

# Método Integral de Evaluación de Competencias Médicas Profesionales (MIDECOMP) del examen profesional FacMed, UNAM

Facultad de Medicina



Alejandra Navarro-Escalera<sup>a,†,\*</sup>, Carlos Alberto Soto-Aguilera<sup>a,‡</sup>, Ana Ivette Mondragón-Pineda<sup>a,§</sup>, Amílcar Alpuche-Hernández<sup>a,¶</sup>, Rocío García-Durán<sup>a,μ</sup>, Mauricio Pilar-Díaz<sup>a,β</sup>, Antonio Cerritos<sup>a,θ</sup>, Armando Ortíz-Montalvo<sup>a,Δ</sup>



## Resumen

**Introducción:** El Examen Profesional fase teórica (EP-FT) evaluó la adquisición de cinco de las competencias establecidas en el perfil de egreso del Plan de Estudios 2010 (PE-2010) de la Facultad de Medicina (FM) UNAM, empleó el Método Integral De Evaluación de Competencias Médicas Profesionales (MIDECOMP).

**Objetivo:** Describir la metodología que se utilizó para el diseño, elaboración, aplicación y análisis del EP-FT de enero 2024 de la Licenciatura de Médico Cirujano de la FM.

**Método:** Este instrumento de evaluación se diseñó con base en una estrategia que integró el análisis secuenciado de las competencias, sus atributos, los resultados de aprendizaje y el nivel taxonómico, en un contexto de la práctica de la medicina general. El diseño, elaboración

y revisión cruzada de los casos y reactivos, se realizaron por comités inter y multidisciplinarios. Para ello y para el ensamblaje de la prueba se utilizó el Sistema Automatizado para Bancos de Reactivos en Medicina (SABERMED). Se aplicó el examen con el Sistema Integral de Aplicación de Exámenes (SIAEX), ambos creados en la FM. Los análisis psicométricos establecieron parámetros que dieron evidencia de la calidad del instrumento y permitieron hacer inferencias sobre los resultados.

**Resultados:** Se utilizó el MIDECOMP para construir el EP-FT. Se logró elaborar un instrumento de evaluación de 300 casos clínicos con reactivos de opción múltiple y tres opciones de respuesta, de acuerdo con el panorama epidemiológico nacional. Se integraron secuencialmente elementos conceptuales, logísticos, soluciones informá-

<sup>a</sup> Departamento de Evaluación Educativa, Secretaría de Educación Médica, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México.

ORCID ID:

<sup>†</sup> <https://orcid.org/0000-0001-5935-3052>

<sup>‡</sup> <https://orcid.org/0000-0003-1104-2962>

<sup>§</sup> <https://orcid.org/0000-0001-8809-5179>

<sup>¶</sup> <https://orcid.org/0000-0002-5388-7632>

<sup>μ</sup> <https://orcid.org/0009-0009-7422-9960>

<sup>θ</sup> <https://orcid.org/0009-0001-8228-7804>

<sup>θ</sup> <https://orcid.org/0000-0001-9346-8793>

<sup>Δ</sup> <https://orcid.org/0000-0003-3548-1513>

Recibido: 11-octubre-2024. Aceptado: 17-diciembre-2024.

\* Autor para correspondencia: Alejandra Navarro Escalera. Departamento de Evaluación Educativa, Secretaría de Educación Médica, Facultad de Medicina, UNAM, Cd. Mx., México.

Correo electrónico: [alnavarro.degas@gmail.com](mailto:alnavarro.degas@gmail.com)

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

ticas y análisis psicométricos que permitieron alcanzar los estándares de calidad para evaluar eficientemente a 1,312 sustentantes. Se obtuvo una consistencia interna (alfa de cronbach) 0.90, dificultad (media de P) 0.63 y discriminación (media de Rpbis) 0.17, que representaron valores adecuados.

**Conclusiones:** La implementación del MIDECOMP permitió diseñar, elaborar, aplicar y analizar el EP-FT del 2024 de manera masiva, eficiente y con estándares de calidad, alineado al PE-2010 de la FM, UNAM.

**Palabras clave:** Competencias; MIDECOMP (siglas en español); metodología; evaluación; altas consecuencias.

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

## Medical Professional Competencies Integral Assessment Method (MIDECOMP) of a high-stakes examination at the FacMed, UNAM

### Abstract

**Introduction:** The theoretical phase of the high-stakes examination (EP-FT) evaluated the acquisition of five competencies established in the graduate profile described in the 2010 Curriculum (PE-2010) of the Faculty of Medicine (FM) UNAM, used the Medical Professional Competencies Integral Assessment Method (MIDECOMP).

**Objective:** Describe the method used for the design, elaboration, delivery, and analysis of the EP-FT of January of 2024 of the Medical Degree of FM.

**Method:** This assessment tool was designed based on a strategy that integrated the sequential analysis of the com-

petencies, their attributes, the learning outcomes, and the taxonomic level, in a generalist medicine based practice environment. The design, development, and crossover review of the cases and the items was performed by inter and multidisciplinary committees. Therefore, for this task and the assembly of the test, we used the Automated System for Medicine Item Bank (SABERMED). For the delivery of the test we utilized the Integral System For Test Delivery (SI-AEX), both of these systems were developed by the FM. The psychometric analysis established the parameters that showed the evidence of the quality of the tool and allowed us to make inferences of the results.

**Results:** The MIDECOMP was used to construct the EP-FT. We were able to design an assessment tool composed of 300 clinical cases with multiple-choice questions with three different options, based on the national epidemiological outlook. We integrated sequentially conceptual and logistics elements, informatic solutions, and psychometric analyses that let us achieve the quality standards to assess 1,312 candidates. We obtained an internal consistency (Cronbach's Alpha) of 0.90, difficulty (P mean) of 0.63, and a discrimination index (Rpbis) of 0.17; these values represented adequate values.

**Conclusions:** The implementation of the MIDECOMP permitted the design, elaboration, delivery, and analysis of the 2024 EP-FT in a massive and efficient manner, with high quality standards, aligned with the PE-2010 of the FM, UNAM.

**Keywords:** Competencies; MIDECOMP (Spanish acronym); methodology; assessment; high-stakes examination.

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

## INTRODUCCIÓN

Históricamente, la evaluación ha sido un elemento fundamental en el proceso educativo y en la toma de decisiones en el ámbito profesional y laboral. A lo largo del tiempo, ha ido evolucionando la comprensión de lo que constituye una evaluación efectiva y justa.

Una clasificación tradicional de la evaluación educativa se centra en su objetivo el cual puede ser diagnóstica, sumativa y formativa.

La evaluación diagnóstica se realiza al principio de un curso o una actividad académica con la finalidad de determinar el nivel de conocimiento, habilidad o actitud del educando. La información obtenida puede ser de utilidad para hacer adecuaciones al contenido o a las actividades académicas planeadas previamente.

La evaluación sumativa se compone de la suma de las valoraciones efectuadas durante un curso para

determinar, al final del mismo, el grado con el que se alcanzaron los objetivos de la enseñanza y con ello otorgar calificaciones.

La evaluación formativa se utiliza para monitorear el progreso del aprendizaje y proporcionar realimentación al estudiante sobre sus logros, deficiencias y áreas de mejora<sup>1</sup>.

Un examen profesional es una evaluación sumativa y de altas consecuencias. El último término es abordado en el documento *Standards for Educational and Psychological Testing* del 2014, como aquellas evaluaciones que tienen implicaciones significativas para los individuos evaluados, por ejemplo la obtención de un título, el acceso a oportunidades educativas o profesionales, y otras decisiones importantes basadas en los resultados de la evaluación. Dichas evaluaciones requieren una gran atención en términos de validez y justicia, ya que los resultados afectan de manera directa y profunda la vida de los individuos<sup>6</sup>.

A estas dos clasificaciones de los tipos de evaluación de un Examen Profesional se agrega la evaluación de competencias, ya que el Examen Profesional fase teórica (EP-FT) de la Facultad de Medicina (FM) de la UNAM, se basa en ellas para el diseño del examen.

Las competencias representan un constructo multivariado conformado por conocimientos, habilidades (cognitivas, destrezas) y actitudes, por lo que su evaluación se hace bastante compleja. Podemos afirmar que no existe un instrumento que evalúe esta multidimensionalidad de la competencia médica, por lo que su evaluación se ve fragmentada en instrumentos específicos para cada una de estas dimensiones.

A nivel internacional, la pirámide de Miller ha servido de guía para identificar los niveles básicos y superiores de las competencias en medicina<sup>3</sup>. Lo anterior ha orientado a los evaluadores para seleccionar los instrumentos de evaluación más adecuados para cada uno de sus niveles (1. Saber, 2. Saber hacer, 3. Demostrar cómo se haría, 4. Hacer).

Para la evaluación los dos primeros niveles se pueden utilizar: preguntas abiertas, de correlación, de opción múltiple, de completamiento o ensayos, entre otros. Para los niveles 3 y 4 por lo general se utilizan estrategias a base de la observación del desempeño apoyadas por: listas de cotejo, listas de

apreciación, rúbricas, mini cex y últimamente el uso de portafolios electrónicos de evidencias y los Exámenes Clínicos Objetivos Estructurados (ECO)<sup>4</sup>.

Cortés T. y cols. también han desarrollado otros instrumentos para tratar de evaluar competencias más complejas como es el caso del cuestionario para evaluar el desarrollo y crecimiento personal de los Alumnos<sup>5</sup>.

En la FM de la UNAM, se tienen antecedentes de la evaluación del área de conocimientos en los exámenes profesionales mediante el sorteo de temas y el cuestionamiento al sustentante por el sínodo. En la década de los 70 con la masificación de la matrícula se optó por la formulación de preguntas aisladas de opción múltiple con cinco opciones de respuesta.

En la década de los 80 se introdujeron casos clínicos con cinco o más preguntas, cada una con cinco opciones de respuesta. A la implementación del Plan de Estudios 2010 (PE-2010) de la FM, el EP-FT se enfocó en la evaluación de competencias y en los últimos diez años basados en la literatura, se ha privilegiado al apoyo de pequeñas viñetas clínicas acompañadas de una o dos preguntas de opción múltiple con tres opciones de respuesta.

En el mundo existen varios modelos de competencias en educación médica como el modelo de la *Canadian Medical Education Directives for Specialists* (CanMEDS), fue creado en 2005 estableciendo un marco de competencias compuesto de 28 competencias principales y 126 habilidades específicas.

Asimismo, el *Accreditation Council for Graduate Medical Education* (ACGME) establece su propio modelo de competencias en seis dominios diferentes y se conjuga con el *Milestone Project* para definir criterios de desempeño específicos para avanzar al siguiente nivel de habilidad o conocimiento. Este último modelo se alinea más al del Método Integral De Evaluación de Competencias Médicas Profesionales (MIDECOMP) que utiliza la FM, para el EP-FT<sup>2</sup>.

Después del reto que implica la evaluación por competencias también nos enfrentamos a nuevos desafíos en los procesos y metodologías de evaluación como el que representa el continuo avance de las ciencias médicas.

Otro de los desafíos más relevantes lo constituye el advenimiento de la Inteligencia Artificial Generativa con su alto potencial de transformación de modelos e instrumentos de evaluación. Es necesario

conocer la herramienta tecnológica para encontrar el nicho adecuado de su utilización.

Un reto más lo constituye el equilibrio entre la evaluación formativa y la sumativa; es ideal para el estudiante contar con evaluaciones de competencias para su formación, antes de entrar a resolver exámenes de altas consecuencias.

De manera evidente, la formación de los evaluadores en este tipo de competencias es indispensable para su amplia participación, tanto en la conformación de viñetas clínicas como en el ejercicio de la calificación, tanto en tiempo real (como sucede con los ECOEs) o posterior a la evaluación, como sucede con los exámenes de tipo ‘respuesta construida’ en donde se capturan las respuestas de los sustentantes para su posterior calificación por parte de los evaluadores.

Por último, también es importante incorporar competencias emergentes como el interprofesionalismo, la evaluación de habilidades blandas, el contexto ético e inclusivo y las competencias espirituales.

De acuerdo con lo antes mencionado la FM, retomó algunos criterios para la elaboración del EP-FT, con la finalidad de seguir mejorando sus estándares de validez y confiabilidad, manteniendo una evaluación justa y objetiva. Iniciando con las recomendaciones de la Conferencia de Ottawa de 2010<sup>7</sup>, donde se reunieron expertos en evaluación para discutir y establecer las siguientes recomendaciones, clave para una evaluación de calidad y efectiva:

1. Claridad y transparencia en los criterios de evaluación: establecer criterios claros y transparentes, que fueran comprensibles para los evaluadores y para los evaluados, con el fin de garantizar una evaluación justa y objetiva.
2. Adaptación al contexto: se subrayó la necesidad de adaptar los criterios de evaluación al contexto específico, reconociendo las diferencias entre distintos entornos educativos y profesionales.
3. Validez y confiabilidad: que las evaluaciones midieran de manera precisa lo que pretendieron medir y que los resultados fueran consistentes y reproducibles.
4. Inclusión de múltiples fuentes de evidencia: se recomendó el uso de diversas fuentes de evidencia, como exámenes escritos, proyectos y desem-

peño práctico del sustentante, para obtener un perfil completo y preciso del rendimiento de los evaluados. El EP de la FM, cuenta con una fase práctica en modalidad ECOE, que se lleva a cabo después del EP-FT.

5. Realimentación constructiva: haber proporcionado realimentación específica, oportuna y constructiva a los evaluados, con el objetivo de apoyar su aprendizaje y desarrollo profesional.

Estas recomendaciones invitaron a hacer conciencia respecto a las políticas, prácticas y toma de decisiones relacionadas con el tema de evaluación en los programas académicos en los diferentes niveles educativos.

De acuerdo con estas recomendaciones, la FM de la UNAM retomó una serie de evidencias que respaldaron la validez y confiabilidad de su examen EP-FT y que permitieron cumplir con sus objetivos de “valorar los conocimientos del egresado de la carrera de medicina, la aplicación de los mismos y que posea criterio profesional”, así como “certificar los conocimientos y la aplicación de los mismos a un amplio espectro de problemas médicos relevantes a los cuales se enfrentará en su práctica clínica real”<sup>8</sup>.

Además de lo anterior la FM contempló en su Método MIDECOMP, los siguientes criterios para la elaboración del EP-FT:

1. Validez de contenido: los exámenes escritos debieron cubrir adecuadamente el contenido del curso que se evaluó. Esta validez se logró al contar con una robusta y ágil tabla de especificaciones que dirigió la elaboración de reactivos para que abordan los temas importantes y representativos de las asignaturas, competencias o habilidades, según correspondiera.
2. Validez de constructo: consistió en que el examen midiera lo que debía medir. Esta validez se estableció mediante el análisis de la relación entre las puntuaciones del examen y otras medidas que teóricamente debieron estar relacionadas.
3. Validez de criterio: fue la medida en que las puntuaciones del examen se correlacionaron con algún criterio externo que se consideró relevante.
4. Confiabilidad: los exámenes escritos debieron ser consistentes en la medida en que evaluaron a los estudiantes. La consistencia interna y la es-

tabilidad de los resultados a lo largo del tiempo fueron indicativos de confiabilidad.

5. Análisis de ítems: evaluaron la calidad de las preguntas individuales en un examen. Las preguntas debieron ser claras, pertinentes, equitativas y con un nivel de dificultad adecuado. Un análisis de ítems ayudó a identificar preguntas problemáticas que pudieran necesitar revisión o eliminación.
6. Pruebas piloto y revisión continua: antes de su implementación, los exámenes escritos suelen ser sometidos a pruebas piloto para evaluar su efectividad y claridad.
7. Comparaciones con otros métodos de evaluación: se compararon los resultados de los exámenes escritos con otras formas de evaluación, como exámenes prácticos (ECOE), evaluaciones clínicas o evaluaciones de desempeño, que proporcionaron una visión más completa del rendimiento de los estudiantes<sup>7</sup>.

Estas evidencias permitieron que el EP-FT se realizara con las mejores prácticas de evaluación descritas en la literatura.

En un inicio este examen se basó en el perfil epidemiológico resultante de los índices de morbi-mortalidad en México, pero a partir del 2010 se emplearon como base los créditos de cada asignatura y en el 2016 se inició un proceso basado en los 12 pasos para el desarrollo efectivo de una prueba propuesta por Steve Downing<sup>9</sup>.

Actualmente, el comité elaborador coteja los temas y subtemas cursados en la licenciatura de médico cirujano de la FM asegurando la representatividad del banco de reactivos con el perfil epidemiológico nacional ya que los programas académicos de las asignaturas están basados y actualizados en este de acuerdo con la Dirección General de Epidemiología.

El diseño actual del EP-FT, evaluó cinco de las ocho competencias del Plan de Estudios 2010 (PE-2010), con las que se sentaron las bases para la elaboración de la tabla de especificaciones del examen.

## OBJETIVO

Describir la metodología MIDECOMP utilizada para el diseño, elaboración, aplicación y análisis del EP-FT de enero 2024 de la Licenciatura de Médico Cirujano de la FM de la UNAM.

## MÉTODO

### 1. Tabla de especificaciones y definición del contenido

En 2016 se formaron comités colegiados transdisciplinarios para elaborar la tabla de especificaciones del EP-FT en tres fases: la primera de diseño, la segunda de validación y la tercera de revisión. En cada una de estas fases participaron diferentes académicos: 18 médicos especialistas, de ciclos básicos o clínicos y un facilitador psicólogo.

En la primera fase, el comité de diseño se encargó de elaborar el perfil de referencia y la tabla de especificaciones de acuerdo con las competencias y sus atributos descritos en el PE-2010. En seis sesiones de trabajo de tres horas cada una, este grupo diseñó los resultados de aprendizaje para cada atributo de competencia, el nivel taxonómico de los reactivos y el número de reactivos por atributo.

En la segunda fase el comité validó el contenido de la tabla de especificaciones, en cuatro sesiones de tres horas cada una. En la tercera fase, el comité de revisión emitió las recomendaciones necesarias para las adecuaciones finales.

La tabla de especificaciones del EP-FT integró las competencias del PE-2010. Se conformó de los atributos de competencia, resultados de aprendizaje y el nivel taxonómico. Además, cada reactivo se clasificó por habilidad clínica (ej., interrogatorio, exploración física, prevención, profesionalismo, etc.), área clínica (ej., medicina familiar, medicina interna, pediatría, etc.) y de carrera (ej., biomédicas, sociomédicas / humanísticas y clínicas).

Para el sistema de clasificación cognitiva se utilizó la taxonomía de operaciones cognoscitivas de Bloom modificada por Anderson y Krathwohl<sup>10</sup>, con cinco niveles: recordar, comprender, aplicar, analizar y evaluar; sin embargo, para este EP se migró a la taxonomía de Castañeda<sup>11</sup>, con el uso de tres niveles cognitivos: comprender y organizar la información, aplicar conceptos principios y procedimientos y resolver problemas.

### 2. Diseño y composición del instrumento

En el EP-FT de 2024 se evaluaron cinco de las competencias (C) del PE-2010<sup>12</sup> (ej., C1 Pensamiento crítico; C4. Aplicación de las ciencias; C5. Habilidades clínicas; C6. Profesionalismo, aspectos éticos y res-

ponsabilidades legales; C7. Promoción de la salud y prevención de la enfermedad), desglosadas en atributos y resultados de aprendizaje contenidos en la tabla de especificaciones.

Para el diseño del EP-FT de 2024 se ponderó el peso específico con la cantidad de reactivos requeridos por competencia y resultado de aprendizaje, con el formato de reactivos de opción múltiple, cada uno con su caso clínico y tres opciones de respuesta.

Este diseño se fundamentó en los estándares descritos por Haladyna y Rodríguez en 2013<sup>13</sup>; así como, en las cinco reglas básicas descritas por el *National Board of Medical Examiners* (NBME)<sup>14</sup>.

De acuerdo con los estándares para la validez de constructo<sup>15,16</sup> el grupo de redactores se conformó por cinco médicos-académicos con especialización en diferentes áreas clínicas: medicina familiar, general, interna, pediatría y epidemiología, adscritos al Departamento de Evaluación Educativa (DEE) de la Secretaría de Educación Médica (SEM), todos contaron con amplia experiencia en evaluación y en elaboración exámenes de altas consecuencias.

Los dos estándares antes descritos permitieron a este grupo esquematizar de forma clara y precisa el proceso de elaboración de estos elementos, orientados a las ciencias de la salud, con base en los resultados de aprendizaje de las competencias contenidas en el PE-2010<sup>13,14,17,18</sup>.

Para la composición del EP-FT 2024 se seleccionaron reactivos del banco, con índices psicométricos adecuados (dificultad y discriminación)<sup>19,20</sup> elaborados en los años 2022 y 2023 y se establecieron los reactivos de *novo*. Esta composición se revisó de manera colegiada y confidencial.

### 3. Elaboración de los casos clínicos, las preguntas de los reactivos y las opciones de respuesta

Se asignaron algunos resultados de aprendizaje y niveles cognitivos al grupo de redactores, quienes elaboraron los reactivos de *novo*, para sustituir a aquellos que no cumplieron con los estándares de calidad psicométrica en aplicaciones previas.

Los casos clínicos construidos simulaban el acto médico o aquellos procesos mentales necesarios para aplicar los conocimientos o resolver un problema<sup>20</sup>.

Estos se diseñaron con los estímulos y pistas necesarios para evocar el proceso cognitivo del sustentante y para establecer el contexto clínico requerido. Es importante denotar, que en algunos casos se utilizaron imágenes, gráficos o tablas, haciendo alusión al caso clínico que se presentó<sup>14,17</sup>.

Además, los casos se conformaron por algunos de los siguientes elementos: características sociodemográficas del paciente; motivo de consulta; semiología de síntomas y semiótica de signos; antecedentes relevantes; signos vitales; hallazgos de la exploración física; estudios de laboratorio / gabinete; diagnóstico y terapéutica empleada.

Para mantener la validez de contenido<sup>15,16</sup> los reactivos se enfocaron en los casos clínicos, evitando pistas gramaticales, fueron claros y la redacción fue de forma positiva sin términos vagos, para que el sustentante pudiera aplicar sus conocimientos o resolver el problema planteado de acuerdo con el resultado de aprendizaje.

Estas características evitaron procesos mentales innecesarios y permitieron observar si los sustentantes verdaderamente alcanzaron el resultado de aprendizaje evaluado<sup>13,14,17,18</sup>.

De igual forma, para alcanzar la validez de contenido<sup>15,16</sup> las opciones de respuestas fueron homogéneas, se mantuvo un formato uniforme para las categorías (semiótica). Los distractores fueron plausibles, verosímiles, independientes y mutuamente excluyentes, para atraer a aquellos sustentantes que no conocieran la respuesta correcta<sup>13,14,17,18</sup>.

### 4. Comités y revisiones

Para continuar con la validez de contenido<sup>15,16</sup> posterior a la elaboración de los reactivos, el grupo de redactores se reunió en comité de forma semanal, revisaron cada uno de los casos clínicos, reactivos y opciones de respuesta.

El objetivo fue revisar que cada uno de los reactivos se apegaran al resultado de aprendizaje asignado, que fueran claros, homogéneos y que el contenido técnico-médico fuera correcto.

También, cotejaron que la información estuviera referenciada y justificada de acuerdo con las guías de práctica clínica y las Normas Oficiales Mexicanas, ya que es el marco jurídico del ejercicio profesional de la medicina en México; además, verificaron que

la información fuera consistente con las guías de manejo internacionales más actualizadas.

Asimismo, se verificó que cada caso clínico tuviera su correcta clasificación del CIE-11, con el objetivo de evaluar si los padecimientos que se incluyeron en los casos clínicos correspondían a lo que el médico general se enfrenta con mayor frecuencia.

Concluida esta revisión, cada uno de los miembros del grupo de redactores subió sus reactivos al Sistema de Administración de Bancos de Exámenes y Reactivos Médicos (SABERMED).

A continuación, se realizó la revisión técnica por dos psicólogas con amplia experiencia en evaluación educativa, adscritas al Departamento de Evaluación Educativa (DEE). Utilizaron una lista de cotejo propuesta por Haladyna y modificada por Rivera<sup>21</sup>.

El objetivo de esta revisión fue evaluar que los reactivos cumplieran con los estándares antes mencionados; que existiera congruencia entre la justificación y la opción de respuesta correcta; la plausibilidad de los distractores y que los reactivos técnicamente cumplieran con la clasificación expuesta en la tabla de especificaciones.

Después los reactivos pasaron a revisión disciplinar a cargo de tres médicos con experiencia en el área clínica, externos a la SEM y al grupo de redactores. Antes de iniciar el proceso firmaron un acuerdo de confidencialidad y de no divulgación.

A cada revisor externo le fueron asignados entre 20 y 30 reactivos; específicamente los recién elaborados, modificados o los reutilizados que resultaron mal puntuados psicométricamente en exámenes previos.

El objetivo de la revisión disciplinar fue verificar que la información médica incluida en los casos clínicos y en la respuesta, fueran correctas, también verificar la plausibilidad médica de los distractores, la congruencia entre la pregunta del reactivo y la competencia, atributo y resultado de aprendizaje; y que el contenido del reactivo tuviera relación con el programa de Internado Médico de Pregrado.

A continuación, se muestra un reactivo de este examen profesional, que cumplió con todos los criterios establecidos con el diseño, elaboración, revisión técnica, revisión disciplinar y obtuvo una

puntuación psicométrica adecuada, con dificultad moderada y muy buena discriminación:

Mujer de 29 años con embarazo de 33 SDG y preeclampsia en tratamiento conservador. Para darle el manejo más adecuado tiene a la mano tres fuentes de información:

- a) Metaanálisis reporta que el fármaco de primera elección es labetalol por su rápida acción terapéutica y menos riesgo de comorbilidades.
- b) Guía de práctica clínica la cual sugiere que el nifedipino es el medicamento de elección por consenso de expertos.
- c) Serie de casos y controles refiere que la alfametildopa tiene mejores resultados.

De acuerdo con la mejor evidencia usted indica a la paciente:

- a) Labetalol
- b) Nifedipino
- c) Alfametildopa

La respuesta correcta es la opción a) y este reactivo fue diseñado para evaluar un aspecto de la competencia 1. Pensamiento crítico, juicio clínico, toma de decisiones y manejo de información. Enfocado a determinar la utilidad de la información médica disponible para la resolución de los problemas más frecuentes de salud.

## 5. Ensamblaje de la prueba

El ensamblaje del EP-FT del 2024 se realizó por medio del sistema SABERMED, creado en el año 2018 por los ingenieros adscritos al DEE de la SEM. El objetivo principal de su creación fue mejorar la construcción, almacenamiento, administración y recuperación de la información de nuestros instrumentos de evaluación.

Cabe destacar que el SABERMED permitió que los usuarios elaborarán de forma rápida, eficiente y segura diferentes reactivos de diversos formatos. El sistema demostró características de seguridad y privacidad robustas, además se adaptó a las necesidades particulares de las evaluaciones diseñadas por el DEE, garantizando la integridad de la información.

El SABERMED se vinculó con el Sistema Integral de Aplicación de Exámenes (SIAEX), creado por ingenieros de la FM de la UNAM, con las características de seguridad y privacidad adecuadas.

De forma electrónica el SABERMED transfirió los reactivos a este sistema aplicador. Por consiguiente, el SIAEX fue la interfase utilizada para la aplicación del EP-FT 2024 y este sistema permitió al sustentante visualizar cada uno de los reactivos del examen de manera aleatoria.

## 6. Logística de la aplicación

La aplicación del examen fue planificada por la Coordinación de Logística (CL) para llevarse a cabo en dos días, en dos turnos de cuatro horas cada uno, en modalidad presencial en las instalaciones de la FM. Se utilizaron computadoras portátiles (*chromebooks*) programadas para permitir ingresar exclusivamente al SIAEX para la aplicación del EP-FT 2024.

La logística para la aplicación de la prueba incluyó instrucciones claras y precisas sobre el acceso a la plataforma, validación de la identidad de los sustentantes, restricción del uso del celular, supervisión constante, reactivos aleatorizados, examen cronometrado y condiciones ambientales adecuadas.

## 7. Análisis estadístico

El análisis se realizó en el área de psicometría del DEE, SEM, con el uso del software psicométrico Iteman versión 4 (Assessment System Corporation; Minnesota; EUA). Los resultados se presentan en la sección correspondiente.

Previo al análisis se verificaron los criterios para la utilización de pruebas paramétricas a partir de

la distribución de los puntajes obtenidos, nivel de medición, representatividad de la muestra y homocedasticidad de la varianza.

El análisis de las respuestas de la prueba se realizó con base en los supuestos de la teoría clásica de los tests, obteniendo los siguientes parámetros: consistencia interna de la prueba<sup>22</sup>; dificultad<sup>23</sup> y discriminación<sup>22</sup>, partiendo de una estructura de competencias, seguida de la identificación de áreas de conocimiento, habilidades clínicas y nivel taxonómico.

Después se identificaron aquellos reactivos con inconsistencias en los parámetros antes mencionados y se pasaron al comité del grupo de redactores, quienes realizaron el análisis cualitativo del contenido.

Posteriormente, emitieron un veredicto sobre qué reactivos permanecieron o se eliminaron del conteo final. Para la calificación de los sustentantes se estableció un punto de corte de acuerdo con la distribución de todos los puntajes.

## 8. Consideraciones éticas

Se eliminaron los identificadores individuales de los sustentantes; se presentaron únicamente resultados agregados, se mantuvo el anonimato de los sustentantes en la elaboración del manuscrito.

## RESULTADOS

### 1. Diseño y composición del examen

El EP-FT se conformó por 300 casos clínicos con reactivos de opción múltiple y tres opciones de respuesta, distribuidos en 5 competencias del PE-2010 de la FM de acuerdo con la tabla de especificaciones del examen (**tabla 1**).

**Tabla 1.** Distribución de reactivos por competencias para el EP-FT

Competencias	Reactivos	Ponderación
Competencia 1. Pensamiento crítico, juicio clínico, toma de decisiones y manejo de información	60	20%
Competencia 4. Conocimiento y aplicación de las ciencias biomédicas, sociomédicas y clínicas en el ejercicio de la medicina	90	30%
Competencia 5. Habilidades clínicas de diagnóstico, pronóstico, tratamiento y rehabilitación	90	30%
Competencia 6. Profesionalismo, aspectos éticos y responsabilidades legales	30	10%
Competencia 7. Salud poblacional y sistema de salud: promoción de la salud y prevención de la enfermedad	30	10%
Total	300	100%

Esta tabla muestra la distribución de reactivos asignados a cada competencia.

**Tabla 2.** Niveles taxonómicos del EP-FT

Taxonomía de Anderson et al. (2001)	Reactivos	Taxonomía de Sandra Castañeda (2006)	Reactivos
Crear	0	Resolver problemas	144
Evaluar	144		
Analizar	88	Aplicar conceptos, principios y procedimientos	122
Aplicar	34		
Comprender	33	Comprender y organizar la información	34
Recordar	1		
Total	300	Total	300

Esta tabla muestra los niveles taxonómicos que se exploran en el EP-FT de acuerdo con la tabla de especificaciones y con las cantidades con Anderson y Castañeda.

**Tabla 3.** Análisis de resultados

Elementos	Distribución de reactivos	Media	Desviación estándar	Puntaje mínimo	Puntaje máximo	Consistencia interna, Alpha de Cronbach	Error estándar de la media	Dificultad	Discriminación
Todos los reactivos	281	177.07	22.66	91	233	0.90	7.24	0.63	0.17
Comp. 1	57	34.37	5.66	15	49	0.66	3.28	0.60	0.17
Comp. 4	87	54.94	8.15	27	78	0.75	4.06	0.63	0.17
Comp. 5	84	53.12	8.04	5	74	0.75	4.02	0.63	0.17
Comp. 6	26	17.43	2.84	0	25	0.49	2.03	0.67	0.17
Comp. 7	27	17.21	2.83	0	26	0.42	2.16	0.64	0.15

Esta tabla muestra los resultados de los análisis psicométricos de la prueba, con enfoque por competencias y excluyendo los reactivos eliminados. Los puntajes mínimo y máximo se establecieron de acuerdo con los aciertos obtenidos, en relación al total de reactivos y a la distribución de reactivos por competencia.

Las taxonomías de operaciones cognitivas que se utilizaron para el diseño de los reactivos del examen fueron las que se observan en la (tabla 2).

Las ponderaciones por nivel taxonómico que fueron implementadas, se consideraron adecuadas para un examen profesional, ya que implicó que los sustentantes demostraran principalmente la aplicación de sus conocimientos y la resolución de problemas y dieron evidencia del aprendizaje construido durante su formación académica<sup>11</sup>.

## 2. Aplicación del examen

El examen se llevó a cabo el 8 y 9 de enero de 2024 de acuerdo con lo establecido en el calendario aprobado por el Honorable Consejo Técnico (HCT) de la FM. La aplicación se dividió en 2 días, proporcionando 4 horas y 165 reactivos por día.

La cantidad de sustentantes que presentaron el EP-FT en enero del 2024 fueron 1312, todos pertenecientes a la FM de la UNAM.

## 3. Calificaciones de aprobación

El área de psicometría envió un reporte de los reactivos que tuvieron un grado determinado de dificultad y discriminación al comité de elaboración de reactivos, que analizó de forma colegiada cada reactivo, manteniendo los que cumplieron con criterios adecuados para un instrumento de altas consecuencias<sup>21</sup> y recomendaron eliminar 19 reactivos que no tuvieron índices psicométricos adecuados de discriminación y dificultad.

## 4. Reporte de los resultados de la prueba

La tabla 3 concentró los resultados psicométricos del EP-FT 2024-1, desglosados por competencias. Final-

mente, el examen se calificó con 281 casos clínicos, debido a los 19 reactivos eliminados.

Los resultados indicaron una consistencia interna alta, un nivel de dificultad moderadamente fácil para su resolución y una discriminación moderada entre puntajes altos y bajos. La competencia 1 presentó mayor dificultad, en tanto que la competencia 6 presentó menor dificultad.

A partir de los resultados psicométricos se integró la información para la toma de decisiones con respecto de la conformación final de los reactivos que fueron contabilizados en la calificación final, y se presentaron los valores que dieron evidencia de parámetros pertinentes para la reutilización o eliminación permanente de los reactivos.

Esta información permitió establecer procedimientos posteriores para la identificación de fortalezas, debilidades y áreas de oportunidad, tanto del examen como del desempeño de los sustentantes.

### 5. Reporte técnico de la prueba

Se generó un reporte técnico que incluyó la estructura del examen por competencias, área clínica, habilidad clínica, área de carrera y nivel taxonómico. También se especificó la cantidad de reactivos nuevos, modificados y reutilizados.

Posterior a la aplicación del examen se generaron informes con los resultados obtenidos por sede (UNAM y escuelas incorporadas).

Los informes institucionales del examen profesional se elaboraron en dos grandes rubros. El primero abordó los antecedentes generales del examen y el segundo integró las características, estructura, resultados y estadísticas de la fase teórica considerando Alpha, medias, desviación estándar y valores mínimo-máximo.

Se compararon los resultados de este EP 2024-1 (**tabla 3**) con otros exámenes como el EP 2023-1 del que se obtuvieron los siguientes resultados psicométricos con 330 reactivos analizados, media de 221.37, desviación estándar de 20.26, puntaje mínimo 122, puntaje máximo 280, dificultad con media de P 0.67, discriminación con Rpbis 0.13, alpha de Cronbach de 0.86 y error estándar de la media 7.52.

También se comparó con el EPT 2021-1, con 325 reactivos analizados, media de 237.02, desviación estándar de 20.21, puntaje mínimo 105, puntaje

máximo 288, dificultad con media de P 0.73, discriminación con Rpbis 0.14, alpha de Cronbach de 0.88 y error estándar de la media 7.07. Observando que al implementar el MIDECOMP se han obtenido mejores resultados psicométricos en los EP-FT.

### DISCUSIÓN

Se ha descrito en la literatura universal la relevancia de la evaluación por competencias en exámenes teóricos tales como *Australian Medical Council* (Australia), *Outcomes for Graduates* (Reino Unido), CanMEDS (Canadá) y el USMLE, que concluyen que para garantizar una evaluación válida y confiable se recomienda<sup>24-27</sup>:

- Realizar casos clínicos cortos con preguntas de opción múltiple de alta complejidad, diseñadas para evaluar no solo el recuerdo de hechos, sino también la aplicación de conceptos basados en competencias.
- Uso de herramientas psicométricas para garantizar que las preguntas sean válidas y confiables.
- Los exámenes deben ser revisados por paneles de expertos para garantizar su relevancia y confiabilidad.

El MIDECOMP ha integrado las recomendaciones previas además como las de la Conferencia de Ottawa de 2010<sup>7</sup>, los 12 pasos de Steve Downing<sup>9</sup>, los estándares de Haladyna y Rodríguez en 2013<sup>13</sup>, las cinco reglas básicas del NBME<sup>14</sup> y el uso de la taxonomía de Castañeda<sup>11</sup>.

Estas características avaladas por sus resultados psicométricos lo hacen un método robusto y adecuado para aplicarlo en diversos contextos de evaluación de competencias médicas, con la respectiva adaptación a las patologías clínicas más comunes de cada lugar y a las competencias que se deseen evaluar en cada institución educativa.

### CONCLUSIONES

La implementación del MIDECOMP permitió diseñar, elaborar, aplicar y analizar el EP-FT del 2024 de manera masiva, eficiente y con estándares de calidad, alineados al PE-2010. Es importante que las evaluaciones de altas consecuencias sigan una metodología que permita medir de manera válida y

confiable a los sustentantes, por lo que recomendamos el uso de esta metodología.

La mejora de los índices psicométricos se ve influenciada por la calidad del proceso de diseño y redacción del instrumento de medición, así como de estructuras que representan la complejidad y contingencia de los aprendizajes evaluados en concordancia con la formación recibida durante su formación profesional, por lo que recomendamos lo siguiente:

En cuanto al diseño de exámenes de altas consecuencias, es recomendable mantener actualizada la tabla de especificaciones en función de los cambios institucionales, tanto en los programas académicos como en las sedes clínicas, para asegurar una evaluación justa y objetiva.

Además, es fundamental que dicha tabla se elabore con el mayor nivel de detalle y precisión, ya que esto garantiza que la evaluación refleje fielmente lo que se enseñó. Esto, a su vez, contribuye a una retroalimentación adecuada del examen.

Respecto a la elaboración de reactivos es recomendable incorporar la inteligencia artificial para generar nuevas viñetas clínicas, siempre revisadas y ajustadas por expertos, ya que podría hacer más eficiente el proceso.

En cuanto a la clasificación de reactivos por enfermedades, con uso del CIE 11 y basadas en el panorama epidemiológico nacional, se recomienda hacer pública dicha información como apoyo a las universidades incorporadas a la UNAM y otras instituciones, para la selección de las patologías más frecuentes del país, así como para que el estudiantado tenga una idea más clara de las patologías con las que se les evaluará.

Para las revisiones de reactivos, se recomienda contar con una buena distribución del tiempo destinado a ello, con la finalidad de disminuir errores por fatiga y cansancio, originados por revisiones constantes de grandes cantidades de reactivos.

Respecto a la logística de aplicación, se ha propuesto una estrategia para reducir el tiempo general de aplicación y aumentar la seguridad del examen: utilizar todos los recintos disponibles de la FM, tanto dentro como fuera del campus, aplicando el examen de forma simultánea.

Esto reduciría un turno y minimizaría la posibilidad de comunicación entre los sustentantes. No obstante, implicaría un aumento en la cantidad de

equipos de cómputo, personal para la supervisión y desplazamientos a los diferentes recintos.

Al adoptar estrategias precisas y objetivas como la metodología descrita, se reflejan las áreas de oportunidad y las competencias mejor asimiladas por los sustentantes, de manera confiable. Esto facilita una retroalimentación efectiva, consolida el aprendizaje y permite modificar el programa académico y los procesos en las prácticas clínicas según sea necesario.

Asimismo, abre oportunidades para investigaciones educativas y de evaluación, proporcionando bases para futuras modificaciones en los planes de estudio y mejoras en los procesos.

### CONTRIBUCIÓN INDIVIDUAL

- ANE: Resumen, objetivos, método, resultados, conclusiones, revisión y correcciones generales al manuscrito, formato para envío, seguimiento.
- CASA: Resumen, traducción, objetivos, método, revisión y correcciones generales al manuscrito, seguimiento.
- AIMP: Resumen, método, revisión y correcciones generales.
- AAH: Resumen, introducción, método, resultados, revisión y correcciones generales.
- RGD: Introducción, método, resultados, revisión y correcciones generales.
- MPD: Método, revisión y correcciones generales al manuscrito.
- AC: Introducción, método, revisión y correcciones generales.
- AOM: Introducción, método, revisión y correcciones generales.

### PRESENTACIONES PREVIAS

Presentación de parte de la metodología en el VIII Congreso Internacional de Educación Médica y VII Congreso Internacional de Simulación en Educación Médica, con reconocimiento del primer lugar en presentación oral.

### FINANCIAMIENTO

Ninguno.

### CONFLICTO DE INTERESES

Ninguno.

**DECLARACIÓN DE IA**Ninguna. **REFERENCIAS**

1. Sánchez Mendiola. La evaluación del aprendizaje de los estudiantes ¿es realmente tan complicada? *Revista Digital Universitaria*. 2018;19,6.
2. Soto-Aguilera Carlos A, Robles-Rivera Karina, Fajardo-Ortiz Guillermo, Ortiz-Montalvo Armando, Hamui-Sutton Alicia. Actividades profesionales confiables (APROC): un enfoque de competencias para el perfil médico. *FEM* (Ed. impresa) [Internet]. 2016 Feb [citado 2024 Dic 10];19(1):55-62. Disponible en: [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S2014-98322016000100010&lng=es](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S2014-98322016000100010&lng=es).
3. Miller G. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63-7. DOI: 10.1097/00001888-199009000-00045.
4. Díaz Barriga Arceo F. Evaluación de competencias en educación superior: experiencias en el contexto mexicano. *Rev Iberoam Eval Educ*. 2019;12(2):49-66. DOI: 10.15366/riee2019.12.2.003.
5. Cortés MT, Petra I, Acosta E, Reynaga J, Fouilloux M, García R, Piedra E. Desarrollo y crecimiento personal: construcción y validación de un instrumento para evaluar esta competencia en alumnos de medicina. *FEM*. 2017;20(2):65-73. DOI: 10.33588/fem.202.881.
6. American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 2014.
7. Boursicot K, Kemp S, Wilkinson T, Findyartini A, Canning C, Cilliers F, Fuller R. Performance assessment: Consensus statement and recommendations from the 2020. Ottawa Conference. *Med Teach*. 2021;(1):58-67. <http://doi.10.1080/0142159X.2020.1830052>
8. Facultad de Medicina, UNAM. Plan de Desarrollo 2008-2016 [Internet]. México: Facultad de Medicina, UNAM; 2024 [citado 2024 ago 28]. Disponible en: [www.facmed.unam.mx/plan\\_2k8\\_2k16/planfm\\_2k8\\_2k16.pdf](http://www.facmed.unam.mx/plan_2k8_2k16/planfm_2k8_2k16.pdf)
9. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Meas Educ*. 2002;15(3):309-34. doi:10.1207/S15324818AME1503\_5
10. Anderson LW, et al. *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives. (First Edition)* Pearson Education Group. Boston: 2001.
11. Castañeda S. *Evaluación del aprendizaje en el nivel universitario: elaboración de exámenes y reactivos objetivos*. Ciudad de México: Facultad de Psicología, UNAM; 2006.
12. Facultad de Medicina UNAM. Plan de Estudios 2010 y Programas Académicos de la Licenciatura de Médico Cirujano. [Internet] Facultad de Medicina UNAM; 2009 [Citado 2023 enero 08]. Disponible en: <http://www.facmed.unam.mx/plan/PEFMUNAM.pdf>
13. Haladyna TM, Rodriguez MC. *Developing and Validating Test Items*. Oxford: Taylor & Francis Group; 2013.
14. National Board of Medical Examiners. *Guía de redacción de preguntas del NBME*. [Internet] Tyson J; 2022 [Citado 2023 noviembre 30]. Disponible en: [https://www.nbme.org/sites/default/files/2022-10/NBME\\_Item-Writing\\_Guide\\_Spanish.pdf](https://www.nbme.org/sites/default/files/2022-10/NBME_Item-Writing_Guide_Spanish.pdf)
15. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-7. doi:10.1046/j.1365-2923.2003.01594.x.
16. Carrillo-Avalos BA, Leenen I, Trejo-Mejía JA, Sánchez-Mendiola M. Bridging Validity Frameworks in Assessment: Beyond Traditional Approaches in Health Professions Education. *Teach Learn Med*. 2023;(18):1-10. doi:10.1080/10401334.2023.2293871.
17. Coughlin PA, Featherstone CR. How to Write a High Quality Multiple Choice Question: A Guide for Clinicians. *Eur J Vasc Endovasc Surg*. 2017;(5):654-658. doi:10.1016/j.ejvs.2017.07.012.
18. Boland RJ, Lester NA, Williams E. Writing multiple-choice questions. *Acad Psychiatry* [Internet]. 2010;34(4):310-6. Disponible en: <http://dx.doi.org/10.1176/appi.ap.34.4.310>
19. Tavakol M, O'Brien DG, Sharpe CC, Stewart C. Twelve tips to aid interpretation of post-assessment psychometric reports. *Med Teach*. 2024;(2):188-195. doi: 10.1080/0142159X.2023.2241624.
20. Abu-Zaid A, Khan TA. Assessing declarative and procedural knowledge using multiple-choice questions. *Med Educ Online*. 2013;(18):21-132. doi: 10.3402/meo.v18i0.21132.
21. Rivera Jiménez J, Flores Hernández F, Alpuche Hernández A, Martínez González A. Evaluación de reactivos de opción múltiple en medicina: evidencia de validez de un instrumento. *Rev Invest Educa Med*. 2017;6(21):8-15. doi:10.1016/j.riem.2016.04.005
22. Alpuche HA. Análisis y utilización de la información obtenida de la evaluación docente: caso de la Facultad de Medicina de la UNAM. En: Hamui A, Motalvo A, Gatica F, editores. *Evaluación de las competencias docentes en las ciencias de la salud*. México: Manual Moderno; 2019:156-177.
23. Muñoz J. *Introducción a la Psicometría: teoría clásica y TRI*. Madrid: Pirámide; 2018.
24. Australian Medical Council. *Standards for Assessment and Accreditation of Primary Medical Programs*. Canberra: AMC; 2012.
25. General Medical Council. *Outcomes for Graduates*. London: GMC; 2018. Disponible en: <https://www.gmc-uk.org/>
26. Frank JR, Snell LS, Sherbino J. *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada; 2015.
27. United States Medical Licensing Examination (USMLE). *USMLE Bulletin of Information*. Philadelphia: Federation of State Medical Boards and National Board of Medical Examiners; 2023.