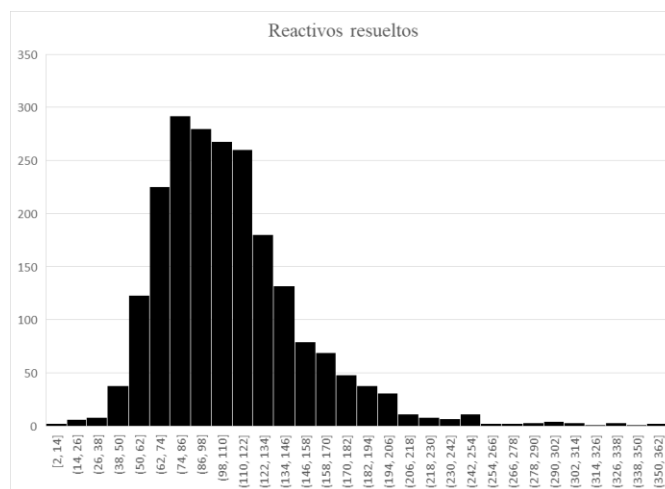


Fig. 2 Histograma de frecuencias de reactivos resueltos.

Finalmente, el registro de tiempos indica que el promedio empleado por respuesta fue de 56.74 segundos (0.94 min), con una desviación estándar de 19.9 segundos. La clase modal se ubica en [0.426, 0.497] minutos, intervalo en el que se resolvieron aproximadamente 17,648 reactivos-alumno.

El histograma de distribución de frecuencias de los reactivos resueltos correctamente se presenta en la Figura 3. Se observa una clase modal en el intervalo [57, 64], con una media de 65 reactivos y una desviación estándar de 98. La

distribución muestra un sesgo a la derecha, identificándose un valor máximo de 197 aciertos, el cual se clasifica como un caso excepcional o posiblemente anómalo.

Por otro lado, al analizar la relación entre los reactivos omitidos y el total de reactivos vistos, se obtuvo la recta de regresión $y = 0.0505x - 2.9911$. El coeficiente de determinación calculado, $R^2 = 0.0577$, indica una correlación sumamente baja entre estas variables. Esto sugiere que no existe una dependencia lineal entre el volumen de reactivos visualizados y la tendencia a omitirlos; por el contrario, existe una alta probabilidad de que el alumno elija una respuesta una vez que se le presenta el reactivo. En términos prácticos, de cada 100 reactivos vistos, apenas 5 son omitidos por los sustentantes.

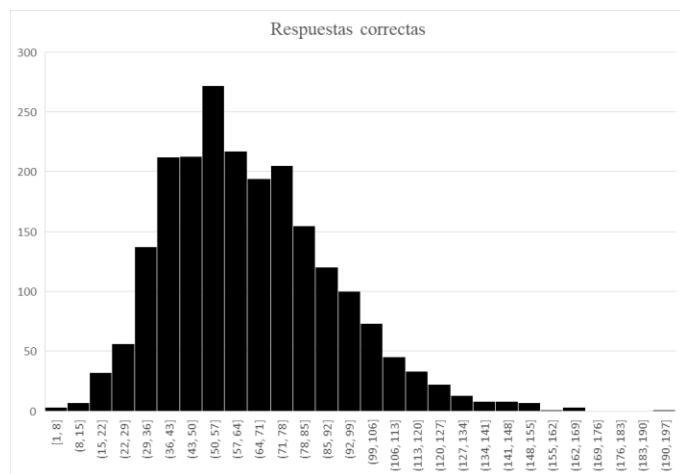


Fig. 3 Distribución de frecuencias de los reactivos resueltos.

Se identificaron indicios de que aproximadamente 20 estudiantes (equivalentes a 9 de cada 1,000) respondieron al azar. De este grupo, solo dos sujetos visualizaron la casi totalidad de los reactivos, mostrando una tendencia a seleccionar posiblemente aquellos de menor dificultad.

En el diagrama de dispersión de la Figura 4, se analiza la correlación entre el número de reactivos vistos y los resueltos correctamente. Los datos arrojan una media de 59.18% de aciertos respecto al total de reactivos visualizados, con una desviación estándar del 2.75%. Por su parte, el modelo de regresión lineal predice una tasa de eficiencia del 77%; es decir, por cada 130 reactivos vistos, se estima que 100 serán resueltos de forma correcta.

El análisis de dispersión revela que la eficiencia de los sustentantes mejora cuando visualizan menos de 200 reactivos. Esta discrepancia entre la media observada y la predicción del modelo sugiere que los resultados no son producto del azar, sino que reflejan el nivel de conocimiento de los estudiantes.

Por otro lado, la Fig. 5 ilustra la dispersión entre los reactivos incorrectos y los reactivos vistos, mostrando una correlación del 64.61%. Se observa que los estudiantes que visualizaron un máximo de 200 reactivos obtuvieron, en su gran mayoría, menos de 100 aciertos; de hecho, superar dicho volumen de reactivos no se tradujo en un incremento significativo en su calificación final. Finalmente, destaca el coeficiente de 56.635 en la ecuación de la recta, lo cual indica que la gran mayoría de los sustentantes registró un mínimo de 56 respuestas incorrectas independientemente del total de reactivos vistos.

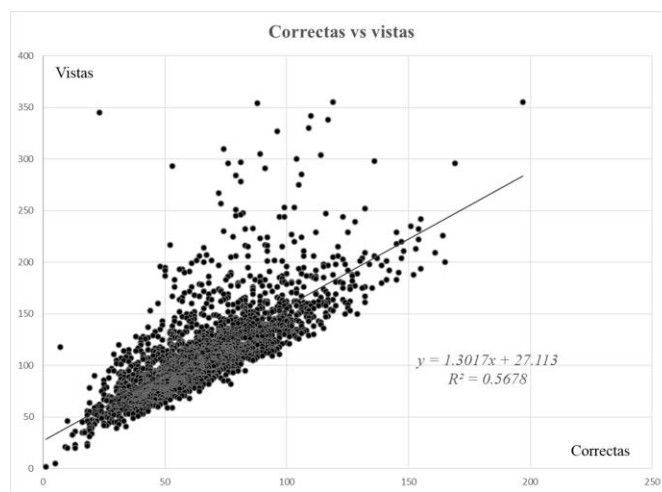


Fig. 4 Diagrama de dispersión y recta de regresión de reactivos correctos vs. vistos.

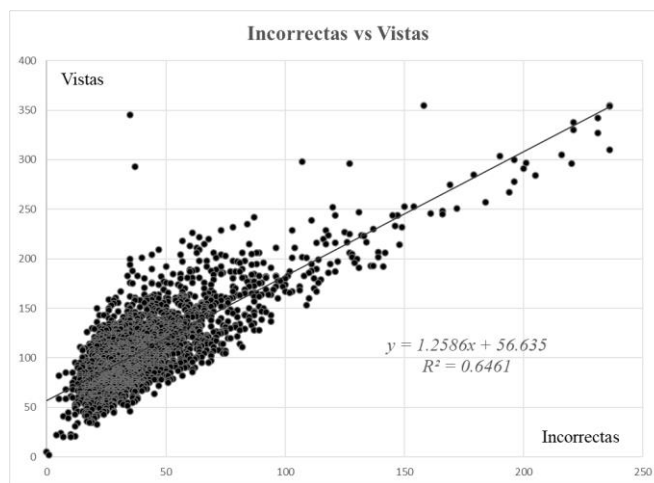


Fig. 5 Diagrama de dispersión y recta de regresión de reactivos incorrectos vs. vistos.

B. Contenido de reporte de análisis de validez

El análisis de validez por reactivo se consolidó en un formato uniforme que sintetiza diversos indicadores técnicos. Un ejemplo de esta estructura se observa en la Fig. 6,

correspondiente al reactivo 23, la cual se divide en las siguientes secciones:

1. Cabecera

- 1.1. **Identificación:** Código único del reactivo.
- 1.2. **Área de conocimiento:** Clasificación del dominio evaluado (Lenguas o Matemáticas).
- 1.3. **Subárea:** Identificación específica dentro de los dos dominios principales.
- 1.4. **Aplicación:** Población objetivo (Ingeniería, Administración o ambas).

2. Conteo de aplicación del reactivo

- 2.1. **Resueltos:** Número total de estudiantes que interactuaron con el reactivo, independientemente del resultado.
- 2.2. **Factor de uso:** Indicador que mide la proporción de aspirantes que resolvieron el reactivo respecto al total de la muestra ($N = 2296$).

3. Deciles (Ingeniería)

- 3.1. **Distribución por deciles:** Clasificación de los sustentantes donde el Decil 1 agrupa a los alumnos con mayor desempeño (más respuestas correctas) y el Decil 10 a aquellos con menor desempeño. Estos datos se presentan mediante un polígono de frecuencias y su respectiva recta de regresión.

4. Correctas por cuartiles

- 4.1. **Análisis comparativo:** Incluye tablas y polígonos de frecuencias absolutas organizados por cuartiles. Se presentan histogramas diferenciados para los sustentantes de las áreas de Ingeniería y Administración, permitiendo una comparativa visual del rendimiento.

5. Análisis de dificultad

Este apartado evalúa la complejidad intrínseca de los reactivos mediante tres enfoques:

- 5.1. **Global:** Incluye indicadores para la totalidad de los sustentantes (N). Se considera el número de respuestas correctas C_i^G y el factor de uso¹ F_i para determinar la proporción de éxito (P-Pearson) $p_i^G = C_i^G / (NF_i)$. El índice de dificultad global $d_i^G = 1 - p_i^G$ representa la proporción de respuestas incorrectas; un valor cercano a 1 indica un reactivo de alta dificultad, mientras que un valor cercano a 0 indica un reactivo fácil.
- 5.2. **Administración:** Aplica los mismos indicadores del punto 5.1, pero segmentados exclusivamente

para la población de dicha área. En cuyo caso:

$$N^A = 160, p_i^A = C_i^A / (N^A F_i) \text{ y } d_i^A = 1 - p_i^A.$$

- 5.3. **Ingeniería:** Aplica los mismos indicadores segmentados para los sustentantes de Ingeniería.

Considerando: $N^I = 2136, p_i^I = C_i^I / (N^I F_i)$ y

$$d_i^I = 1 - p_i^I.$$

Interpretación del Índice de Dificultad (d):

Se utiliza una escala de tipo Likert para su clasificación cualitativa:

- $0 \leq d < 0.20$: Muy fácil.
- $0.20 \leq d < 0.40$: Fácil.
- $0.40 \leq d < 0.60$: Dificultad media.
- $0.60 \leq d < 0.80$: Difícil.
- $0.80 \leq d \leq 1.00$: Muy difícil.

6. Calidad global de los distractores

Cada reactivo cuenta con cuatro opciones de respuesta (A, B, C y D) desplegadas aleatoriamente.

- **Indicadores:** Se analiza el número de elecciones por opción y su valor porcentual (%e).
- **Balance ideal:** Aunque lo óptimo es una distribución equitativa, la elección de la respuesta correcta condiciona los resultados. Por ello, se busca un equilibrio en la proporción de elección de las opciones incorrectas.

En este indicador se considera la interpretación cualitativa y la sugerencia dada para una opción de respuesta incorrecta es el criterio de aceptación siguiente.

Criterios de aceptación (%e):

- $\%e > 0.20$: Opción buena (Aceptable).
- $0.10 \leq \%e \leq 0.20$: Opción regular.
- $0 \leq \%e < 0.10$: Opción mala.

7. Análisis de discriminación del reactivo

La discriminación es el principal indicador de calidad, pues distingue entre alumnos de alto y bajo rendimiento.

- 7.1. **Pendiente de regresión:** Se calcula sobre los datos de deciles y cuartiles. Para ser válido, la pendiente debe ser negativa (indicando que, a menor desempeño general, menor probabilidad de acierto en el reactivo). Una pendiente positiva señala un reactivo atípico que debe ser descartado.
- 7.2. **Índices de discriminación inter-cuartil e inter-decil:**

Corresponden respectivamente con

$$d_i^Q = (Q_i^1 - Q_i^4) / Q_i^1$$

¹ El factor de uso F_i es la proporción real de estudiantes que activaron el reactivo i en su evaluación, respecto del total N , según el área del sustentante.

$$d_i^D = (D_i^1 - D_i^{10}) / D_i^1$$

Estos representan la razón del diferencial entre los aspirantes con mejores resultados y los de resultados más bajos.

- **Índice inter-decil d_i^D** : Es una prueba más "ácida" o rigurosa al comparar solo los extremos (el 10% superior vs. el 10% inferior), mientras d_i^Q lo hace con el 25% de cada extremo.
- **Interpretación:** Ambos deben ser estrictamente positivos. Si el valor es menor a 0.10, el reactivo se considera atípico y debe eliminarse. Debe considerarse la siguiente escala de discriminación de la tabla 2.

TABLA II
ESCALA DE CALIDAD DE LA DISCRIMINACIÓN

Valor del índice	Clasificación	Acción sugerida
$d < 0$	Muy mala	Desechar o modificar drásticamente
$0 \leq d < 0.10$	Pobre	Modificar drásticamente
$0.10 \leq d < 0.30$	Regular	Revisar y modificar ligeramente
$0.30 \leq d < 0.40$	Buena	Revisión opcional
$d \geq 0.40$	Excelente	Conservar reactivo

7.3. Índice de discriminación total dt_i :

Compara la razón de respuestas correctas del cuartil más alto contra el más bajo respecto al total de participantes que lo accedieron. Este indicador suaviza el efecto de la discriminación en función del crecimiento de la población evaluada.

Se determina para deciles como $dt_i^D = (D_i^1 - D_i^{10}) / N_i$, mientras para cuartiles $dt_i^Q = (Q_i^1 - Q_i^4) / N_i$.

Para interpretarlo cualitativamente se puede emplear el punto 7.2 con el cambio de escala 4 y 10 respectivamente. Este indicador suaviza la discriminación dependiendo del crecimiento de la población.

8. Conclusión para cada reactivo

Al finalizar el formato resumen, con la información desplegada se presenta la conclusión para cada reactivo en el formato de la Fig. 6, por ejemplo, para el reactivo 23: "El reactivo 23, es excelente, su dificultad es medianamente difícil, con excelente discriminación y sus opciones son buenas".

C. Reporte de datos de análisis de calidad y validez

Después del análisis de cada uno de los 375 reactivos que componen el Banco del examen de selección de aspirantes y un reporte de cerca de 400 páginas, se tiene:

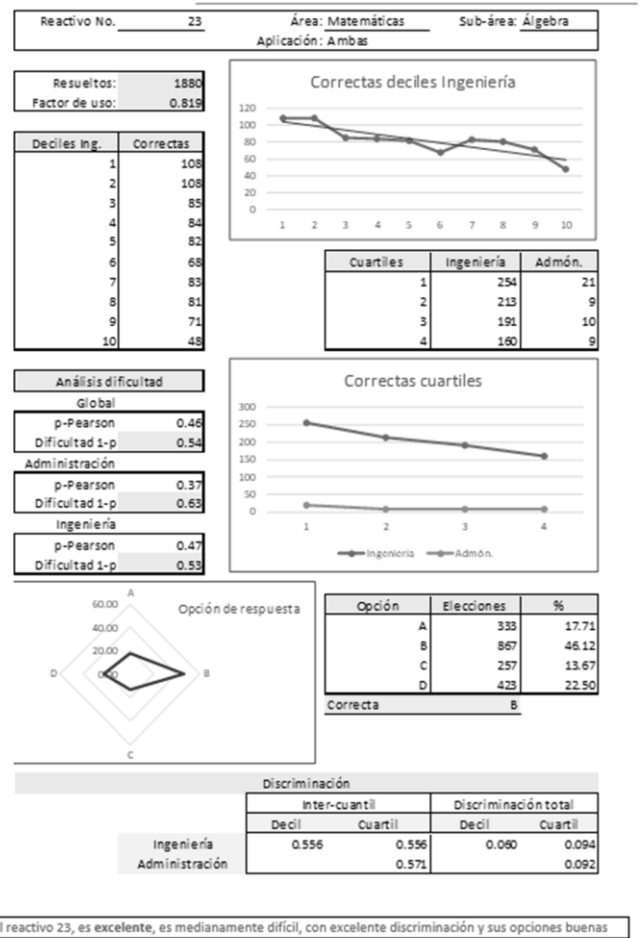


Fig. 6 Ejemplo de formato de reporte de análisis de reactivo.

1. Distribución del grado de dificultad: Muy fácil, 39,10%; Fácil, 88, 24%; Media, 105, 28%; Difícil, 135, 36% y Muy difícil, 8, 2%. Con base en el grado de dificultad que experimentaron los aspirantes se considera que es adecuada, ya que el porcentaje de los tres niveles más bajos representa 62% mientras la media del número de respuestas de los estudiantes fue 65 reactivos correctos de 109 resueltos, también como valor medio (Fig. 7).

2. Distribución del grado de discriminación: Muy mala, 20, 5%; Pobre, 5, 2%; Regular, 19, 5%; Buena, 23, 6% y Excelente, 308, 82%. Con base en el grado de discriminación, se tiene un nivel excelente en el 82% de los casos y pobre o mala, apenas el 6% que se deben modificar (Fig. 8).

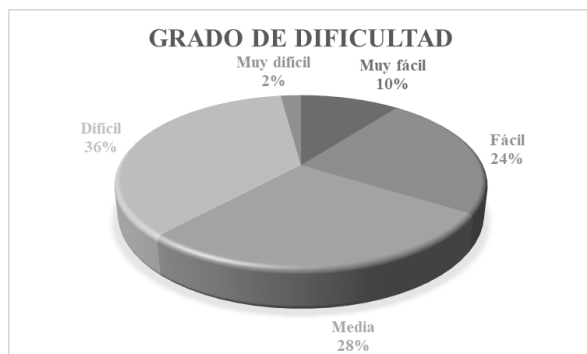


Fig. 7 Distribución de reactivos por grado de dificultad

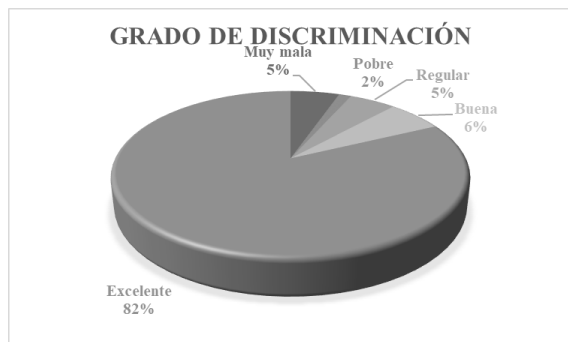


Fig. 8 Distribución por grado de discriminación.

3. Distribución de la calidad de las opciones de respuesta: Buenas, 183, 49%; Regulares, 160, 43% y Malas, 32, 8%. Apenas el 8% de los reactivos presenta opciones clasificadas como malas para los cuales se propone modificar alguna de las opciones para mejorar su calidad (Fig. 9).



Fig. 9 Distribución por calidad de las opciones.

4. Distribución de la calidad general del reactivo: Excelente, 86, 23%; Buena, 219, 58%; Aceptable, 49, 13%; Mala, 21, 6% y Pésima, 0, 0%. De manera importante la calidad general de los reactivos permite utilizar los reactivos clasificados como aceptables, apenas el 6% clasificados como malos, esto son apenas 20 reactivos que deben modificarse (Fig. 10).

5. La revisión más detallada de los reactivos clasificados con calidad mala, indica que hay inconsistencia sobre todo en el indicador de discriminación en contraste con la dificultad, visiblemente detectados por la pendiente negativa de la recta de regresión. Por lo que se sugiere no eliminarlos, porque son un indicador fuerte del nivel de deficiencia que presentan los estudiantes y que debe ser atendido en curso de nivelación, esto es la ubicación.

6. Debido a que la proporción de aspirantes de ingeniería (2136) es mucho mayor que la de administración (160), dicha cantidad de población afecta los cálculos de discriminación de los reactivos. El objetivo de este reporte no consideró esta variable de segmentación, pero se puede revisar como parte del formato de cada reactivo.



Fig. 10 Distribución por calidad del reactivo.

IV. CONCLUSIONES

El análisis estadístico descriptivo del modelo de selección demuestra un proceso robusto y confiable. A diferencia de los instrumentos con un número fijo de reactivos, este modelo permite un ordenamiento jerárquico preciso de los aspirantes según su nivel de logro, eliminando la problemática de empates técnicos y la falta de discriminación psicométrica.

La aleatoriedad en la asignación de reactivos garantiza que cada estudiante se enfrente a un examen prácticamente único². Aunque el diseño permite la omisión de ítems, los datos revelan que esta variable es marginal, ya que más del 95% de los casos no registraron omisiones.

Un hallazgo crítico es la mitigación del sesgo por respuesta aleatoria. En exámenes de opción múltiple convencionales, existe el riesgo de que el aspirante responda al azar para avanzar; sin embargo, en este modelo, los resultados reflejan conocimiento real: solo 9 de cada 1000 alumnos optaron por el azar. Por lo tanto, se confirma que el número actual de reactivos es óptimo y se recomienda

² Cuando se activa la evaluación para un estudiante, se genera un vector aleatorio que contiene los ordinales de todos los reactivos disponibles para su área. De tal forma que este vector indica en cada momento cuál es el reactivo que se habrá de hacer visible. Para Ingeniería este vector tiene 255 ordinales, lo que implica 255! exámenes diferentes con idéntico grado de dificultad.

mantener el diseño original del instrumento. Se observa en la Fig. 11, de acuerdo con la recta de regresión, que la proporción de aciertos disminuye cuando los sustentantes resuelven más reactivos.

En conclusión, el instrumento institucional de selección es válido, confiable y robusto, cumpliendo con los objetivos y la pregunta de investigación planteados. La información obtenida permite no solo la ubicación del estudiante, sino la detección temprana de deficiencias académicas. Esto fundamenta el diseño de materiales de nivelación personalizados, eje central del proyecto MicroCell.

Finalmente, la validez demostrada hace viable la escalabilidad del modelo a una base mayor de estudiantes. La integración futura de procesos de Big Data e Inteligencia Artificial (IA) permitirá evolucionar hacia un sistema de ubicación individualizado y dinámico, consolidando las estrategias de apoyo necesarias para solventar las brechas formativas de cada futuro ingeniero.

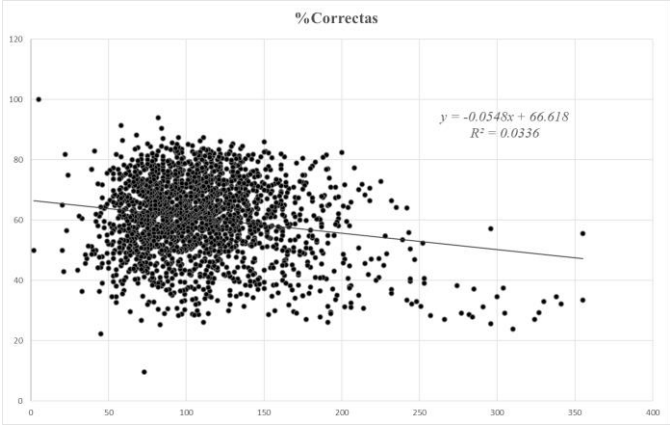


Fig. 11 Distribución del porcentaje de respuesta correctas, respecto de los reactivos vistos.

V. TRABAJO FUTURO

La densidad de los datos que se muestra en la tabla 3, invita al análisis de los temas que fueron menos favorecidos con los aciertos de los estudiantes, lo que desencadena el trabajo futuro en que se centra el proyecto MicroCell.

El trabajo futuro consiste en contrastar los programas de bachillerato, contra los prerrequisitos reales que necesitan las carreras de ingeniería y junto con los datos obtenidos sobre las deficiencias de los estudiantes, obtenidas del examen de ubicación, se obtienen los temas prioritarios para los que se deberán desarrollar estrategias de nivelación, con el objeto de favorecer la calidad de la recepción de la nueva matrícula y a la vez mejorar el desempeño de los estudiantes a lo largo de sus estudios, ya que esas deficiencias deberán ser compartidas por los alumnos ya inscritos en la licenciatura.

Resulta importante aclarar que en el momento de este reporte ya existen 2 años de trabajo del proyecto MicroCell, y

éste ya considera debilidades adicionales detectadas como consecuencia de la pandemia del COVID19.

TABLA III
CUANTILES DE LA DISTRIBUCIÓN DE RESPUESTAS CORRECTAS

Decil	Reactivos correctos	Cuartil	Reactivos correctos	
	Ingeniería		Ingeniería	Administración
1	197-101	1	197-82	139-75
2	100-87			
3	86-77	2	81-62	76-60
4	77-69			
5	69-62	3	62-47	59-48
6	62-56			
7	56-50	4	47-1	47-9
8	50-43			
9	43-35			
10	35-1			

AGRADECIMIENTO/RECONOCIMIENTO

El cuerpo académico ITPUE-CA6 agradece a todos los estudiantes que presentaron este examen, y que sin saberlo contribuyeron en este proceso, al igual que a las autoridades del Tecnológico Nacional de México campus Instituto Tecnológico de Puebla por permitirnos desarrollar la parte experimental del examen de ubicación.

REFERENCIAS

[1] M. Sánchez, M. García, A. Martínez y E. Buzo, "El Examen de Ingreso a la Universidad Nacional Autónoma de México: Evidencias de Validez de una Prueba de Alto Impacto y Gran Escala", *Rev. Iberoamericana. Evaluación Educativa*, vol. 13, no. 2, pp. 107-128, 2020. doi: 10.15366/rie2020.13.2.006.

[2] C. García y M. Alvarado, "Análisis de Validez de los Reactivos del Examen de Selección", *Academia Journals Puebla 2023*, vol. 15, México, 2023. ISSN 1946-5351.

[3] A. Torres, J. Acuña, G. Acevedo y J. Villanueva, "Caracterización del perfil de ingreso a la universidad. Consideraciones para la toma de decisiones", *Rev. Iberoam. Investig. Desarro. Educ.*, vol. 9, no. 18, ene.-jun. 2019. doi: 10.23913/ride.v9i18.435.

[4] R. Anderson, L. Casabury, K. Lee y A. Rezaee, "What is Development Engineering?", *Introduction to Development Engineering*, Cham, Suiza: Springer, 2023, pp. 1-28. doi: 10.1007/978-3-030-86065-3_1.

[5] S. Rider, M. Peters, M. Hyvönen y T. Besley, "World Class Universities: A Contested Concept", *Evaluating Education: Normative Systems and Institutional Practices*, Singapur: Springer, 2021. doi: 10.1007/978-981-15-7598-3.

[6] P. Pineda y S. Mishra, "The semantics of diversity in higher education: differences between the Global North and Global South", *Higher Education*, vol. 84, mayo 2022. doi: 10.1007/s10734-022-00870-4.

[7] L. Cremonini, D. Westerheijden y J. Enders, "Disseminating the right information to the right audience: cultural determinants in the use (and misuse) of rankings", *Higher Education*, vol. 55, no. 3, 2008. doi: 10.1007/s10734-007-9062-8.

[8] N. B. Dhawan, D. Z. Belluigi y G. E. I. Idahosa, "'There is a hell and heaven difference among faculties who are from quota and those who are non-quota': under the veneer of the 'New Middle Class' production of Indian public universities", *Higher Education*, oct. 2022. doi: 10.1007/s10734-022-00932-7.

- [9] H. Van't Land, A. Corcoran y D.-C. Iancu, "The Promise of Higher Education", *Evaluating Education: Normative Systems and Institutional Practices*, Cham, Suiza: Springer, 2021. doi: 10.1007/978-3-030-67245-4.
- [10] O. F. Salazar, C. M. Vélez y J. J. Zuleta, "Evaluación de conocimientos con exámenes de selección múltiple: ¿tres o cuatro opciones de respuesta? Experiencia con el examen de admisión a posgrados médico-quirúrgicos en la Universidad de Antioquia", *Iatreia*, vol. 28, no. 3, pp. 300-311, jul.-sep. 2015.
- [11] H. Oviedo y A. Campos-Arias, "Metodología de investigación y lectura crítica de estudios. Aproximación al uso del Coeficiente alfa de Cronbach", *Rev. Colomb. Psiquiatr.*, vol. 24, no. 4, pp. 572-580, 2005.
- [12] C. Guzmán y O. V. Serrano, "Las puertas del ingreso a la educación superior: el caso del concurso de selección a la licenciatura de la UNAM", *Rev. Educ. Sup.*, vol. 40, no. 157, pp. 31-53, ene.-mar. 2011.
- [13] O. I. Trejos, L. E. Muñoz y G. R. Solarte, "Aplicación del modelo 4Q para seleccionar aspirantes a Ingeniería de Sistemas a partir del perfil profesional y ocupacional", *Entre Cienc. e Ing.*, vol. 15, no. 29, pp. 37-45, ene.-jun. 2021. doi: 10.31908/19098367.1703.