

Evidencias de validez del proceso de admisión a una escuela de medicina en México

Blanca Ariadna Carrillo-Avalos^{a,‡}, Iwin Leenen^{b,§}, Juan Andrés Trejo-Mejía^{c,¶}, Melchor Sánchez-Mendiola^{c,d,¶,*}

Facultad de Medicina



Resumen

Introducción: Los procesos de admisión a las escuelas de medicina deben demostrar la validez de la interpretación de sus resultados al ser evaluaciones de alto impacto.

Objetivo: Aplicar un modelo que integra los marcos de validez de Messick y Kane para obtener evidencias de validez del proceso de admisión en una escuela de medicina.

Método: Se realizó un estudio observacional analítico en la Facultad de Medicina de la Universidad Autónoma de San Luis Potosí de México, en las generaciones 2013 y 2014. El proceso de admisión incluye examen psicométrico (EP), examen de conocimientos (EC) y EXANI-II.

Se realizó en dos pasos: 1) argumento de usos e interpretaciones y 2) argumento de validez, conformado por tres etapas: validez instrumental, verificación de la interpretación y la decisión, utilidad de las acciones. En cada etapa se establecieron hipótesis sobre las inferencias de acuerdo a Kane, para fundamentarse con las fuentes de Messick.

Resultados: Se analizaron resultados de 1,373 aspirantes en 2013; y 1,554 de 2014; así como los 145 alumnos admitidos en cada generación. Se identificó un factor g de inteligencia, que tiene repercusión en los resultados del proceso de admisión. El EC explica el 15% de la varianza de las calificaciones del primer año (2013), el EP y el EXANI-II no tienen influencia en el primer año de

^a Departamento de Ciencias Morfológicas, Facultad de Medicina, Universidad Autónoma de San Luis Potosí, SLP, México.

^b Facultad de Psicología, Universidad Nacional Autónoma de México, Cd. Mx., México.

^c Facultad de Medicina, Universidad Nacional Autónoma de México, Cd. Mx., México.

^d Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia, Universidad Nacional Autónoma de México, Cd. Mx., México.

ORCID ID

[‡] <https://orcid.org/0000-0003-4111-4795>

[§] <https://orcid.org/0000-0003-4807-540X>

[¶] <https://orcid.org/0000-0002-0680-6836>

[¶] <https://orcid.org/0000-0002-9664-3208>

Recibido: 8-septiembre-2023. Aceptado: 21-noviembre-2023.

* Autor para correspondencia: Melchor Sánchez-Mendiola.

CUAIED, UNAM. Anexo Centro Ciencias de la Complejidad. 1er

Piso. Circuito Cultural S/N, Ciudad Universitaria. Del. Coyoacán.

Cd. Mx. C.P. 04510. México.

Correo electrónico: melchorsm@unam.mx

Este es un artículo Open Access bajo la licencia CC BY-NC-ND

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

la carrera (2013 y 2014). Hay correlación entre algunos subcomponentes que miden un constructo semejante. Existe diferencia en los resultados de Biología (EC) entre hombres y mujeres (2014). El nivel de satisfacción de los alumnos acerca del proceso es bueno.

Discusión y conclusiones: Las fuentes de evidencia de validez proporcionaron información relevante para informar y mejorar el proceso de admisión, se requieren más datos y un equipo interdisciplinario para recabar más tipos de evidencias y de mejor calidad. Se recomienda utilizar marcos de referencia claros en estos procesos y planear la recolección de datos usando dichos marcos.

Palabras clave: Examen de admisión; validez; evaluación de alto impacto; escuelas de medicina; México.

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Validity evidence in the admission process to a medical school in Mexico

Abstract

Introduction: Medical school admission processes need to demonstrate the validity of their scores' interpretations, since they are high-stakes assessments.

Objective: To apply an integrative model using Messick's and Kane's frameworks to obtain validity evidence for a medical school admission process.

Method: An analytical observational study was performed at the Autonomous University of San Luis Potosí Faculty of Medicine in Mexico, with the 2013 and 2014 cohorts. The admission process included a psychometric

test (PT), a general knowledge test (KT), and EXANI-II. The model was applied in two steps: 1) uses and interpretations argument, and 2) validity argument, which has three stages: instrumental validity, interpretation and decision verification, and action usefulness. Hypotheses were established at each stage using Kane's inferences, which were supported using Messick's sources of evidence.

Results: The results of 1,373 applicants from 2013 and 1,554 from 2014 were analyzed, as well as the academic performance of the 145 students admitted in each cohort. A general intelligence factor (g) was identified as influencing the results of admission process. KT explains 15% of the variance in the first-year grades (2013), PT and EXANI-II have no impact on first-year's academic performance (2013 and 2014). There was a correlation among some subcomponents that measure a similar construct. There was a difference in Biology results (KT) between men and women. Students' degree of satisfaction with the admission process is adequate.

Discussion and conclusions: The sources of validity evidence provide relevant information to support and improve the admission process, more data and an interdisciplinary team to retrieve more sources of quality evidence are needed. It is recommended to use well-defined frameworks in these processes and plan data collection and retrieval using these models.

Keywords: College admissions; validity; high-stakes assessment; medical school; Mexico.

This is an Open Access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

INTRODUCCIÓN

Las evaluaciones de alto impacto como los exámenes de admisión a las escuelas de medicina, tienen consecuencias importantes sobre instituciones y personas¹. Los procesos de admisión son muy diversos y suelen incluir información de varios tipos, como exámenes, entrevistas, juicio situacional y promedio del bachillerato. Dadas las altas consecuencias de estas evaluaciones, es importante demostrar que los

resultados reflejan el nivel real de los constructos evaluados y que las decisiones son justas^{2,3}.

En estos contextos se recomienda realizar un análisis integral con un marco de referencia sólido para aportar evidencias de validez²; sin embargo, es común que las publicaciones sobre validez de los exámenes de admisión se enfoquen en solo uno de los aspectos a evaluar, como la validez predictiva, y no suelen especificar el marco de referencia uti-

lizado⁴⁻¹³. En México estos análisis son escasos y estudian un solo tipo de prueba¹⁴⁻¹⁷. Los procesos de admisión y los instrumentos utilizados en las escuelas de medicina latinoamericanas en general han sido poco estudiados, por lo que es importante identificar y evaluar sus fuentes de evidencias de validez, así como los marcos conceptuales utilizados para interpretar los resultados de los exámenes. Debido a la alta demanda de espacios educativos y la responsabilidad de formar profesionales de la salud competentes para atender las necesidades de la población, el proceso de admisión requiere demostrar que cumple sus objetivos a través de la validación.

El marco de referencia de validez tradicional, que incluía la validez de contenido, de constructo y de criterio¹⁸, ha sido reemplazado por marcos de referencia contemporáneos, como los de Messick y Kane. El marco de Messick unificó los tres tipos de validez del marco tradicional y estipula cinco fuentes de evidencia de validez^{19,20}, mientras que el de Kane implica dos pasos generales con inferencias para desarrollar el argumento de validez²¹⁻²³. Los autores propusieron un nuevo modelo²⁴, que integra ambos marcos e incluye aportaciones de otros autores como Russell²⁵. En el presente trabajo aplicamos este modelo con el objetivo de obtener y evaluar las evidencias de validez del proceso de admisión en una escuela de medicina mexicana. La pregunta de investigación que orientó el trabajo fue: ¿cómo utilizar el concepto actual de validez para recoger y analizar fuentes de evidencia del proceso de admisión en una escuela de medicina?

MÉTODO

Contexto e instrumentos

El estudio se realizó con los datos de los aspirantes que presentaron el proceso de admisión en los años 2013 y 2014 para ingresar a la Facultad de Medicina de la Universidad Autónoma de San Luis Potosí (UASLP) en México. Este proceso incluyó tres componentes: examen psicométrico (EP), elaborado por el Centro de Salud Universitario de la UASLP; examen de conocimientos (EC), elaborado por una comisión interna de la Facultad de Medicina de la UASLP; y el Examen Nacional de Ingreso a la Educación Superior (EXANI-II, elaborado por el Centro Nacional de Evaluación para la Educación Superior,

A.C.)²⁶. Cada prueba estuvo conformada por preguntas de opción múltiple (POM) diferentes en cada ocasión, e incluyó a su vez varias secciones a las que nos referimos como subcomponentes (**tabla 1**).

Modelo de validez

Alineado al marco de Kane, el enfoque utilizado implica dos pasos generales: (1) argumento de usos e interpretaciones y (2) argumento de validez. El segundo paso se conforma de tres etapas en las que se establecen hipótesis en el marco de las inferencias de Kane, para ser comprobadas por medio de las fuentes de evidencia estipuladas por Messick.

1. **Argumento de usos e interpretaciones.** Se especifican objetivo del instrumento, usuarios propuestos, usos que se darán a las puntuaciones, constructo que se pretende medir e interpretación de las puntuaciones (con referencia a norma o a criterio). Después de que se ha aplicado el instrumento, se describen las características generales de la población de sustentantes.
2. **Argumento de validez.** En este paso se establecen las hipótesis basadas en las inferencias de Kane. El análisis de la evidencia para demostrar las hipótesis con base en las fuentes de evidencia de Messick se lleva a cabo a través de un argumento de validez que consta de las tres etapas de Russell:

- I. **Validez instrumental.** Esta etapa se debe realizar antes de utilizar el instrumento. El responsable de proveer evidencias es el desarrollador de la prueba. Se verifica si el instrumento cumple el objetivo del constructo medido, por medio de evidencia sobre las características psicométricas de las puntuaciones, representación del constructo y las amenazas a la validez²⁷. Las inferencias de Kane, sus hipótesis correspondientes, las fuentes de evidencia de Messick para comprobarlas, y el plan de recolección de evidencia se señalan en la **tabla 2**.
- II. **Verificación de la interpretación y la decisión.** En esta etapa se verifica si la interpretación de los resultados de la prueba es apropiada y si puede informar la toma de decisiones acerca de las acciones subsecuentes. El responsable de aportar la información es

Tabla 1. Dominios del conocimiento evaluados mediante los procesos de admisión de las generaciones 2013 y 2014

2013			
Examen	Total de ítems	Subcomponente	# de ítems
Psicométrico	154 POM de 5 opciones	Razonamiento verbal	30
		Lectura	20
		Razonamiento abstracto 1	60
		Razonamiento abstracto 2	44
Conocimientos	160 POM de 5 opciones	Biología	35
		Química	35
		Español	25
		Inglés	35
		Físico-matemático	30
EXANI-II	100 POM de 4 opciones	Razonamiento verbal	25
		Razonamiento numérico	25
		Razonamiento matemático	25
		Español	25
2014			
Examen	Total de ítems	Subcomponente	# de ítems
Psicométrico	163 POM de 5 opciones	Razonamiento verbal	50
		Lectura	26
		Razonamiento abstracto 1	60
		Razonamiento abstracto 2	27
Conocimientos	160 POM de 5 opciones	Biología	35
		Química	35
		Español	25
		Inglés	35
		Físico-matemático	30
EXANI-II	100 POM de 4 opciones	Pensamiento matemático	25
		Pensamiento analítico	25
		Estructura de la lengua	25
		Comprensión lectora	25

POM: preguntas de opción múltiple.

quien realiza la prueba, y se lleva a cabo una vez que se ha aplicado esta (tabla 2).

III. Utilidad de las acciones. Esta etapa implica evaluar las consecuencias de las acciones, ya sean voluntarias o involuntarias, y el grado en que las positivas superan a las negativas (tabla 2).

Datos recopilados y análisis estadísticos

Se obtuvieron los resultados de los procesos de admisión de 2013 y 2014 para la licenciatura de médico cirujano, así como las calificaciones durante la carrera de los 145 alumnos que ingresaron en cada generación. Los criterios de inclusión para los sustentantes

fueron haber presentado todas las pruebas que conformaron el proceso de admisión correspondiente. Para el análisis de regresión logística se incluyeron las calificaciones de todos los alumnos inscritos por primera vez en el primer año de la carrera de médico cirujano en la Facultad de Medicina de la UASLP durante los años 2013 y 2014. Se excluyeron los datos de los sustentantes que presentaron solo una o dos de las pruebas, así como de los estudiantes que se inscribieron, pero estaban recursando alguna asignatura o se inscribieron y no cursaron las asignaturas.

Con estos datos se realizó estadística descriptiva para el argumento de usos e interpretaciones. Se

Tabla 2. Plan de recolección de evidencias para las tres etapas de validación

I. Validez instrumental				II. Verificación de la interpretación y la decisión			III. Utilidad de las acciones					
	A. Puntuación	B. Generalizabilidad	C. Extrapolación	A. Puntuación	B. Generalizabilidad	C. Extrapolación	D. Implicaciones					
Inferencias de Kane e hipótesis	La regla de puntuación es apropiada. La calidad de redacción y del formato de los ítems es buena.	Los ítems son una muestra adecuada del dominio.	La puntuación observada se relaciona con un criterio relevante.	La puntuación refleja el proceso cognitivo apropiado. El instrumento midió la(s) dimensión(es) planeada(s).	El tamaño de la muestra fue adecuado. No existen/ sí existen diferencias entre grupos. Confiabilidad. La puntuación observada se relaciona con la de instrumentos que miden el mismo constructo/no se relaciona con la de instrumentos que miden un constructo diferente.	El proceso de admisión predice el primer año de la carrera. La relación con otras variables se puede generalizar a un dominio más amplio del conocimiento.	La interpretación de los resultados es adecuada. Los alumnos tienen éxito académico en la carrera. Es razonable y justificable el tipo de escala de la prueba.					
Fuente de evidencia de Messick	1. Contenido			2. Procesos de respuesta	3. Estructura interna		4. Relación con otras variables	5. Consecuencias				
Elemento	a. Definición del dominio	b. Proceso de desarrollo del instrumento	c. Relevancia del dominio	d. Representación del dominio	No aplica	a. Dimensión	b. Invarianza de la medida	c. Confiabilidad	a. Relación prueba-criterio	b. Generalización de la validez	No aplica	
Evidencia para un Instrumento con POM	Ofrece detalles con respecto de lo que la prueba mide. Transforma el constructo teórico en un dominio de contenido concreto.	Propiedades de los ítems.	Grado en que cada ítem en una prueba es relevante para el dominio que se evalúa.	Expertos. Alineación con currículo. Alineación de contenido y nivel cognitivo evaluado.	Demuestra que se están llevando a cabo los procesos cognitivos esperados al contestar la prueba.	Identificar dimensiones de los ítems.	Análisis del ítem: dificultad y discriminación. Ausencia de sesgo sistemático: demuestra que las puntuaciones serán las mismas cuando se comparan entre grupos con diferentes características como raza, edad o sexo, de acuerdo con la taxonomía de equivalencia.	En cada ocasión que se aplica el instrumento se obtendrán resultados semejantes.	Las puntuaciones predicen el desempeño en un criterio relevante, de manera predictiva o concurrente.	Metaanálisis de los estudios de validación anteriores en condiciones semejantes.	Los usuarios están de acuerdo con la interpretación de los resultados.	Establecer la trayectoria académica de los alumnos y de la generación en conjunto. Utilidad de las escalas con referencia a la norma.

Continúa en la siguiente página...

Tabla 2. Continuación...

	I. Validez instrumental			II. Verificación de la interpretación y la decisión			III. Utilidad de las acciones	
	A. Puntuación	B. Generalizabilidad	C. Extrapolación	A. Puntuación	B. Generalizabilidad	C. Extrapolación	D. Implicaciones	
Lista de cotejo para demostrar las hipótesis	La tabla de especificaciones de la prueba: • Describe detalladamente las áreas del contenido y las habilidades cognitivas para cuya medición se ha diseñado el instrumento. • Enlista las subáreas y los niveles cognitivos que se miden. • Muestra los estándares específicos de contenido, objetivos curriculares, o habilidades contenidas dentro de los diferentes niveles cognitivos.	Revisión de los ítems por expertos que aseguren su exactitud técnica. Revisión de los ítems por expertos de medición para determinar qué tan bien se adhieren a los estándares de principios de escritura de ítems de calidad. Revisión de sensibilidad para evitar varianzas irrelevantes al constructo. Piloteo de los ítems con análisis estadístico para seleccionar los ítems más apropiados para uso operativo.	Revisión de ítems por expertos que evalúen: • Que todos los aspectos importantes del dominio sean medidos por la prueba. • Si la prueba presenta contenido trivial o irrelevante.	Expertos externos e independientes califican cada ítem para determinar si: • Representa completa y suficientemente al dominio. • Concuerda con el estándar de contenido o un elemento de la tabla de especificaciones de la prueba. • Concuerda con el nivel cognitivo que se pretende alcanzar durante las clases.	Explicación de los dominios del conocimiento que deben ser evaluados para el ingreso a la licenciatura en Medicina y los porcentajes necesarios para cada campo. Demostrar la combinación de constructos que deben ser evaluados en los aspirantes a ingresar a la licenciatura en Medicina.	Entrevistas cognitivas, pensar en voz alta. Tiempos de respuesta. Análisis factorial confirmatorio.	Índice de dificultad por ítem y general de la prueba. Funcionamiento diferencial del ítem. Alfa de Cronbach o KR20-21. Análisis de sendesros.	Metaanálisis de estudios de validación anteriores en condiciones semejantes Encuesta a los usuarios: alumnos, profesores y administrativos de la facultad. Calificaciones de cada alumno para cada asignatura de 1° a 5° año. Bibliografía en donde se demuestre la utilidad del uso de la escala con referencia a la norma en este tipo de evaluaciones.
POM: preguntas de opción múltiple.								

realizó análisis de regresión logística para conocer la relación de las evaluaciones de admisión con el desempeño académico, tomando como variable dependiente a las calificaciones del primer año de la carrera y como variable independiente a las evaluaciones del proceso de admisión. También se desarrolló un modelo a través de un análisis factorial confirmatorio (AFC) para identificar las dimensiones que causan la variabilidad de las puntuaciones de los componentes del proceso de admisión. Además, se elaboró un modelo para predecir la calificación en los distintos años a partir de los resultados de los años anteriores y el examen inicial a través de análisis de senderos. Estos se realizaron con el software SAS (versión 9.4), y se consideró $p < 0.05$ como significativa.

Consideraciones éticas

Aprobación del comité de investigación (CI-006-2019) y del comité de ética en investigación (CEI-2019-004), de la Facultad de Medicina de la UASLP.

RESULTADOS

1. Argumento de usos e interpretaciones

- a) El objetivo de la prueba.
- Proceso en general: “...asegurar que los aspirantes que ingresen a la universidad cuenten con las competencias requeridas para la realización de sus estudios, utilizando procesos confiables, rigurosos, certificados, transparentes y en constante modernización”²⁸.
 - EP: son “...evaluaciones psicológicas estandarizadas que miden las aptitudes básicas para el estudio”²⁹.
 - EC: “Evalúa los conocimientos, las destrezas y las habilidades requeridas de los aspirantes a ingresar de acuerdo con el perfil del alumno pretendido en las licenciaturas...”²⁸.
 - EXANI-II: “...establecer el nivel de potenciali-

dad de un individuo para lograr nuevos aprendizajes...”³⁰.

- b) Usuarios propuestos: la Facultad de Medicina de la UASLP.
- c) Usos: seleccionar a “los aspirantes que cuentan con las competencias requeridas para la realización de sus estudios de licenciatura en médico cirujano”²⁸.
- d) Constructo medido.
- EP: “1. Razonamiento verbal: evalúa la capacidad de utilizar sinónimos y antónimos, paráfrasis incompletas, y definiciones de palabras. 2. Retención y comprensión: mide la habilidad para comprender y retener por medio de lectura de comprensión. 3. Razonamiento abstracto: valora la capacidad de razonamiento lógico e inmediato ante problemas cotidianos por medio de figuras y búsqueda de semejanzas y diferencias”²⁹.
 - EC: “Explora la capacidad de comprensión y razonamiento en 5 diferentes áreas del conocimiento: Físico-Matemático, Biología, Química, Inglés y Español”³¹.
 - EXANI-II: “...evalúa la habilidad para analizar y resolver problemas con base en principios elementales de las matemáticas... [y] ...la capacidad de comunicación del sustentante”³².
- e) Las interpretaciones de los resultados. Se realiza con referencia a norma, se calcula un puntaje ponderado a partir de los resultados en las distintas pruebas (EP 15%, EC 45%, EXANI-II 40%)²⁶ y los 145 aspirantes con mayor puntaje agregado son admitidos.

En la **tabla 3** se observa la proporción entre hombres y mujeres de aspirantes y alumnos admitidos de ambas generaciones.

Tabla 3. Proporción entre hombres y mujeres de aspirantes y alumnos admitidos en 2013 y 2014

	2013					2014				
	Total	M	%	H	%	Total	M	%	H	%
Aspirantes	1,373	748	54.5	625	45.5	1,554	916	59.0	638	41.0
Admitidos	145	76	52.4	69	47.6	145	66	45.5	79	54.5

M: mujeres; H: hombres.

2. Argumento de validez

I. Validez instrumental

Puntuación - Contenido - Proceso de desarrollo del instrumento. El EC de la Facultad de Medicina²⁶, fue elaborado en 2013 y en 2014 por una comisión interna que verifica que cada ítem se adhiera a los estándares de principios de escritura de ítems de calidad. La guía del EXANI-II en ambos años^{30,32} menciona que incluye preguntas de prueba y de control.

Puntuación - Contenido - Relevancia del dominio. El EC es elaborado por una comisión especial que se asegura de la pertinencia de los ítems, mientras que en la guía del EXANI-II de 2013 y de 2014^{30,32} se aclara que son los expertos en evaluación del aprendizaje quienes se encargan de verificar la relevancia del dominio.

Generalizabilidad - Contenido - Representación del dominio. En la Guía del Examen Nacional de Ingreso a la Educación Superior de 2013 indica que se elaboraron “más de 60 versiones del EXANI-II de selección... todas equivalentes en contenido y grado de dificultad. Cada cuadernillo contiene una mezcla distinta de reactivos y opciones de respuesta...”³⁰, mientras que en 2014 fueron más de 70 versiones³².

También menciona que “El EXANI-II incluye únicamente preguntas del tipo opción múltiple, con cuatro opciones de respuesta”, y que cada respuesta correcta equivale a un punto. También menciona que quienes elaboran los ítems son expertos en evaluación educativa³².

II. Verificación de la interpretación y la decisión

- **Puntuación - Estructura interna -Dimensionalidad.** Se realizó AFC para cada generación por separado. Se identificó un factor general, al que se llamó factor común de inteligencia g, que causa gran parte de la varianza de cada subcomponente y que explica las calificaciones y las correlaciones entre las calificaciones de los 13 subcomponentes. También se identificó la influencia de cuatro subfactores: Verbal, Matemático, Día 1 y Día 2.
- 2013: La mayor parte de las correlaciones entre subcomponentes son moderadas (+0.4 a +0.6), las más altas (≥ 0.7) son Química (EC) con Biología (EC) y con Físico-Matemático (EC), y de Matemáticas (EXANI-II) con Físico-Matemático (EC) y con Razonamiento Numérico (EXANI-II). Las

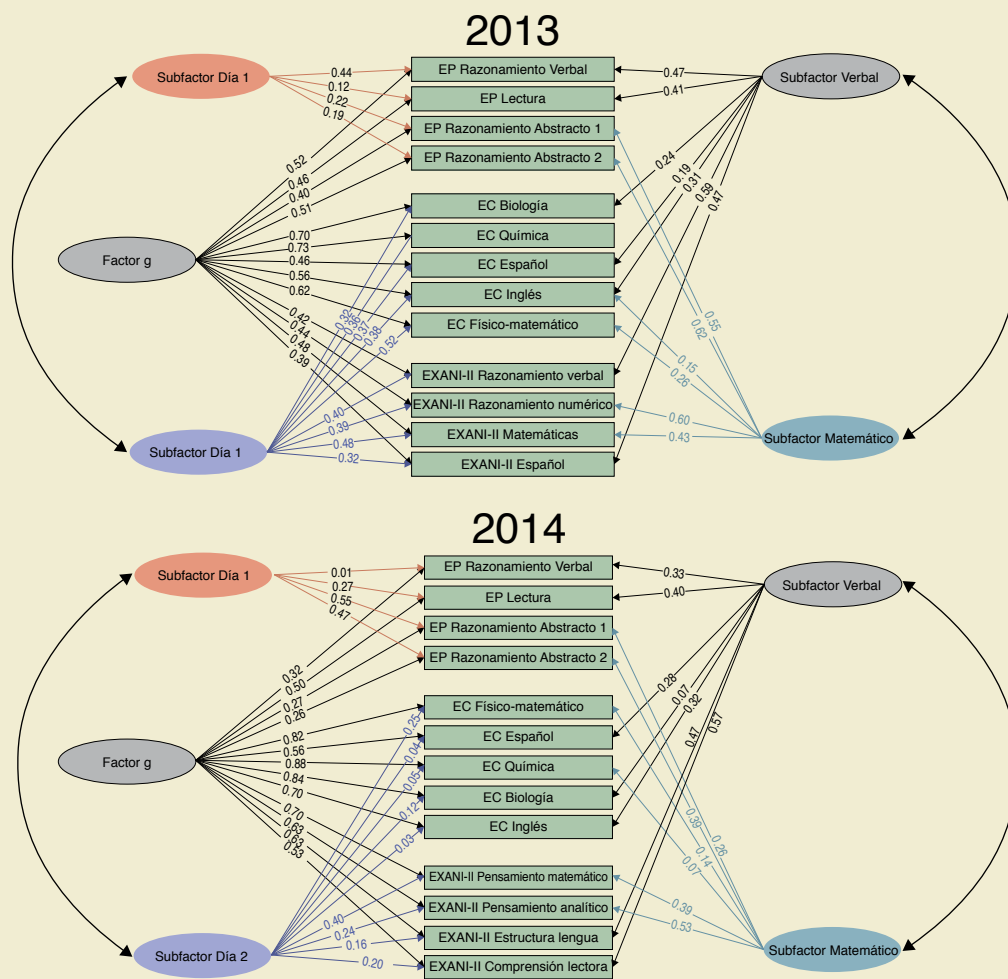
Tabla 4. Matriz de correlaciones entre los subcomponentes del proceso de admisión de 2013 (sombreado en gris) y 2014

			Psicométrico				Conocimientos					EXANI-II			
	2013		RV	L	RA1	RA2	F-M	E	Q	B	I	RV	RN	M	E
		2014	RV	L	RA1	RA2	F-M	E	Q	B	I	PM	PA	EL	CL
Psicométrico	RV	RV		0.38	0.49	0.38	0.5	0.46	0.34	0.46	0.46	0.53	0.45	0.43	0.41
	L	L	0.28		0.29	0.42	0.39	0.37	0.28	0.37	0.36	0.42	0.31	0.33	0.35
	RA1	RA1	0.10	0.34		0.51	0.38	0.38	0.24	0.35	0.39	0.39	0.5	0.42	0.35
	RA2	RA2	0.20	0.36	0.43		0.43	0.43	0.28	0.39	0.45	0.41	0.53	0.52	0.33
Conocimientos	F-M	F-M	0.27	0.44	0.24	0.25		0.65	0.47	0.59	0.64	0.56	0.52	0.57	0.45
	E	E	0.28	0.38	0.18	0.21	0.49		0.47	0.56	0.68	0.5	0.53	0.61	0.43
	Q	Q	0.31	0.46	0.26	0.25	0.74	0.49		0.5	0.51	0.45	0.39	0.45	0.39
	B	B	0.30	0.46	0.26	0.24	0.67	0.49	0.74		0.6	0.57	0.49	0.54	0.47
	I	I	0.32	0.49	0.24	0.26	0.06	0.51	0.63	0.61		0.58	0.63	0.69	0.48
EXANI-II	RV	PM	0.32	0.44	0.28	0.03	0.73	0.43	0.67	0.56	0.56		0.58	0.56	0.57
	RN	PA	0.31	0.42	0.03	0.34	0.65	0.45	0.60	0.53	0.52	0.74		0.67	0.49
	M	EL	0.35	0.49	0.23	0.27	0.61	0.49	0.58	0.55	0.59	0.62	0.58		0.50
	E	CL	0.36	0.51	0.23	0.27	0.55	0.42	0.50	0.47	0.54	0.58	0.56	0.64	

RV: razonamiento verbal, L: lectura, RA1: razonamiento abstracto 1, RA2: razonamiento abstracto 2, B: biología, Q: química, E: español, I: inglés, FM: físico-matemático, RN: razonamiento numérico, M: matemáticas, PM: pensamiento matemático, PA: pensamiento analítico, EL: estructura de la lengua, CL: comprensión lectora.

Todos los valores tienen $p < 0.0001$.

Figura 1. Pesos del factor de inteligencia y los subfactores en los subcomponentes del proceso de admisión de 2013 y 2014



EP: examen psicométrico, EC: examen de conocimientos.

Los índices de bondad para el modelo final en 2013 fueron $\chi^2 = 74.3$ y $gL = 37$ con $p < 0.001$; $SRMR = 0.012$, $RMSEA = 0.026$, y $AGFI = 0.98$. Para 2014 fueron $\chi^2 = 61.7$ y $gL = 37$ con $p < 0.01$; $SRMR = 0.010$, $RMSEA = 0.020$, y $AGFI = 0.985$.

relaciones más débiles (≤ 0.3) fueron de Lectura y Razonamiento Abstracto 1 (EP) con el resto de los subcomponentes (tabla 4).

- 2014: Razonamiento Verbal, Razonamiento Abstracto 1 y Razonamiento Abstracto 2 (EP) correlacionan débilmente con todos los subcomponentes. El resto de correlaciones son moderadas, excepto por Físico-Matemático (EC) con Química (EC) y con Pensamiento Matemático (EXANI-II), Química con Biología (EC), y Pensamiento

Matemático con Pensamiento Analítico (EXANI-II), que son fuertes (tabla 4).

A través del AFC se obtuvo el peso de cada subfactor y del factor común sobre cada subcomponente y para cada generación (figura 1).

Generalizabilidad - Estructura interna - Invarianza de la medida. La prueba de la diferencia de parámetros entre hombres y mujeres en cada subcomponente y para cada generación mostró diferen-

Tabla 5. Resultados del análisis de regresión logística para 2013 y 2014, por componente y subcomponente

		2013		2014	
		Parámetro	p	Parámetro	p
	Psicométrico	-0.011	0.04	0.010	0.02
	EXANI-II	0.009	0.53	0.053	<0.001
	Conocimientos	0.025	<0.001	0.030	<0.001
Psicométrico	Razonamiento verbal	-0.014	0.38	-0.003	0.72
	Lectura	-0.018	0.35	-0.018	0.40
	Razonamiento Abstracto 1	-0.007	0.48	0.018	0.01
	Razonamiento Abstracto 2	-0.022	0.17	0.026	0.16
Conocimientos	Biología	0.027	0.13	0.040	0.002
	Química	0.037	0.02	0.019	0.38
	Español	0.001	0.93	0.012	0.60
	Inglés	0.017	0.21	0.025	0.06
	Físico-matemático	0.028	0.14	0.046	0.01
EXANI-II	Razonamiento verbal	-0.032	0.40		
	Razonamiento lógico-matemático	0.035	0.30		
	Matemáticas	0.018	0.65		
	Español	0.0004	0.98		
	Pensamiento matemático			0.052	0.05
	Pensamiento analítico			0.047	0.76
	Estructura de la lengua			0.010	0.005
	Comprensión lectora			0.084	0.002

cias importantes entre hombres y mujeres en el EP, sobre todo con respecto de las cargas del subfactor Día 1 para ambas generaciones (**figura 1**). Además, en el caso de Biología en 2014, el efecto de inteligencia global sí tiene diferencia entre hombres y mujeres ($p = 0.014$).

Generalizabilidad - Estructura interna - Confiabilidad. Se reportó índice de confiabilidad de Kuder Richardson de 0.78 en 2014. No hay información para 2013.

Extrapolación - Relación con otras variables - Relación prueba-criterio. Para la generación 2013, el resultado del proceso de admisión explica 4% de la varianza de las calificaciones de todos los alumnos de primer año ($p = 0.01$), y 7% de la varianza de solo los aprobados ($n = 126$, $p = 0.0024$). Las calificaciones del proceso de admisión de la generación 2014 explican 24% de la varianza de las calificaciones de todos los alumnos de primer año ($p < 0.001$), y 31% de la varianza en el caso de solo los aprobados ($n = 115$, $p < 0.001$).

Los resultados del análisis de regresión logística para 2013 y 2014 por componente y subcomponente se registran en la **tabla 5**.

Para conocer la relación del resultado del proceso de admisión con el criterio (calificaciones de primer año) se realizó análisis de senderos, que mostró que el puntaje del primer año sí predice fuertemente el puntaje en el segundo año; sin embargo, incluir los puntajes del examen de admisión no añade fuerza de predicción significativa a la basada únicamente en el puntaje del primer año.

III. Utilidad de las acciones

Para comprobar que la interpretación de los resultados es adecuada como fuente de evidencia de la utilidad de las acciones, se aplicó una encuesta en septiembre y octubre de 2021 (**anexo**) a los alumnos aceptados en los años 2013 y 2014³³. Sus objetivos fueron: conocer el nivel de aceptabilidad con el que los sustentantes perciben el cumplimiento de los instructivos, y medir la satisfacción con las etapas del proceso de admisión.

Los encuestados de ambas generaciones tenían entre 16 y 21 años cuando presentaron el examen de admisión, y entre 25 y 29 años edad cuando contestaron la encuesta. Respondieron 78 personas, 40 mujeres y 38 hombres. El instrumento se divide en tres momentos: antes del proceso de admisión, durante la aplicación de los exámenes, y posterior a la evaluación. Con respecto al primer momento, entre 68 y 93% de los encuestados está satisfecho o muy satisfecho con los aspectos relacionados con la convocatoria y la utilidad del Instructivo para Aspirantes de Nuevo Ingreso; el grado de satisfacción más bajo fue con la utilidad de la Guía temática del Examen de Conocimientos de la Facultad de Medicina, para conocer qué se va a evaluar (48.7%) y la relación de la guía de estudios y la bibliografía con el contenido de los exámenes (52.5%). Entre 78 y 97% de los encuestados está satisfecho o muy satisfecho con los aspectos del momento de la aplicación de los exámenes de admisión, excepto por la comodidad del mobiliario (65.3%). En el momento posterior a la evaluación, entre 61.5 y 67.9% conocían el número de espacios disponibles, cómo se califican los exámenes y cómo se conforman las listas de aceptados antes de presentar las evaluaciones. No se aplicó una encuesta semejante a los académicos o administrativos a cargo, en virtud de que muchos de ellos ya no laboran en la institución.

Para saber si los alumnos tienen éxito académico en la carrera se hizo un seguimiento de la trayectoria académica por generación. De los 145 alumnos que ingresaron en 2013, 89 terminaron hasta el internado; de ellos 77 aprobaron el EGEL de Medicina General (Examen de Egreso de Licenciatura del CE-NEVAL - primera parte del examen profesional), y 72 aprobaron el ECOE (Examen Clínico Objetivo Estructurado - segunda parte del examen profesional) y se titularon a tiempo. De los 145 alumnos que ingresaron en 2014, 80 terminaron hasta el internado, 77 aprobaron el EGEL y 72 aprobaron el ECOE y se titularon a tiempo.

Evaluar la evidencia y formular un juicio

Finalmente, los resultados se analizan para producir un juicio acerca del grado de validez de los usos e interpretaciones de las puntuaciones al determinar la aceptación o rechazo de las hipótesis³⁴. Las fuentes

de evidencia de validez proporcionaron información relevante para informar y mejorar el proceso de admisión de 2013 y 2014, aunque hay amplias áreas de oportunidad en el proceso, mismas que se anotan al final de la discusión.

DISCUSIÓN

La validez de los procesos de admisión es un área poco estudiada en México, este estudio de validación identificó datos e información de diversas fuentes de evidencia de validez en el proceso de admisión de una escuela de medicina de una universidad pública, encontrando evidencias de soporte para las inferencias que se realizan como resultado del proceso, pero también áreas de oportunidad. Para conocer el grado de validez de los usos e interpretaciones de las puntuaciones se analizaron los datos de los procesos de admisión de 2013 y 2014 por medio de un modelo que integra tres marcos de referencia. Se dividió el proceso en tres etapas y en cada una se establecieron hipótesis con base en las inferencias descritas por Kane, las que se probaron por medio de las fuentes de evidencia de Messick. En esta sección se comenta la evidencia a favor y en contra de la validez de la interpretación de los resultados del proceso de admisión, así como las evidencias faltantes.

La evidencia de validez a favor del proceso de admisión estudiado es principalmente la basada en la estructura interna (correlaciones entre componentes y el AFC), la relación entre el examen de conocimientos y los resultados en el primer año, así como la satisfacción de los usuarios. Solo se encontraron correlaciones fuertes (+0.7 a +0.9, -0.7 a -0.9)^{35,36} entre subcomponentes de EP y de EXANI-II (matemáticas) en 2014. Existe relación moderada (+0.4 a +0.6, -0.4 a -0.6) entre los subcomponentes de ciencias naturales del EC y entre los subcomponentes de matemáticas del EXANI-II, y débil (+0.1 a +0.3) o moderada entre los subcomponentes que evalúan constructos semejantes dentro del EP. Sobre todo en 2014, el factor *g* tuvo una fuerte influencia en la varianza de las calificaciones de los subcomponentes orientados a las ciencias naturales y el de inglés. La influencia de este factor sobre el subcomponente de Inglés puede deberse a que en este se evalúa el dominio de la lengua y también se hacen preguntas sobre otros subcomponentes en inglés. Además, la

influencia de este factor sobre los subcomponentes del EP fue moderada (2013) o baja (2014), y sobre los del EXANI-II moderada (2013) o alta (2014). En contraste, en un estudio en que se compararon dos versiones del MCAT (*Medical College Admission Test*) de EUA, encontraron correlaciones moderadas entre componentes que miden constructos semejantes y además identificaron cuatro factores que nombraron conocimiento biomédico, conceptos de ciencia básica, razonamiento cognitivo y desempeño general⁴.

La diferencia de coeficientes entre hombres y mujeres con respecto a Biología, en 2014, es interesante, ya que en otras publicaciones se han reportado con respecto a ciencias naturales en general o a Física. Además, en estos análisis se ha observado que entre los aspirantes hay más mujeres que hombres, pero son admitidos más hombres que mujeres (lo que se observa en la generación de 2014). Puede tratarse de varianza irrelevante al constructo, tal vez producida por ansiedad u otras variables que, aunque impactan en el desempeño de las mujeres en evaluaciones de alto impacto, no implican un mal desempeño académico posterior³⁷⁻⁴².

En cuanto a la relación con otras variables, el dominio de relación de la prueba con el criterio se examinó a través de regresión logística. El componente de conocimientos produce gran varianza en las calificaciones del primer año (por cada punto en el EC, el promedio sube 0.025 en primer año en 2013 y 0.030 en 2014), mientras que el EP tiene influencia negativa; y el EXANI-II, nula. En la literatura se encuentran resultados variables que incluyen mayor impacto del razonamiento verbal sobre el desempeño académico durante y hasta el final de la carrera⁴³, o mayor importancia de los conocimientos científicos (semejante al EC) sobre el desempeño académico durante la carrera⁴⁴. Si bien cada institución educativa decide los elementos que toma en cuenta en sus procesos de admisión, resultados como el encontrado en este estudio sobre los instrumentos EP y EXANI-II (o herramientas similares), motiva a reflexionar sobre su pertinencia y aplicación, así como la interpretación y ponderación de sus resultados.

La satisfacción de los estudiantes con respecto del proceso de admisión fue buena en general; el bajo resultado en nivel de satisfacción con la guía del examen es motivo de reflexión sobre su elaboración y la infor-

mación que contiene. El acceso a materiales de ayuda puede ser motivo de inequidad entre los aspirantes, favoreciendo a quienes tienen mejor situación económica⁴⁵. Desde el punto de vista de aceptabilidad del examen de conocimientos como prueba de admisión, hay reportes con resultados mixtos; es más frecuente que les parezca irrelevante el examen de razonamiento no verbal⁴⁶. La verificación de la calidad de los ítems debe hacerse para probar que no hay varianza irrelevante al constructo, por medios como poner ítems a prueba². Solo el EXANI-II utiliza ítems a prueba, pero el EC, que sí causa varianza en las calificaciones del primer año, no aporta información al respecto. Otra fuente débil de evidencia es la de confiabilidad, ya que solo contamos con el reporte de la UASLP en 2014, pero no tenemos datos de 2013. Esto muestra la importancia de contar con archivos históricos y documentación formal objetiva de este tipo de procesos en las instituciones educativas de nuestro país, con el objetivo de proveer información organizada que permita realizar consultas e investigación.

En cuanto a la representación del dominio para extrapolación, aunque en Estados Unidos se ha enfatizado la importancia de evaluar los campos del conocimiento de ciencias naturales y la capacidad de razonamiento verbal y matemático para el proceso de admisión a medicina⁴⁷, no se cuenta con este tipo de análisis en nuestras instituciones. Otros estudios sugieren llevar a cabo pruebas de juicio de situaciones para aportar escenarios más auténticos de la vida real^{48,49}.

Se observó en este estudio que el EXANI-II tiene un efecto nulo en las calificaciones del primer año y el EP lo tiene negativo en 2013; esto podría significar que altas puntuaciones en el EP predicen menor desempeño en el primer año, lo que podría confirmarse o descartarse con otras investigaciones. Diferentes estudios de validez predictiva con respecto de pruebas de admisión han mostrado que la relación entre el proceso de admisión y las calificaciones se va perdiendo después del primer año, pues existen muchas variables que tienen impacto sobre las calificaciones de la carrera^{39,50}. Sin embargo, en la literatura existe evidencia del poder predictivo del examen de razonamiento verbal y de pruebas semejantes al EC⁷, así como de la prueba de razonamiento verbal por sí sola⁴³.

Si se cumplieran los objetivos del EP y del EXANI-II, se estarían admitiendo estudiantes aptos para estudiar y con gran potencial para aprender cosas nuevas, que idealmente los llevaría a tener un buen desempeño académico. Nuestros resultados no demuestran esta predictibilidad, por lo que sugerimos reevaluar el desarrollo del EP, así como proporcionar más información a los sustentantes acerca de su estructura y de la modalidad de las preguntas. En cuanto al EXANI-II, su diseño ha cambiado por completo recientemente, por lo que sería necesario llevar a cabo un análisis similar con la nueva versión.

Faltó evidencia documental que demostrara la definición del dominio, su revisión por expertos, así como demostración de que la combinación de constructos evaluados es la apropiada. Además faltaron datos para evaluar confiabilidad, discriminación y funcionamiento diferencial de los ítems, entrevistas cognitivas.

En el caso estudiado, el proceso de admisión es con propósitos de selección, por lo que se aplica la evaluación con referencia a norma; sin embargo, sería deseable demostrar que las tres pruebas discriminan entre aspirantes de alto y bajo desempeño. El ingreso a las escuelas y facultades de medicina en México y países similares es un asunto complejo de discutir y analizar, ya que no es factible aplicar la evidencia académica publicada sobre evaluación sumativa de alto impacto en organizaciones que tienen múltiples factores que inciden en el proceso, como presupuesto, espacios físicos, número y competencia de los profesores, entre otros muchos. Estos factores son determinantes de la cantidad máxima de estudiantes que se pueden recibir⁵¹⁻⁵⁴.

Limitaciones

Los procesos de admisión que se evaluaron ocurrieron hace 9 y 10 años, por lo que parte de la información ya no se encuentra disponible o podría presentar diversos sesgos como el de tiempo (por ejemplo, las respuestas de la encuesta de satisfacción). La muestra para el análisis de regresión es pequeña, y va disminuyendo conforme se avanza en la carrera. Al mismo tiempo, el efecto de la restricción de rango (solo analizamos los resultados de los alumnos admitidos, los 145 sustentantes con mayor calificación del proceso) implica que los resultados

del análisis de regresión, por ejemplo, puedan parecer no significativos, aunque para el tamaño de la muestra sean adecuados. En instituciones educativas existen situaciones de normatividad y manejo de datos personales que dificultan el acceso a los datos crudos y a los procesos de validación. El estudio se realizó en una sola institución pública, por lo que los resultados no necesariamente son extrapolables a otras universidades públicas o privadas, nacionales o extranjeras.

CONCLUSIONES

Los exámenes de alto impacto para las profesiones de la salud tienen consecuencias sobre varios sectores de la población, por lo que es responsabilidad de las y los líderes institucionales garantizar que estos sean justos y equitativos, así como que la interpretación de sus resultados sea válida. La validación a través de la aplicación de un modelo que integra los marcos de referencia modernos permite identificar áreas de oportunidad para análisis futuros, al demostrar cuáles datos deben recuperarse de inmediato y las medidas a tomar durante el desarrollo y la aplicación de las pruebas. Se requiere de un equipo interdisciplinario, con planeación adecuada.

La validación con el modelo propuesto es un proceso sistemático que permite identificar claramente las áreas de oportunidad, por medio de la recolección de evidencias en el orden con que se generan. El uso de marcos de referencia como el utilizado aporta bases más sólidas a las conclusiones a las que se llegan y permite aportar sugerencias útiles a las autoridades correspondientes, tales como mejorar las guías para los aspirantes, revisar la pertinencia y el desarrollo de los componentes del proceso de admisión, llevar a cabo estudios de validación y compartir los resultados, así como asegurar la recolección de datos para obtener evidencia de calidad.

CONTRIBUCIÓN INDIVIDUAL

- BACA: Diseño del estudio, recolección, análisis e interpretación de datos, elaboración y revisión del primer borrador y del reporte final.
- IL: Análisis estadístico, interpretación de datos, revisión de primer borrador y del reporte final.
- JATM: Revisión de primer borrador y del reporte final.

- MSM: Diseño del estudio, interpretación de datos, revisión de primer borrador y del reporte final.

AGRADECIMIENTOS

A la Facultad de Medicina de la UASLP.

PRESENTACIONES PREVIAS

Ninguna.

FINANCIAMIENTO

Ninguno.

CONFLICTO DE INTERESES

Ninguno. 🔍

REFERENCIAS

1. Instituto Nacional para la Evaluación de la Educación. Criterios Técnicos Para El Desarrollo, Uso y Mantenimiento de Instrumentos de Evaluación. 2017. <http://tinyurl.com/2cnrfs7w>
2. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. 6th ed. American Educational Research Association, American Psychological Association & National Council on Measurement in Education; 2014.
3. Ferrara S. Our Field Needs a Framework to Guide Development of Validity Research Agendas and Identification of Validity Research Questions and Threats to Validity. *Measurement (Mahwah NJ)*. 2007;5(2-3):156-164. doi:10.1080/15366360701487500
4. Violato C, Gauer JL, Violato EM, Patel D. A study of the validity of the new MCAT exam. *Academic Medicine*. 2020; 95(3):396-400. doi:10.1097/ACM.0000000000003064
5. Paton LW, McManus IC, Cheung KYF, Smith DT, Tiffin PA. Can achievement at medical admission tests predict future performance in postgraduate clinical assessments? A UK-based national cohort study. *BMJ Open*. 2022;12(2):1-12. doi:10.1136/bmjopen-2021-056129
6. Simpson PL, Scicluna HA, Jones PD, et al. Predictive validity of a new integrated selection process for medical school admission. *BMC Med Educ*. 2014;14(1):1-10. doi:10.1186/1472-6920-14-86
7. Bala L, Pedder S, Sam AH, Brown C. Assessing the predictive validity of the UCAT—A systematic review and narrative synthesis. *Med Teach*. 2022;44(4):401-409. doi:10.1080/0142159X.2021.1998401
8. Barajas-Ochoa A, Ramos-Remus C, Castillo-Ortiz JD, et al. Flaws in the design of the Examen Nacional para Aspirantes a Residencias Médicas produce inequity. *Salud Pública Mex*. 2019;61(2):125-135. doi:10.21149/9790
9. Crawford C, Black P, Melby V, Fitzpatrick B. An exploration of the predictive validity of selection criteria on progress outcomes for pre-registration nursing programmes—A systematic review. *J Clin Nurs*. 2021;30(17-18):2489-2513. doi:10.1111/jocn.15730
10. Cunningham C, Patterson F, Cleland J. A literature review of the predictive validity of European dental school selection methods. *European Journal of Dental Education*. 2019; 23(2):73-87. doi:10.1111/eje.12405
11. Hafferty FW, O'Brien BC, Tilburt JC. Beyond High-Stakes Testing: Learner Trust, Educational Commodification, and the Loss of Medical School Professionalism. *Academic Medicine*. 2020;95(6):833-837. doi:10.1097/ACM.0000000000003193
12. Panda N, Bahdila D, Abdullah A, Ghosh AJ, Lee SY, Feldman WB. Association between USMLE Step 1 Scores and In-Training Examination Performance: A Meta-Analysis. *Academic Medicine*. 2021;96(12):1742-1754. doi:10.1097/ACM.0000000000004227
13. Salehi PP, Azizzadeh B, Lee YH. Pass/Fail Scoring of USMLE Step 1 and the Need for Residency Selection Reform. *Otolaryngology - Head and Neck Surgery (United States)*. 2021; 164(1):9-10. doi:10.1177/0194599820951166
14. Villegas Vizcaino R. ¿Es el EXANI II un predictor o un factor de exclusión social? *Educación y Ciencia*. 2017;6(48):43-52. Accessed July 11, 2023. <https://bit.ly/42s5bgS>
15. Bárcquez Antillán IL, Vivian Mascareño MF. La validez predictiva de la prueba de aptitud académica (PAA) respecto al desempeño académico de los estudiantes de la Universidad La Salle Noroeste, A.C. In: Consejo Mexicano de Investigación Educativa AC, ed. X Congreso Nacional de Investigación Educativa. Consejo Mexicano de Investigación Educativa, A.C.; 2009. Accessed July 12, 2023. <https://bit.ly/490R58T>
16. Morales Ibarra R, Barrera Baca A, Mandujano Garnett E. Validez predictiva y concurrente del EXANI-II, en la Universidad Autónoma del Estado de México. In: Consejo Mexicano de Investigación Educativa AC, ed. X Congreso Nacional de Investigación Educativa. Consejo Mexicano de Investigación Educativa, A.C.; 2009. Accessed July 12, 2023. <https://bit.ly/3SLdXDG>
17. Martínez Villarreal RT. Valor Predictivo Del Examen Nacional de Ingreso En La Licenciatura de Medicina En La Universidad Autónoma de Nuevo León, México. Facultad de Medicina de la Universidad Complutense de Madrid; 2014. Accessed July 12, 2023. <https://hdl.handle.net/20.500.14352/38125>
18. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281-302. doi:10.1037/h0040957
19. Messick S. Validity of test interpretation and use. In: *Encyclopedia of Educational Research*. 6th ed. Macmillan; 1990:1487-1495. <https://doi.org/10.1002/j.2333-8504.1990.tb01343.x>
20. Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd ed. Macmillan; 1989:13-104. doi:10.1002/j.2330-8516.1987.tb00244.x
21. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: A practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575. doi:10.1111/medu.12678
22. Kane M. Validating score interpretations and uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing*. 2012;29(1):3-17.

- doi:10.1177/0265532211417210
23. Kane MT. An argument-based approach to validity. *Psychol Bull.* 1992;112(3):527-535. doi:10.1037/0033-2909.112.3.527
 24. Carrillo Avalos BA, Leenen I, Sánchez Mendiola M. Un acercamiento a la unificación de los marcos de referencia de validez de Messick y de Kane. In: XVI Congreso Nacional de Investigación Educativa. Consejo Mexicano de Investigación Educativa A.C.; 2021.
 25. Russell M. Clarifying the Terminology of Validity and the Investigative Stages of Validation. *Educational Measurement: Issues and Practice.* 2022;41(2):25-35. doi:10.1111/emip.12453
 26. UASLP. Guía Temática del Examen de Conocimientos Facultad de Medicina. Published online 2014.
 27. Carrillo-Avalos BA, Sánchez Mendiola M, Leenen I. Amenazas a la validez en evaluación: implicaciones en educación médica. *Investigación en Educación Médica.* 2020;(34):100-107. doi:10.22201/facmed.20075057e.2020.34.221
 28. UASLP. Plan Institucional de Desarrollo 2013-2023.; 2014.
 29. UASLP. Instructivo Para Aspirantes de Nuevo Ingreso 2015-2016.; 2015.
 30. CENEVAL. Guía Del Examen Nacional de Ingreso a La Educación Superior (EXANI-II) 2014. 19th ed. Centro Nacional de Evaluación para la Educación Superior, A. C.; 2013.
 31. UASLP. Guía Temática del Examen de Conocimientos Facultad de Medicina. 2017.
 32. CENEVAL. Guía Del Examen Nacional de Ingreso a La Educación Superior (EXANI-II); 2013.
 33. Miranda López F, Pérez Güemes EE, Villamil Serrano E, Márquez Gutiérrez Y, González Chávez GA, Hernández Gómez JE. Encuesta de Satisfacción de Los Procesos de Evaluación de Ingreso y Promoción En Educación Básica y Media Superior 2017; 2018. <https://bit.ly/485bwAe>
 34. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation.* 2016;1(1):1-12. doi:10.1186/s41077-016-0033-y
 35. Dancy C, Reidy J. *Statistics without Maths for Psychology.* 7th ed. Pearson; 2017. www.pearson-books.com
 36. Rios J, Wells C. Evidencia de validez basada en la estructura interna. *Psicothema.* 2014;26(1):108-116. doi:10.7334/psicothema2013.260
 37. Luschin-Ebengreuth M, Dimai HP, Ithaler D, Neges HM, Reibnegger G. Medical University admission test: a confirmatory factor analysis of the results. *Wien Klin Wochenschr.* 2016;128(9-10):376-383. doi:10.1007/s00508-015-0911-7
 38. Habersack M, Dimai HP, Ithaler D, Reibnegger G. Time: an underestimated variable in minimizing the gender gap in medical college admission scores. *Wien Klin Wochenschr.* 2015;127(7-8):241-249. doi:10.1007/s00508-014-0649-7
 39. Ganjoo R, Schwartz L, Boss M, McHarg M, Dobrydneva Y. Predictors of success on the MCAT among post-baccalaureate pre-medicine students. *Heliyon.* 2020;6(4). doi:10.1016/j.heliyon.2020.e03778
 40. Haladyna TM, Downing SM. Construct-Irrelevant Variance in High-Stakes Testing. *Educational Measurement: Issues and Practice.* 2004;23(1):17-27. <https://bit.ly/49nXOtd>
 41. Leiner JEM, Scherndl T, Ortner TM. How do men and women perceive a high-stakes test situation? *Front Psychol.* 2018;9(DEC). doi:10.3389/fpsyg.2018.02216
 42. Arenas A, Calsamiglia C. Gender Differences in High-Stakes Performance and College Admission Policies. 2022. <http://hdl.handle.net/10419/265771>
 43. Violato C, Donnon T. Does the medical college admission test predict clinical reasoning skills? A longitudinal study employing the Medical Council of Canada clinical reasoning examination. *Acad Med.* 2005;80(10):S14-S16. doi:10.1111/suppl/S14 [pii]
 44. Davies DJ, Sam AH, Murphy KG, Khan SA, Choe R, Cleland J. BMAT's predictive validity for medical school performance: A retrospective cohort study. *Med Educ.* 2022;56(9):936-948. doi:10.1111/medu.14819
 45. Lambe P, Waters C, Bristow D. The UK clinical aptitude Test: Is it a fair test for selecting medical students? *Med Teach.* 2012;34(8). doi:10.3109/0142159X.2012.687482
 46. Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical education? A systematic review. *Med Educ.* 2016;50(1):36-60. doi:10.1111/medu.12817
 47. Association of American Medical Colleges, Howard Hughes Medical Institute. Scientific Foundations for Future Physicians Report of the AAMC-HHMI Committee.; 2009. Accessed July 9, 2023. <https://bit.ly/3UBXIIt>
 48. Patterson F, Roberts C, Hanson MD, et al. 2018 Ottawa consensus statement: Selection and recruitment to the healthcare professions. *Med Teach.* Published online 2018:1-11. doi: 10.1080/0142159X.2018.1498589
 49. UKCAT Consortium. UCAT Official Guide 2023.; 2023. Accessed July 9, 2023. <https://bit.ly/3w41bal>
 50. Hernández-Mata JM, Hernández-Castro R, Nieto-Caraveo A, Hernández Sierra JF. Factores de riesgo para la deserción de estudiantes en la Facultad de Medicina de la Universidad Autónoma de San Luis Potosí (UASLP), México. *Gac Med Mex.* 2005;141(5):445-447. Accessed July 11, 2023. <https://bit.ly/3w6hHqB>
 51. Royal KD, Guskey TR. Editorial: On the appropriateness of norm-and criterion-referenced assessments in medical education. *Ear Nose Throat J.* 2015;94(7):150-152. doi: 10.1177/014556131509400701
 52. Tavakol M, Dennick R. The foundations of measurement and assessment in medical education. *Med Teach.* 2017;39(10):1010-1015. doi: 10.1080/0142159X.2017.1359521
 53. Lok B, McNaught C, Young K. Criterion-referenced and norm-referenced assessments: compatibility and complementarity. *Assess Eval High Educ.* 2016;41(3):450-465. doi: 10.1080/02602938.2015.1022136
 54. Sireci SG, Greiff S. Editorial: On the importance of educational tests. *European Journal of Psychological Assessment.* 2019;35(3):297-300. doi: 10.1027/1015-5759/a000549

ANEXO

Instrumento para la encuesta de satisfacción y percepción sobre el proceso de admisión

ENCUESTA DE OPINIÓN SOBRE EL PROCESO DE ADMISIÓN

Esta encuesta tiene como objetivo conocer su opinión acerca del proceso de admisión que llevó a cabo para ingresar a la licenciatura en médico cirujano; por favor conteste con respecto a lo que usted sabía y percibió cuando se presentó a estos exámenes. Sabemos que esta evaluación se llevó a cabo hace mucho tiempo, y posiblemente ya no recuerde muchos detalles; aún así, por favor intente responder lo mejor posible. Si realmente no se

acuerda de lo que se le pregunta, puede marcar la opción correspondiente.

Los datos que se obtengan serán utilizados por la M en C Blanca Ariadna Carrillo Ávalos en un estudio de validez relacionado con dicho proceso de admisión realizado en los años 2013 y 2014.

Si tiene alguna duda al respecto o desea conocer los resultados, por favor envíe un correo a carrillo@uaslp.mx.

Sus respuestas son muy valiosas, ¡gracias por contestar esta encuesta!

I. Datos demográficos

- ¿Cuál es su edad? (Variable continua, numérica)
- ¿Cuántos años tenía cuando hizo el examen de admisión para ingresar a la licenciatura en médico cirujano? (Variable continua, numérica)
- ¿Cuál es su sexo? (Variable dicotómica: Hombre/mujer)

II. Momento previo a la evaluación

Por favor indique su nivel de satisfacción acerca de los rubros siguientes, seleccionando una opción:

Rubro	Alto nivel de satisfacción		Bajo nivel de satisfacción		No aplica
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho	
1. El tiempo de emisión y vigencia de la convocatoria para el proceso de admisión					
2. Los medios de difusión de la convocatoria del proceso de admisión					
3. La utilidad del Instructivo para Aspirantes de Nuevo Ingreso para conocer el proceso en general					
4. La claridad con la que la UASLP resolvió las dudas sobre la convocatoria					
5. La atención de la autoridad educativa para realizar el registro, recepción y revisión de la documentación					
6. La utilidad del Instructivo para Aspirantes de Nuevo Ingreso para conocer cómo se va a evaluar					

7. La utilidad de Guía temática del Examen de Conocimientos de la Facultad de Medicina para conocer qué se va a evaluar						
8. El tiempo con el que contó para tener acceso a la bibliografía y guía de estudios						
9. La relación de la guía de estudios y la bibliografía, con el contenido de los exámenes						

III. Momento de aplicación de los exámenes de admisión

Por favor indique su nivel de satisfacción acerca de los rubros siguientes, seleccionando una opción:

Rubro	Alto nivel de satisfacción		Bajo nivel de satisfacción		No me acuerdo
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho	
10. Los aspectos que se evalúan en los exámenes					
11. La precisión de la redacción de los planteamientos en las preguntas					
12. La cantidad total de preguntas del examen					
13. La extensión de las preguntas del examen					
14. La contextualización de las preguntas del examen					
15. La localización de la sede					
16. La accesibilidad de la sede					
17. La comodidad del mobiliario de las aulas					
18. La iluminación y la temperatura de las aulas					
19. La precisión de las indicaciones brindadas por el aplicador durante el examen					
20. La atención del aplicador ante las dudas de los sustentantes					
21. El trato brindado a los sustentantes por el aplicador					

IV. Conocimiento general del proceso

Por favor seleccione una opción con referencia a qué tanto conoce acerca del rubro que se evalúa:

Rubro	Alto nivel de conocimiento		Bajo nivel de conocimiento	
	Sí		No	
22. Sabía el número de lugares que se concursan antes del examen				
23. Sabía cómo se califican los exámenes antes de presentarlos				
24. Cómo se conforman las listas de aceptados				

DIMENSIONES

Las dimensiones abarcan los tres momentos: el momento previo a la evaluación (para conocer el grado de satisfacción de los encuestados en cuanto a los mecanismos e instrumentos con los que contaron para presentar los exámenes de admisión), el momento de aplicación de los exámenes de admisión (para conocer la percepción de los sustentantes en cuanto a la pertinencia de las características de los exámenes y las sedes) y el momento posterior a la evaluación (para conocer si los encuestados conocen cómo se califican los exámenes y cómo se distribuyen los lugares para estudiar).

Momento previo a la evaluación	
Dimensión	Preguntas
Convocatoria	1. El tiempo de emisión y vigencia de la convocatoria para el proceso de admisión
	2. Los medios de difusión de la convocatoria del proceso de admisión
	3. La utilidad del Instructivo para Aspirantes de Nuevo Ingreso para conocer el proceso en general
Atención de la UASLP	4. La claridad con la que la UASLP resolvió las dudas sobre la convocatoria
	5. La atención de la autoridad educativa para realizar el registro, recepción y revisión de la documentación
Utilidad de los documentos referentes a la evaluación	6. La utilidad del Instructivo para Aspirantes de Nuevo Ingreso para conocer cómo se va a evaluar
	7. La utilidad de Guía temática del Examen de Conocimientos de la Facultad de Medicina para conocer qué se va a evaluar
Bibliografía y guía de estudios	8. El tiempo con el que contó para tener acceso a la bibliografía y guía de estudios
	9. La relación de la guía de estudios y la bibliografía, con el contenido de los exámenes

Momento de aplicación de los exámenes de admisión	
Dimensión	Preguntas
Exámenes	10. Los aspectos que se evalúan en los exámenes
	11. La precisión de la redacción de los planteamientos en las preguntas
	12. La cantidad total de preguntas del examen
	13. La extensión de las preguntas del examen
	14. La contextualización de las preguntas del examen
Sede	15. La localización de la sede
	16. La accesibilidad de la sede
	17. La comodidad del mobiliario de las aulas
	18. La iluminación y la temperatura de las aulas
Aplicadores	19. La precisión de las indicaciones brindadas por el aplicador durante el examen
	20. La atención del aplicador ante las dudas de los sustentantes
	21. El trato brindado a los sustentantes por el aplicador

Conocimiento general del proceso	
Dimensión	Preguntas
Información de resultados del proceso de admisión	22. Sabía el número de lugares que se concursan antes del examen
	23. Sabía cómo se califican los exámenes antes de presentarlos
	24. Cómo se conforman las listas de aceptados

Las escalas de Likert correspondientes fueron:

Momento previo a la evaluación (primera fase) y Momento de aplicación de los instrumentos (segunda fase)	
1. Muy satisfecho	Alto nivel de satisfacción
2. Satisfecho	
3. Insatisfecho	Bajo nivel de satisfacción
4. Muy insatisfecho	
Momento posterior a la evaluación (tercera fase)	
1. Si	Nivel de conocimiento adecuado
2. No	Nivel de conocimiento inadecuado