



Comité Normativo Nacional de Consejos de Especialidades Médicas

Estándares de Calidad en Evaluación

31 de enero de 2025

Comité Normativo Nacional de Consejos de Especialidades Médicas, A.C.

JUNTA DE GOBIERNO DE CONACEM

José Ignacio Santos Preciado

PRESIDENTE

Felipe Cruz Vega

SECRETARIO

Carlos Eduardo Aranda Flores

TESORERO

ACADEMIA NACIONAL DE MEDICINA

Celia Mercedes Alpuche Aranda

Patricia Elena Clark Peralta

Melchor Sánchez Mendiola

ACADEMIA MEXICANA DE CIRUGÍA

Pierre Jean Aurelus

Juan Manuel Guzmán González

Rafael Medrano Guzmán

CONSEJOS DE ESPECIALIDADES

Oscar Díaz Flores

Médicos Anatomopatólogos

Mario César Peláez Luna

Gastroenterología

Álvaro Alejandro Zavala Reina

Neurofisiología Clínica

CONSEJERO JURÍDICO

David Sánchez Mejía

SECRETARIO TÉCNICO

Miguel Ángel Vásquez Luna

GRUPO DE TRABAJO

ESTÁNDARES CONACEM

Dr. José Ignacio Santos Preciado,
Presidente del CONACEM

Dra. Carmen Zavala García,
Presidenta del Consejo Mexicano de
Medicina Interna, A.C.

Dr. Isaías Hernández Torres,
Coordinador Comité Evaluación del
Consejo Mexicano de Certificación
en Medicina Familiar, A.C.

Dr. Juan Domingo Porras Hernández,
Vocal del Comité de Evaluación del
Consejo Mexicano de Cirugía
Pediátrica, A.C.

Dr. Alejandro Rendón Molina,
Director de Vigencia de Certificación
del Consejo Mexicano de Ginecología
y Obstetricia, A.C.

Dr. Pierre Jean Aurelus
Vocal del CONACEM

Dr. Melchor Sánchez Mendiola
Vocal del CONACEM

Lic. Miguel Ángel Vásquez Luna
Secretario Técnico del CONACEM

Estándares de calidad en evaluación del CONACEM

31 de enero de 2025

Comité Normativo Nacional de Consejos de Especialidades Médicas, A.C., todos los derechos reservados.

Introducción

La certificación médica tuvo sus orígenes en Inglaterra en 1858, expandiéndose posteriormente a otros países como Estados Unidos (1916) y Canadá (1929) (Rhodes, 2007). En México, este proceso inició en 1963 con el Consejo de Médicos Anatomopatólogos, marcando el comienzo de una nueva era en la regulación de la práctica médica especializada en el país (Espinosa de los Reyes Sánchez, 2004).

El Comité Normativo Nacional de Consejos de Especialidades Médicas (CONACEM) se constituyó formalmente en 1995, integrado por la Academia Nacional de Medicina de México, la Academia Mexicana de Cirugía y representantes de los consejos de especialidades médicas (Espinosa de los Reyes Sánchez, 2004). En 2006, el Congreso de la Unión fortaleció su papel al modificar la Ley General de Salud, particularmente en su artículo 81, mismo que establece el marco legal para la certificación de especialidades en México (Durán, 2011). Este artículo estipula que CONACEM tiene la responsabilidad de supervisar el entrenamiento, habilidades, destrezas y calificación de la pericia necesaria para la certificación y vigencia de certificación en las diferentes especialidades médicas reconocidas por este Comité y en las instituciones de salud oficialmente reconocidas ante las autoridades correspondientes.

En la actualidad, CONACEM supervisa 47 Consejos de Especialidades Médicas (CEM) que certifican 47 especialidades y 76 subespecialidades. Esta labor se realiza con el apoyo de más de 900 médicos especialistas que participan de manera *pro bono* en los distintos procesos de evaluación (<https://www.conacem.org.mx>).

Para el proceso de certificación de médicos especialistas CONACEM recomienda utilizar estándares de calidad. Los estándares representan niveles previamente establecidos de calidad en la conducción y uso de procedimientos de evaluación, y sirven como referencia para determinar si un sistema, programa o institución cumple con los objetivos y resultados esperados (Harvey, 1993). Un estándar establece los criterios fundamentales para valorar el mérito y la calidad de las prácticas de evaluación, fungiendo como referente objetivo para garantizar la transparencia, validez y mejora continua en los procesos de certificación y evaluación profesional.

Los presentes estándares de calidad en evaluación surgen de la necesidad de homologar los criterios para la evaluación entre pares y reducir la heterogeneidad existente en los procesos de evaluación de los diferentes CEM, dan seguimiento al trabajo realizado en años recientes (Olmedo, 2022; Olivares et al, 2024). La intención de los estándares es promover prácticas sólidas de desarrollo y aplicación de pruebas en la certificación de especialistas médicos y servir como base para evaluar la calidad de esas prácticas.

Este documento establece las especificaciones, requisitos y pautas que deben seguirse para garantizar que los procesos de certificación cumplan con su propósito

fundamental: asegurar que los especialistas certificados posean las competencias necesarias para ejercer su práctica de manera segura y efectiva. Los estándares se agrupan en cinco dominios generales, cada uno representando un aspecto clave de calidad en evaluación (Figura 1).

1. Planificación y diseño del examen: este dominio establece los fundamentos normativos, metodológicos y operativos que garantizan la calidad técnica y psicométrica de los instrumentos de evaluación, promoviendo su validez, confiabilidad y pertinencia.

2. Implementación y aplicación de los instrumentos: aborda los lineamientos logísticos, técnicos y operativos que aseguran la ejecución eficiente, equitativa y segura de las evaluaciones, minimizando riesgos de fraude y errores en todas sus etapas.

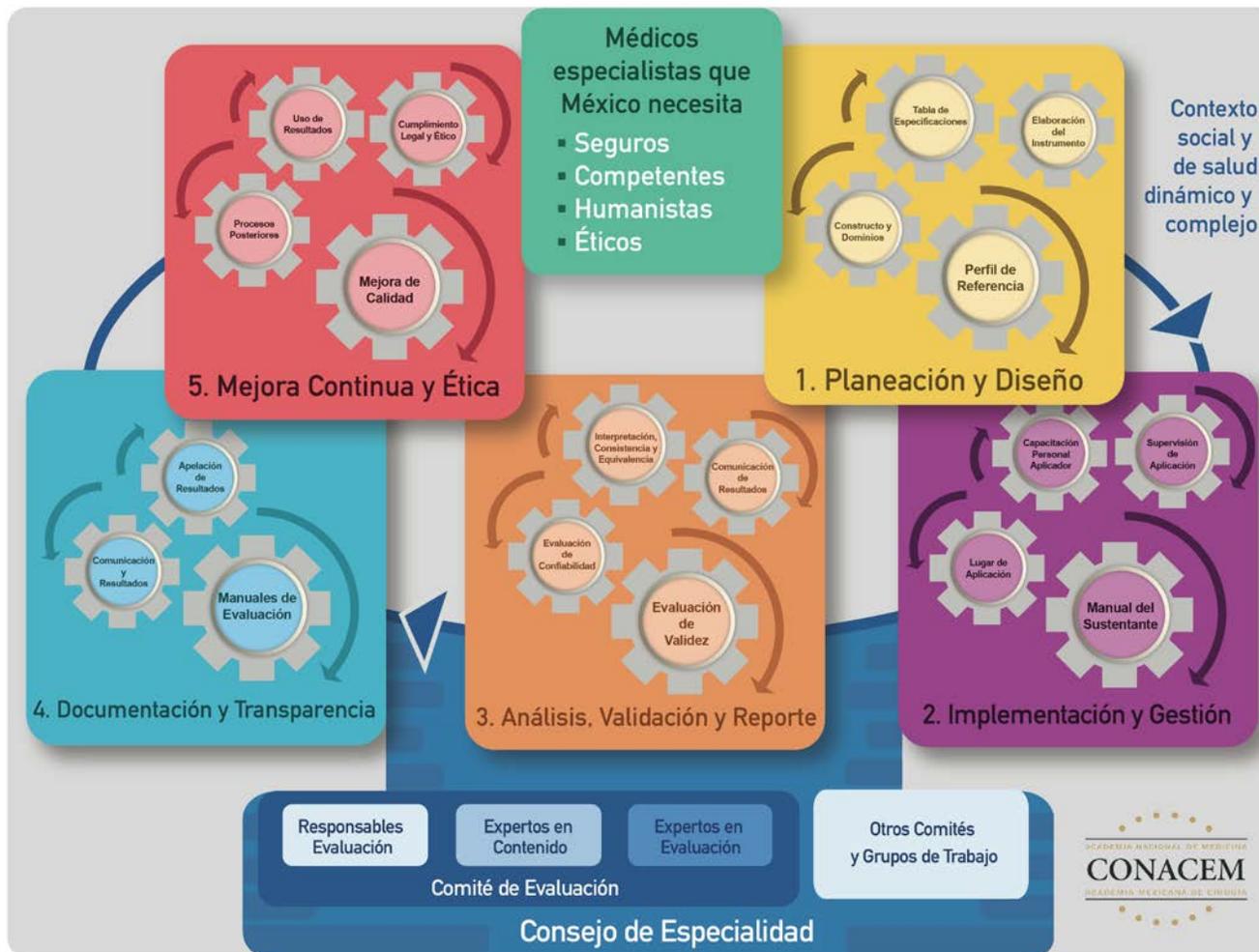
3. Análisis, validación y reporte de resultados: incluye las directrices necesarias para el análisis detallado de los instrumentos, la validación de sus resultados y la comunicación clara y transparente de los mismos, garantizando decisiones fundamentadas y justas.

4. Documentación y transparencia: define los requisitos para la documentación precisa y accesible de los procesos evaluativos, así como para la comunicación segura, equitativa y confidencial de los resultados, asegurando la confianza de todas las partes involucradas.

5. Mejora continua y ética de la evaluación: se enfoca en la revisión sistemática, la innovación y el cumplimiento de principios éticos y legales que respalden la validez, calidad y justicia de las evaluaciones, respetando la dignidad y derechos de los sustentantes.

Estos estándares se alinean con las mejores prácticas internacionales, tomando como referencia a organismos homólogos como el *American Board of Medical Specialties* y el *Royal College of Physicians and Surgeons of Canada*, adaptándose al contexto y necesidades específicas del sistema de salud mexicano.

Figura 1. Esquema de los estándares de evaluación CONACEM



1. Estándares para la Planificación y Diseño de los Instrumentos

OBJETIVO

Diseñar y consolidar un marco normativo, metodológico y operativo que sustente la planificación y diseño de los exámenes de certificación, asegurando la calidad de los instrumentos de evaluación mediante procesos estandarizados, validados y documentados. Este marco debe incluir mecanismos que promuevan la equidad y la inclusión, la identificación y corrección de sesgos, y la implementación de estrategias sistemáticas de mejora continua que garanticen la relevancia, transparencia y confiabilidad de los resultados obtenidos.

I. Organización estructural de Consejos de Especialidades

1.1. Estructura académica: los Consejos de Especialidades deben estar constituidos por profesionales especializados en la disciplina correspondiente. Quienes lo integren deben contar con certificación vigente en la especialidad y tener antecedentes de docencia o demostrar práctica activa en la especialidad que aplique.

1.2. Comité de evaluación (CE): cada Consejo debe tener un Comité de Evaluación (CE) en su organigrama y manual de funciones, que actuará como órgano rector en el diseño, aplicación y análisis de los instrumentos de evaluación. Este comité debe estar compuesto por un grupo interdisciplinario de personas expertas que representen las instancias académicas, técnicas y de evaluación. Los Consejos determinarán, de acuerdo con sus características y posibilidades, el tamaño y la composición de este Comité.

- Responsabilidades: el CE será responsable de todas las etapas del proceso de evaluación, desde la planificación hasta la implementación, revisión y análisis de los exámenes. Este comité debe garantizar que todos los elementos de los instrumentos de evaluación cumplan con estándares de calidad y equidad, y que reflejen las competencias necesarias para la certificación correspondiente.

- Integrantes: el CE debe integrarse por miembros del CEM que estén a cargo de todos los procesos que involucran la elaboración, análisis y reporte de los resultados de los instrumentos de evaluación utilizados. Las funciones de cada integrante deben estar claramente descritas y definidas. El CE debe contar con asesores expertos en evaluación de altas consecuencias, que pueden ser miembros del CE que tengan las credenciales para ello, y/o por personal especialista externo. En el caso de los especialistas externos, su rol se limitará a la asesoría en evaluación, sin voto en las decisiones finales del CE. Asimismo, se recomienda que los integrantes del CE permanezcan en sus funciones al menos durante el periodo correspondiente a dos mesas directivas, en un esquema de rotación secuencial, para que siempre haya personas que tengan

la experiencia vivida en el proceso, con el objetivo de garantizar la continuidad y propiciar mejoras en los procesos.

1.3. Comités y grupos de trabajo: los Consejos podrán designar los Comités y grupos de trabajo que sean necesarios para su adecuado funcionamiento, como pueden ser Comités de Ética, de Vigencia de certificación, de Investigación, entre otros.

II. Reglamentación y Perfil de Puesto de los Integrantes

1.4. Reglamentación: los Consejos de especialidades médicas deben tener una reglamentación interna que defina las atribuciones y responsabilidades de cada elemento de su estructura. Este reglamento debe estar disponible para todos los miembros del consejo, revisarse y actualizarse periódicamente, a más tardar cada cinco años, para reflejar los avances en la práctica de la especialidad y en los métodos de evaluación.

El reglamento debe incluir una descripción de los procedimientos para la creación, revisión, aplicación y mantenimiento de los exámenes de certificación y vigencia de certificación. Además, debe especificar el perfil de los miembros del CE y en su caso de quienes funjan como asesores expertos en evaluación, como prestadores de servicios profesionales.

1.5. Documentación de credenciales y experiencia del CE y asesores externos: los consejos deben contar con un perfil de puesto que acredite las credenciales y experiencia de los miembros del CE y de los asesores expertos en evaluación. Esta documentación puede incluir, a juicio del Consejo y de manera no limitativa:

- Documentos académicos: cédula profesional, diplomas universitarios y certificados vigentes de especialidad médica.
- Nombramientos y constancias laborales: evidencia de nombramientos como profesores en instituciones de educación superior, constancias laborales en instituciones de salud reconocidas por los integrantes del Consejo correspondiente y evidencia de práctica clínica activa.
- Capacitación en evaluación: documentación de capacitación en diseño, elaboración y análisis de instrumentos de evaluación.
- Evidencia de experiencia en el campo: documentación que demuestre la participación activa en actividades, proyectos o comités de evaluación.

1.6. Documentación y conservación de evidencia de las actividades de evaluación: el CE debe mantener un registro de las actividades relacionadas con la planificación, diseño, elaboración, análisis y mantenimiento de los exámenes de certificación. Estos registros deben incluir minutas de reuniones, informes de avances y cualquier otra documentación pertinente. Esta documentación debe conservarse en el archivo del Consejo, por un periodo mínimo de 10 años.

- Minutas de reuniones: cada reunión del CE debe documentarse, especificando objetivo de la reunión, fecha, asistentes, roles dentro del comité y decisiones tomadas con su justificación. Las minutas deben ser firmadas física o digitalmente por los asistentes.

III. Planeación y Diseño de Instrumentos de Certificación y Vigencia de Certificación

1.7. Fundamentación en planes de estudio y características de la especialidad: los instrumentos de certificación deben diseñarse por el CE utilizando los planes de estudio universitarios y la evidencia disponible en torno a la *lex artis* de la especialidad correspondiente, con énfasis en la epidemiología de enfermedades pertinente a la especialidad.

1.8. Perfil de referencia: el CE debe elaborar el perfil de referencia del examen. Este perfil debe incluir una descripción de los conocimientos, destrezas y habilidades, o en su caso las competencias, y características profesionales mínimas que deben poseer los y las profesionales de la especialidad para desempeñar una práctica independiente y segura, en México.

- El perfil de referencia debe considerar la diversidad de planes de estudio presentes en las distintas Instituciones de Educación Superior (IES) del país, así como la evidencia pertinente para la buena práctica, identificando y destacando los elementos comunes esenciales.

1.9. Definición y especificaciones del examen: las definiciones del constructo y los dominios evaluados, así como las especificaciones técnicas del examen, deben ser claras y precisas para que el CE pueda emitir juicios informados sobre la correspondencia de los reactivos/elementos de los instrumentos utilizados, con el dominio que representan.

- La tabla de especificaciones del examen debe ser diseñada de acuerdo con las buenas prácticas de evaluación sumativa de alto impacto, y estar disponible para diseñadores de reactivos y revisores.

IV. Elaboración del instrumento

1.10. Especificaciones para la elaboración de reactivos o retos: los reactivos o retos elaborados para el instrumento de certificación o vigencia de certificación deben ser diseñados de acuerdo con las buenas prácticas de evaluación sumativa de alto impacto referidas en la bibliografía. El proceso de elaboración de reactivos o retos debe describirse con detalle.

- La revisión de los contenidos de los reactivos o retos idealmente debe ser realizada por personal distinto a los escritores de los mismos, para identificar errores y áreas de oportunidad en el instrumento.

1.11. Pilotaje y análisis de reactivos: antes de su inclusión en los exámenes, los reactivos deben ser sometidos a un proceso de pilotaje para documentar sus parámetros psicométricos. Solo los reactivos que cumplan con estándares de calidad predefinidos por el CE serán incorporados en las versiones operativas de los instrumentos.

1.12. Incorporación de reactivos piloteados: los reactivos que se están piloteando pueden ser incluidos en los instrumentos de evaluación, y podrán ser incluidos, a criterio del CE, en la puntuación final siempre y cuando muestren rendimiento adecuado en el análisis psicométrico. Los reactivos con inadecuado desempeño psicométrico deberán ser sometidos a revisión especial por el CE para decidir si pueden ser empleados para la puntuación final.

1.13. Análisis de potenciales sesgos: después del pilotaje y aplicación de la evaluación, los instrumentos deben ser analizados para identificar sesgos potenciales y reales. Se puede utilizar el análisis del Funcionamiento Diferencial del Ítem (DIF), basado en factores de sesgo como género, edad, procedencia de la sede, institución de salud e IES.

- En caso de que se detecten posibles sesgos en los instrumentos, ya sea a través del análisis estadístico o por señalamientos de los usuarios, estos deben ser revisados por el CE, para decidir si los reactivos pueden ser ajustados o si deben eliminarse.

1.14. Bancos de reactivos: el Consejo debe tener un banco de reactivos de los elementos que utilice en sus instrumentos de evaluación, siguiendo las buenas prácticas de evaluación educativa de altas consecuencias. Los elementos de exámenes resguardados en el banco de reactivos deben contar con la información pertinente de su desempeño y fechas de uso.

1.15. Especificaciones para la elaboración de reactivos o retos con inteligencia artificial generativa (IAGen): existe evidencia suficiente para fomentar el desarrollo de reactivos o retos con el apoyo de herramientas de IAGen y modelos de lenguaje de gran tamaño (LLMs), lo anterior siempre bajo la supervisión de expertos en contenido

y evaluación, debiendo validarse la pertinencia, precisión y calidad de los insumos generados para garantizar su alineación con los objetivos del instrumento, así como sus implicaciones legales y de derechos de autor (Mao et al, 2024).

1.16. Análisis de potenciales sesgos: los reactivos generados con el apoyo de IAGen y LLMs deberán someterse a un análisis exhaustivo para identificar sesgos potenciales introducidos por estas herramientas, considerando factores como género, edad, procedencia y contexto institucional. En caso de detectarse sesgos atribuibles al uso de IA, los reactivos deberán ser revisados o eliminados por el comité pertinente.

2. Estándares para la Implementación y Gestión de las Evaluaciones

OBJETIVO

Diseñar e implementar lineamientos operativos y normativos que aseguren la producción, (planificación, logística y ejecución) de los instrumentos de certificación y vigencia, garantizando condiciones de equidad, accesibilidad, seguridad y transparencia, así como la aplicación de las evaluaciones en diferentes modalidades y sitios. Esto incluye minimizar errores, sesgos y riesgos de fraude mediante controles estrictos y mecanismos que promuevan la validez del proceso.

I. Preparación de los escenarios de aplicación físicos o virtuales

2.1. Selección del escenario de aplicación: el CE seleccionará con antelación los lugares de aplicación de la evaluación. Estos lugares serán evaluados previamente por al menos dos consejeros y/o asesores técnicos del Consejo, quienes documentarán su idoneidad mediante listas de verificación que aseguren:

- Accesibilidad.
- Facilidades para aspirantes con necesidades especiales.
- Seguridad tanto para los sustentantes como para el contenido de la evaluación.
- Disponibilidad de equipos adecuados, en caso de usar computadoras, internet o tecnología como simulación.
- Resguardo de pertenencias.

El CE podrá implementar diversas modalidades de evaluación, tales como: exámenes impresos, orales o prácticos de manera presencial; exámenes digitales en dispositivos electrónicos en sitios institucionales con vigilancia presencial; y evaluaciones remotas desde lugares con acceso a Internet, supervisadas en línea. En el caso de las evaluaciones en línea, se garantizará la idoneidad de los sistemas tecnológicos utilizados, verificando que cuenten con medidas de seguridad, herramientas de monitoreo remoto y guías claras para la preparación técnica del sustentante.

2.2. Condiciones físicas y operativas del lugar: se dispondrá de señalización clara, incluyendo adaptaciones para personas con necesidades especiales.

- La distribución de los espacios permitirá un flujo ordenado para las distintas etapas del proceso: llegada, identificación, resguardo de pertenencias, espera, evaluación, recuperación de pertenencias y salida.

- El área de evaluación debe estar organizada para garantizar que cada sustentante realice su examen de forma individual, sin distracciones ni posibilidad de malas prácticas.

2.3. Manual o instructivo para el sustentante: cada Consejo deberá elaborar un manual o instructivo que incluya información relevante sobre:

- Fechas y modo de inscripción, incluyendo costos.
- Modalidad del examen y sedes.
- Requisitos técnicos en caso de modalidad en línea.
- Ubicación, procedimientos y conductas esperadas en caso de modalidad presencial.
- Contenido temático del examen.
- Tipo de examen (opción múltiple, práctico, examen clínico objetivo estructurado -ECO- , oral), con breve descripción de la modalidad.
- Escala o puntaje de calificación.
- Fecha de publicación de resultados.
- Consecuencias ante malas prácticas por parte del sustentante.
- Procedimiento para solicitar revisión del examen.
- Procedimiento de identificación de los sustentantes.

II. Capacitación del personal aplicador

2.4. Competencias del personal aplicador: el CE garantizará que todo el personal asignado para la evaluación cuente con las competencias necesarias para desempeñar sus funciones el día del examen.

2.5. Acompañamiento del sustentante: cada Consejo asignará personal específico para asistir a los sustentantes en el entorno de la evaluación en cada etapa del proceso, desde el inicio hasta el final, en la modalidad presencial, en línea o híbrida.

2.6. Historial de capacitación: cada Consejo documentará el historial de capacitación de los evaluadores, asesores técnicos o de evaluación, incluyendo evidencias de participación en entrenamientos y simulaciones.

III. Supervisión del proceso

2.7. Simulaciones previas: con suficiente antelación, se realizarán simulaciones completas en las sedes de evaluación o en los sistemas tecnológicos utilizados (como simuladores o pacientes simulados), para identificar y corregir errores, sesgos y riesgos de fraude.

2.8. Condiciones previas al examen: la logística, materiales, espacios, señalización e insumos necesarios deberán estar listos al menos 24 horas antes de la evaluación.

2.9. Atención a necesidades especiales: se supervisará que las necesidades especiales manifestadas por los sustentantes durante el proceso de inscripción sean atendidas adecuadamente.

2.10. Supervisión en el día de la evaluación: el personal del Consejo y/o asesores técnicos supervisará todas las etapas del proceso de evaluación utilizando listas de cotejo o los instrumentos que consideren pertinentes para dicho fin, a modo tal que se observe y documente el cumplimiento de los estándares establecidos.

En el caso de evaluaciones en línea, se implementará un protocolo de supervisión remota que contemple la verificación de identidad digital, monitoreo en tiempo real mediante herramientas audiovisuales y sistemas de detección de irregularidades. Además, se proporcionará soporte técnico en tiempo real para resolver problemas operativos que puedan surgir durante la evaluación.

2.11. Encuesta de satisfacción: los sustentantes deberán evaluar su experiencia mediante una encuesta de satisfacción al finalizar la evaluación.

2.12. Revisión posterior al examen: el CE revisará la documentación recolectada al finalizar la jornada, identificando aciertos, errores, aprendizajes y oportunidades de mejora. Estos aspectos deberán integrarse al documento de mantenimiento del examen para futuras ediciones.

3. Estándares para el Análisis, Validación y Reporte de Resultados

OBJETIVO

Establecer procesos estandarizados y basados en buenas prácticas para el análisis, validación y comunicación de los resultados de las evaluaciones, garantizando que las puntuaciones obtenidas reflejen de manera justa y precisa los conocimientos, habilidades, destrezas, actitudes y/o valores evaluados. Asimismo, asegurar la transparencia y la equidad mediante el uso de métodos confiables que permitan identificar errores de medición, promover la mejora continua y comunicar resultados de forma oportuna y comprensible.

I. Estándares para Evaluar la Validez

3.1 Revisión y alineación del contenido: los exámenes deben reflejar fielmente la *lex artis* médica y las competencias centrales definidas para la especialidad médica que corresponda.

3.2 Fuentes de evidencia de validez: los exámenes deben basarse en las cinco fuentes de evidencia de validez propuestas por Messick y Downing (AERA, 2014; Downing, 2003):

- **Contenido:** reflejar fielmente la *lex artis* médica y las competencias centrales de la especialidad médica correspondiente.
- **Proceso de respuesta:** analizar cómo los sustentantes interpretan y responden los reactivos o retos, garantizando que las respuestas sean congruentes con las competencias evaluadas.
- **Estructura interna:** realizar análisis psicométricos de los instrumentos (análisis de reactivos, medición de confiabilidad, dimensionalidad, entre otros), para verificar que la prueba refleje adecuadamente los constructos teóricos.
- **Relaciones con otras variables:** evidenciar que los resultados del examen se correlacionen de manera significativa con otras mediciones o desempeños relevantes.
- **Consecuencias:** evaluar el impacto del examen en la formación profesional, asegurando que promueva prácticas competentes, seguras y éticas.

3.3 Modelo de argumentación de validez: el diseño y análisis de los exámenes deberán fundamentarse en las inferencias de validez propuestas por Kane y descritas por Cook (2015), asegurando un marco robusto y sistemático para la toma de decisiones durante todo el proceso evaluativo:

- **Inferencia de calificación:** asegurar que las observaciones (respuestas a reactivos o retos) se traduzcan en juicios válidos sobre la ejecución de la tarea evaluada.
- **Inferencia de generalización:** garantizar que los juicios sobre la ejecución se generalicen de manera confiable al conjunto más amplio de capacidades o competencias del sustentante.
- **Inferencia de extrapolación:** demostrar que los juicios sobre la competencia reflejan la preparación del sustentante para el ejercicio autónomo y seguro de la práctica profesional.
- **Inferencia de implicaciones:** validar que los juicios sobre la autonomía del sustentante respalden decisiones sumativas relacionadas con la certificación, considerando la seguridad, competencia, ética y humanismo del profesional evaluado.

3.4 Aplicación práctica: se fomentará que los CEM integren evidencias basadas en estos modelos de validez, para fortalecer la calidad de la evaluación y su impacto en la formación y desarrollo profesional de los especialistas.

II. Estándares para Evaluar Error de Medición, Confiabilidad, Longitud del Examen y sus componentes

3.5. Evaluación del error de medición: se recomienda que el error de medición sea evaluado de manera continua en cada ciclo de aplicación de la prueba, utilizando el Error Estándar de Medición (EEM) como indicador. El valor del EEM debe mantenerse dentro de límites aceptables, de acuerdo con los estándares internacionales de evaluación psicométrica, garantizando que el grado de incertidumbre sea lo suficientemente bajo como para no comprometer la validez de las decisiones basadas en los resultados del examen (Lane et al, 2016).

3.6. Confiabilidad de las evaluaciones: se recomienda que cada evaluación incluya un reporte detallado de los métodos de estimación de confiabilidad utilizados, tales como coeficiente alfa de Cronbach, KR-20, entre otros. Esto proporciona evidencia de la reproducibilidad de los resultados obtenidos (Reidl-Martínez, 2013).

3.7. Análisis psicométrico de los ítems: se recomienda realizar un análisis de los ítems basado en alguno de los modelos teóricos disponibles (Teoría Clásica del Test, Teoría de Respuesta al Ítem, etc.) Se recomienda que los análisis incluyan el índice de dificultad, el índice de discriminación y la correlación punto biserial en el caso de la TCT, así como los parámetros de dificultad, discriminación y pseudoazar para la TRI (Haladyna & Rodríguez, 2013; Lane et al, 2016).

3.8. Concordancia interevaluador en evaluaciones clínicas: para evaluaciones clínicas que involucren múltiples evaluadores, se recomienda asegurar un nivel de concordancia interevaluador superior a 0.8. La confiabilidad entre evaluadores debe evaluarse mediante el coeficiente de correlación intraclase o el coeficiente kappa de Cohen, garantizando así la coherencia y objetividad en la calificación del desempeño de los examinados.

3.9. Longitud de las pruebas: se recomienda que la longitud de cada prueba se base en estudios psicométricos previos que aseguren un equilibrio adecuado entre confiabilidad y practicidad. Debe garantizarse que el número de ítems sea suficiente para obtener una alta confiabilidad (valores superiores a 0.9 en certificaciones) sin comprometer la experiencia del examinado debido a la fatiga.

3.10. Tiempo de respuesta: Se recomienda estimar el tiempo necesario para responder los exámenes basándose en análisis preliminares, como el juicio de expertos en evaluación y la complejidad de los reactivos, complementado con datos históricos u obtenidos de estudios piloto. El tiempo total asignado al examen deberá permitir que los sustentantes respondan de manera reflexiva, evitando presión indebida y garantizando la validez de las decisiones basadas en los resultados (Margolis & Feinberg, 2020).

III. Estándares para la Interpretación, Equivalencia y Consistencia de Versiones

3.11. Uso de ítems ancla para garantizar la equivalencia en exámenes con múltiples versiones: en los exámenes que presentan múltiples versiones, se recomienda el uso de ítems ancla como estrategia para asegurar la equivalencia entre las diferentes formas del examen. Estos ítems permiten minimizar las diferencias en la dificultad percibida, facilitar la equiparación de las puntuaciones mediante modelos de Teoría de Respuesta al Ítem (TRI) y garantizar resultados justos y comparables entre los candidatos.

3.12. Integración y actualización del banco de reactivos: se recomienda mantener un banco de reactivos actualizado y calibrado psicométricamente, asegurando que cuente con suficientes ítems de calidad para permitir la creación de múltiples versiones del examen. Los ítems deben ser evaluados continuamente para asegurar su relevancia y adecuación al contexto actual de la práctica médica.

3.13. Establecimiento de puntos de corte: el CE debe documentar sistemáticamente el método utilizado para establecer los puntos de corte de los instrumentos. Se debe utilizar un método formal de establecimiento de punto de corte, como el método de

Angoff modificado, Ebel, Hofstee, Beuk, Bookmark, entre otros, usando el juicio de expertos. Este proceso debe ser documentado de manera transparente, y es fundamental asegurarse de que los puntos de corte sean consistentes con el nivel de competencia requerido para la práctica profesional.

IV. Estándares para la Interpretación y Comunicación de Resultados

3.14. Transparencia y accesibilidad: los CEM deben asegurar que los procedimientos para la interpretación de resultados y las escalas utilizadas sean auditables y comprensibles. Los resultados deben ser revisados periódicamente para garantizar que se alineen con los estándares internacionales, asegurando la transparencia en la evaluación y comunicación de los resultados.

3.15. Retroalimentación personalizada: los resultados deben ser comunicados de manera que promuevan la mejora continua. Para ello, se recomienda proporcionar un informe suficiente a cada sustentante, donde se especifiquen sus fortalezas y áreas de mejora. Este enfoque contribuye a un proceso de evaluación más formativo y centrado en el desarrollo de competencias.

4. Estándares para la Documentación y Transparencia

OBJETIVO

Establecer un marco normativo y metodológico para la elaboración, actualización y consulta de los manuales de evaluación, así como para la comunicación efectiva, segura y equitativa de los resultados. Este marco debe garantizar que cada etapa del proceso evaluativo se documente y comunique con altos estándares de calidad, transparencia, confiabilidad y confidencialidad, fomentando la mejora continua y la confianza de los sustentantes y partes interesadas.

I. Estándares para los Manuales de Evaluación

4.1. Manuales de procedimientos generales: los Comités de Evaluación deben contar con manuales de procedimientos generales que describan los procesos administrativos, operativos y logísticos relacionados con las actividades del Consejo. Estos manuales deben incluir:

- Introducción.
- Objetivo general.
- Identificación e interacción de procesos administrativos y operativos.
- Descripción de procedimientos generales aplicables al Consejo (como reuniones, formación de comités, asignación de roles).
- Registro de ediciones y actualizaciones.
- Distribución interna y validación de los manuales.

4.2. Manuales específicos de evaluación: cada Comité de Evaluación debe contar con manuales detallados que documenten las etapas clave del proceso evaluativo y aseguren la calidad técnica y psicométrica de los instrumentos de evaluación. Estos manuales deben incluir:

- Fundamentación teórica del examen: explicación de las bases teóricas y psicométricas que sustentan el diseño y desarrollo de los instrumentos de evaluación.
- Modelos psicométricos utilizados: descripción de las metodologías empleadas para garantizar la validez y confiabilidad del instrumento.
- Resultados de análisis de confiabilidad y validez: evidencia cuantitativa que respalde la calidad del examen.

Asimismo, deben incluir procedimientos específicos para:

- Planeación, diseño, elaboración, selección y piloteo de reactivos.
- Aplicación del examen y manejo de protocolos de seguridad.
- Análisis de resultados y retroalimentación.
- Establecimiento de puntos de corte.
- Comunicación de resultados y entrega de certificados.

4.3. Disponibilidad y accesibilidad: El Manual del sustentante que contemple los diferentes elementos del proceso de certificación, desde la emisión de la convocatoria hasta el final, debe estar disponible para consulta pública. Esto garantiza transparencia y permite a los interesados comprender los fundamentos y procedimientos que sustentan la evaluación.

4.4. Actualización periódica: los manuales deben revisarse y actualizarse periódicamente para garantizar su vigencia y congruencia con las mejores prácticas internacionales y los avances en la evaluación educativa y psicométrica.

II. Estándares para la Comunicación de Resultados

4.6. Transparencia y plazos definidos: los resultados de los exámenes deben comunicarse dentro de un plazo previamente establecido y razonable, asegurando que todos los sustentantes reciban información precisa, clara y oportuna. La comunicación de los plazos debe ser parte integral de la convocatoria inicial.

4.7. Retroalimentación personalizada: se debe proporcionar un informe suficiente a cada sustentante que incluya su puntaje global, el punto de corte, áreas de fortaleza y oportunidades de mejora. Esta retroalimentación fomenta la reflexión y el desarrollo profesional continuo, además de fortalecer la percepción de equidad del proceso evaluativo.

4.8. Reportes poblacionales y contextuales: se recomienda reportar análisis grupales o poblacionales por región, institución u otras categorías, en apego a las necesidades específicas de cada especialidad, en atención a los centros formadores y, cuando corresponda, a los colegios profesionales pertinentes.

4.9. Confidencialidad y seguridad de la información: la comunicación de los resultados debe cumplir con los estándares más altos de confidencialidad y seguridad. Solo los candidatos y las partes autorizadas deben tener acceso a los resultados, los cuales deberán transmitirse a través de plataformas seguras que garanticen la protección de los datos personales.

4.10. Proceso de apelación: debe estar establecido un mecanismo claro y accesible para que los sustentantes que no estén conformes con sus resultados puedan apelar. Este proceso debe incluir guías detalladas, plazos razonables y una revisión técnica transparente que garantice la equidad y la objetividad en la resolución de las apelaciones.

5. Estándares para la Mejora Continua y Ética de la Evaluación

OBJETIVO

Asegurar que las evaluaciones, en cada una de sus ediciones, sean objeto de mantenimiento, análisis continuo e innovación, promoviendo su validez, confiabilidad y calidad. Asimismo, garantizar que los procesos evaluativos cumplan con los principios de transparencia, equidad, confidencialidad y justicia, respetando la dignidad de los sustentantes y fomentando prácticas éticas, legales y responsables en el manejo de información y en la toma de decisiones.

I. Procesos Posteriores a la Aplicación y Calificación

5.1. Protocolos de seguridad en la gestión de materiales: se implementarán protocolos rigurosos para la recolección, traslado y disposición de los materiales de evaluación, garantizando medidas de seguridad similares a las utilizadas para su recepción y almacenamiento. El acceso a los materiales estará limitado al menor número posible de personas, quienes deberán firmar una carta de confidencialidad.

5.2. Monitoreo y resguardo de materiales: los materiales serán monitoreados y resguardados según los mecanismos definidos por cada consejo, y el Comité de Evaluación (CE) verificará el cumplimiento de los protocolos mediante listas de cotejo.

5.3. Organización y registro de resultados: las inscripciones y materiales de respuesta (hojas, bases de datos, rúbricas, etc.) se revisarán y asociarán cuidadosamente con cada sustentante, asegurando la precisión de los resultados y su correcta atribución.

5.4. Uso de resultados para investigación: cuando los resultados se utilicen con fines de investigación, se deberá:

- Informar a los sustentantes, por ejemplo, mediante un descargo de responsabilidad (“disclaimer”) al inicio de la evaluación.
- Preservar la confidencialidad mediante el registro de datos de manera anónima.

5.5. Mantenimiento y calidad del banco de reactivos: el mantenimiento del banco de reactivos, casos clínicos estandarizados, escenarios simulados, listas de cotejo, rúbricas y demás instrumentos incluirá el historial psicométrico de cada elemento para asegurar la calidad en cada calificación o puntaje.

5.6. Capacitación de evaluadores y supervisores: se debe documentar la capacitación continua de evaluadores y supervisores, así como la evaluación de proveedores o asesores externos mediante cédulas específicas.

5.7. Procedimientos de revisión de resultados: el sustentante podrá solicitar la revisión de sus resultados; los procedimientos, alcance e implicaciones de la revisión estarán claramente descritos en el manual del sustentante.

5.8. Manejo de irregularidades: los criterios y procedimientos para invalidar o anular resultados por irregularidades estarán documentados, y el sustentante afectado será informado del proceso con pleno respeto a su confidencialidad y derecho a réplica.

5.9. Protección de información del sustentante: se documentarán procedimientos para resguardar la información relacionada con la evaluación, especificando tiempos de conservación y métodos de destrucción segura, de acuerdo con la legislación mexicana aplicable.

II. Uso de los Resultados

5.10. Certificación profesional: los resultados serán utilizados para certificar que un profesional cumple con las competencias mínimas para ejercer su especialidad en México de manera segura e independiente.

5.11. Uso de resultados para investigación, control y mejora de calidad: los resultados de las evaluaciones podrán ser utilizados para acciones de mejora y control de la calidad, así como para investigación educativa, garantizando siempre el anonimato y la confidencialidad de los datos de los sustentantes. El Consejo deberá notificar a los sustentantes que, al participar en el examen, aceptan de manera general que sus resultados pueden ser empleados con estos fines, en apego a principios éticos y normativos vigentes, sin necesidad de detallar proyectos específicos.

5.12. Compartición de datos agregados: los resultados agregados y anónimos podrán compartirse con el público o terceros claramente definidos en un aviso de privacidad, siempre asegurando la protección de los datos personales.

III. Mejora Continua de la Evaluación

5.13. Revisión continua del desempeño de los procesos evaluativos: cada consejo deberá revisar y actualizar sistemáticamente sus procedimientos y resultados después de cada edición, documentando las mejoras a implementar.

5.14. Innovación y mejores prácticas: se promoverá la adopción de mejores prácticas y tecnologías en la evaluación, incrementando su validez, confiabilidad y eficiencia.

5.15. Participación de sustentantes y personal: sustentantes y personal participante podrán expresar sus opiniones sobre el proceso evaluativo, fomentando una retroalimentación continua.

5.16. Actualización del personal: el personal involucrado en las evaluaciones deberá mantenerse actualizado sobre cambios en procedimientos, tecnologías y buenas prácticas.

IV. Cumplimiento Legal y Ético

5.17. Respeto por los derechos humanos: cada consejo garantizará el respeto por la dignidad y los derechos de las personas en sus procesos de evaluación, adhiriéndose a los estándares éticos más altos.

5.18. Protección de datos personales: el manejo de la información y los datos personales cumplirá con la normativa aplicable en México, garantizando la privacidad y confidencialidad en todas las etapas.

5.19. Cultura de integridad: se fomentará una cultura de profesionalismo, transparencia, responsabilidad e integridad en todas las actividades relacionadas con la evaluación.

5.20. Auditorías periódicas: los consejos estarán dispuestos a someterse a auditorías periódicas que permitan evaluar el cumplimiento normativo y promover mejoras en sus prácticas.

Referencias

- AEA. (2012). *European Framework of Standards for Educational Assessment 1.0*. The Association of Educational Assessment – Europe. Edizioni Nuova Cultura. Recuperado de <https://aea-europe.net/standards-for-assessment/>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Estándares para pruebas educativas y psicológicas* (M. Lieve, Trad.). Washington: American Educational Research Association. Recuperado de <https://www.testingstandards.net/open-access-files.html>
- Brennan, R. L., National Council on Measurement in Education, & American Council on Education. (2006). *Educational Measurement* (4ª ed.). Praeger Publishers.
- CONACEM. (2015). *Manual de procedimientos 2015*. Recuperado el 14 de enero de 2022 de <https://www.conacem.org.mx/>
- Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*, 49(6), 560–575. <https://doi.org/10.1111/medu.12678>
- Davis-Becker, S., & Buckendahl, C. W. (Eds.). (2017). *Testing in the Professions: Credentialing Policies and Practice* (1ª ed.). Routledge. <https://doi.org/10.4324/9781315751672>
- Downing S. M. (2003). Validity: on meaningful interpretation of assessment data. *Medical Education*, 37(9), 830–837. <https://doi.org/10.1046/j.1365-2923.2003.01594.x>
- Drasgow, F. (Ed.). (2015). *Technology and Testing: Improving Educational and Psychological Measurement* (1ª ed.) Routledge. <https://doi.org/10.4324/9781315871493>
- Durán, A. A. (2011). La certificación del médico especialista. Consejos de Especialidad Médica o Colegios Médicos. *Ginecología y Obstetricia de México*, 79(10), 593-600. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=31138>
- Ercikan, K., & Pellegrino, J. W. (Eds.). (2017). *Validation of Score Meaning for the Next Generation of Assessments: The Use of Response Processes* (1ª ed.). Routledge. <https://doi.org/10.4324/9781315708591>
- Espinosa de los Reyes Sánchez, V.M. (2004). *La Academia Nacional de Medicina de México y la certificación de los especialistas por los Consejos de Especialidades Médicas*. Coedición: Facultad de Medicina, UNAM y Academia Nacional de Medicina.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and Validating Test Items* (1ª ed.). Routledge. <https://doi.org/10.4324/9780203850381>
- Hao, J., et al. (2024), Transforming Assessment: The Impacts and Implications of Large Language Models and Generative AI. *Educational Measurement: Issues and Practice*, 43: 16-29. <https://doi.org/10.1111/emip.12602>

- Harvey, L., & Green, D. (1993). Defining Quality. *Assessment & Evaluation in Higher Education*, 18(1), 9–34. <https://doi.org/10.1080/0260293930180102>
- Hassan, S. (2024). Standard Setting in Medical Education: Standards, Methods, and Psychometrics. In: Zaidi, S.H., Hassan, S., Bigdeli, S., Zehra, T. (eds) *Global Medical Education in Normal and Challenging Times. Advances in Science, Technology & Innovation*. Springer, Cham. https://doi.org/10.1007/978-3-031-51244-5_16
- Instituto Nacional para la Evaluación de la Educación (INEE). (2017). *Criterios técnicos para el desarrollo y uso de instrumentos de evaluación educativa, 2014-2015*. Recuperado de https://www.inee.edu.mx/wp-content/uploads/2019/02/CRITERIOS_TECNICOS_PARA_EL_DESARROLLO_Y_USO_DE_INSTRUMENTOS_10_ABRIL_2014.pdf
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2016). *Handbook of Test Development* (2ª ed.). Taylor & Francis.
- Mao, J., Chen, B. & Liu, J.C. Generative Artificial Intelligence in Education and Its Implications for Assessment. *TechTrends* **68**, 58–66 (2024). <https://doi.org/10.1007/s11528-023-00911-4>
- Margolis, M.J., & Feinberg, R.A. (Eds.). (2020). *Integrating Timing Considerations to Improve Testing Practices* (1st ed.). Routledge. <https://doi.org/10.4324/9781351064781>
- Olmedo Canchola, V. H. (2022). Los nuevos indicadores de calidad del CONACEM. *Certeza CONACEM*, (1), 4-6. <https://certeza.conacem.org.mx/numero-1/el-comite>
- Olivares Olivares, S. L., Navarro Olivos, E., Miranda Fernández, K. A., & Rosas Jurado, M. G. (2024). 4. Desarrollo de una guía de procesos para evaluación de estudiantes y egresados de especialidades médicas. En: E. Sánchez Flores, & A. M. Torres Hernández, *La retribución social de posgrados en México: Un diálogo nacional* (págs. 81-112). Guadalajara, México: Universidad de Guadalajara.
- Reidl-Martínez, L. (2013). Confiabilidad en la medición. *Investigación en Educación Médica*, 2(6), 107-111. [https://doi.org/10.1016/S2007-5057\(13\)72695-4](https://doi.org/10.1016/S2007-5057(13)72695-4)
- Rhodes, R. S., & Biester, T. W. (2007). Certification and maintenance of certification in surgery. *Surgical Clinics of North America*, 87(4), 825-836. <https://doi.org/10.1016/j.suc.2007.06.004>
- Sánchez-Mendiola, M., & Delgado-Maldonado, L. (2017). Exámenes de alto impacto: implicaciones educativas. *Investigación en Educación Médica*, 6(21), 52-62. <https://doi.org/10.1016/j.riem.2016.12.001>
- Sánchez Mendiola, M., Martínez González, A. Eds. (2022). *Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos*. 1ª ed. Ciudad de México, UNAM. 774 pág. ISBN versión impresa 978-607-30-6076-9. ISBN versión digital: 978-607-30-6071-4. <https://www.ceide.unam.mx/index.php/evaluacion-y-aprendizaje-en-educacion-universitaria/>
- Zapata-Rivera, D. (Ed.). (2018). *Score Reporting Research and Applications* (1ª ed.). Routledge. <https://doi.org/10.4324/9781351136501>