



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MÉDICAS,
ODONTOLÓGICAS Y DE LA SALUD
CAMPO DE CONOCIMIENTO: EDUCACION MÉDICA**

**EVALUACIÓN DE LA COMPETENCIA CLÍNICA CON EL EXAMEN CLÍNICO
OBJETIVO ESTRUCTURADO (ECOE) EN EL INTERNADO MÉDICO DE
PREGRADO**

**TESIS QUE PARA OPTAR POR EL GRADO DE:
DOCTORADO EN CIENCIAS DE LA SALUD**

**PRESENTA
JUAN ANDRÉS TREJO MEJÍA**

**TUTOR PRINCIPAL: Dr. Adrián Martínez González
Facultad de Medicina. UNAM**

COTUTORES

**Dr. Ignacio Méndez Ramírez
Instituto de Matemáticas Aplicadas y Sistemas. UNAM
Dr. Melchor Sánchez Mendiola
Facultad de Medicina. UNAM**

México D.F. Octubre de 2014



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

I.	Resumen	4
II.	Introducción	6
	Evaluación	6
	Concepto de competencias	7
	Evaluación de la competencia clínica con el ECOE	11
	El ECOE a nivel nacional	14
	Validez	15
	Confiabilidad	16
	Teoría de la generalizabilidad	19
	Internado médico de pregrado	27
III	Justificación	31
IV	Planteamiento del problema	32
V	Objetivos generales	33
	Objetivos específicos	33
	Hipótesis	34
VI	Metodología	35
	Tipo de estudio	35
	Población	35
	Criterios de inclusión y exclusión	35
	Cálculo de la muestra	35
	Variables. Clasificación metodológica	38
	Operacionalización de variables	38
	Instrumentos de recolección de información	40
	Etapas en la conformación de los instrumentos	41

	Fuentes de recolección de información	45
	Plan de análisis estadístico	46
VII	Organización general del estudio	48
VIII	Aspectos éticos	49
IX	Recursos	50
	Recursos Humanos	50
	Recursos Materiales	50
	Financieros	48
X	Resultados	52
	Resultados de las correlaciones	56
	Resultados confiabilidad	59
XI	Discusión	64
	Limitaciones	75
XII	Conclusiones	78
XIII	Referencias bibliográficas	81
	Anexo 1 Material del alumno (Instrucciones)	90
	Anexo 2 Material del evaluador (Lista de cotejo)	91
	Anexo 3 Material del paciente estandarizado (Libreto)	92
	Anexo 4 Tabla de especificaciones	93
	Anexo 5 Matriz de un examen ECOE	94
	Anexo 6 Hoja de lector óptico	95
	Anexo 7 Presupuesto por examen	96

I. Resumen

Introducción: El Examen Clínico Objetivo Estructurado (ECO) es un método ampliamente utilizado para evaluar la competencia clínica en ciencias de la salud. Existe poca evidencia publicada de su uso en escuelas de medicina en México. En América Latina no hay referencias sobre la confiabilidad del ECO a través de la teoría de la generalizabilidad.

Objetivos: Evaluar la competencia clínica de estudiantes de medicina, antes y después del Internado Médico mediante el ECO.

Valorar la confiabilidad del ECO con el alfa de Cronbach y con la Teoría de la Generalizabilidad.

Métodos: Estudio de cohorte prospectivo, pre-pos-test de un solo grupo. La población evaluada fueron estudiantes de la Facultad de Medicina de la UNAM en México, que cursaban el año de Internado Médico. El instrumento fue un ECO de 18 estaciones, tres de cada área disciplinal.

El pre-test fue el mismo para todos y el pos-test tuvo 4 versiones y se aplicó en 6 sedes clínicas simultáneamente en dos días.

Se empleó un diseño cruzado de efectos aleatorios para identificar las principales fuentes de variación. Las cuatro facetas fueron los estudiantes, las estaciones, las sedes y las versiones. Los estudiantes estaban anidados en versión y sede. Los evaluadores, los pacientes estandarizados y las estaciones fueron considerados como una faceta única de análisis.

Resultados: Se realizó el ECO pre-test para evaluar la competencia clínica de 278 estudiantes al ingresar al Internado Médico y un ECO pos-test 10 meses después. La muestra constituyó el 30.4% del total de los Internos de pregrado. La media global de la puntuación del ECO pre-test fue 55.6 ± 6.6 y la media del pos-test 63.2 ± 5.7 ($p < 0.001$), con una d de Cohen de 1.2.

De los componentes de la competencia clínica, la interpretación radiológica obtuvo las puntuaciones más bajas en relación con los otros componentes.

Hubo diferencia significativa entre el pre y el pos-test por institución de procedencia, la mayor la diferencia fue en los estudiantes del IMSS.

La confiabilidad del estudio con el Alfa de Cronbach fue de 0.62 en el pre-test y de 0.64 en el pos-test. El coeficiente de generalizabilidad fue de 0.87 en el pre-test y 0.93 en el pos-test.

Los mayores componentes de varianza fueron el error residual, las estaciones y los estudiantes. Las sedes y las versiones tuvieron una varianza mínima.

Discusión y Conclusiones: La competencia clínica de los estudiantes de medicina, medida con el ECOE, fue mayor al final del Internado médico de pregrado. Esta diferencia sugiere que el Internado puede influir en el desarrollo de la competencia clínica, en estudiantes de medicina. No hay una variable que por sí sola tuviera un valor predictivo alto para el resultado del pos-test.

Los ECOEs formativo y sumativo de altas consecuencias muestran evidencia de confiabilidad de acuerdo a los dos análisis complementarios realizados, el Alfa de Cronbach y la Teoría G. El análisis de las fuentes de error permite una mejora continua del ECOE al derivar inferencias confiables para la toma de decisiones en su planeación.

Palabras clave: educación médica de pregrado; evaluación del aprendizaje; competencia clínica; internado médico; ECOE.

II. Introducción

El avance del conocimiento y el acelerado desarrollo tecnológico han generado modificaciones importantes y diversificación en el ejercicio de la medicina para dar respuesta a las necesidades, expectativas y exigencias de la sociedad actual, por lo que se requiere de un médico que haya alcanzado un nivel de competencia clínica, que sea capaz de proporcionar una atención integral y al mismo tiempo mantener el aspecto humanístico a través de su relación con el paciente.

El proceso de formación de los médicos es largo y complejo. En su parte medular, implica la asimilación por parte del estudiante de conocimientos científicos, habilidades, actitudes y valores. Además del adiestramiento multidisciplinario requerido y de las competencias clínicas, la formación debe contemplar ciertas estrategias bajo la guía y supervisión de un conjunto de profesores experimentados para que el estudiante adquiera de manera progresiva, la capacidad de solucionar problemas de forma independiente de acuerdo con el perfil del egresado, para un desempeño exitoso en la sociedad contemporánea¹.

El Plan de la Facultad de Medicina plantea un perfil profesional del egresado de la carrera de Médico Cirujano que tiene como propósito formar un médico capaz de ejercer la práctica de la medicina general con los conocimientos, la calidad técnica, la ética profesional y el humanismo que requieren el cuidado, la promoción y la recuperación de la salud.

Para lograr el propósito anterior se requiere la participación de profesores del área biomédica, sociomédica y humanística así como del área clínica que en su justo balance formen un médico competente para ejercer como profesional de la medicina general, por ello las funciones y actividades que realiza el docente y el sistema de evaluación del aprendizaje para valorar el perfil del egresado juegan un papel esencial.

Evaluación

La evaluación es el enjuiciamiento sistemático de la valía o el mérito de un objeto². Implica un proceso sistemático de construcción de conocimiento a partir de la realidad, que orienta la toma de decisiones con el objetivo de provocar cambios positivos en ella. Dicho proceso parte de la recopilación y el análisis de

información y, en el contexto educativo, se orienta a la emisión de juicios de valor sobre algún sujeto, objeto o intervención educativos. Desempeña un papel central para determinar si los estudiantes han alcanzado las competencias que han sido especificadas y si el programa educacional ha sido eficaz.

Es una parte esencial e integral de la educación médica y entre sus propósitos según su tipo destacan: La evaluación diagnóstica identifica si el estudiante cuenta con los conocimientos, habilidades, actitudes, valores y aptitudes necesarias para desarrollar de manera satisfactoria una actividad. La evaluación formativa otorga a los estudiantes realimentación con respecto a su progreso e identifica las áreas fuertes y aquellas que requieren de mejora. La evaluación sumativa o acreditadora determina si los estudiantes alcanzan los estándares aceptables y si pueden continuar o no la siguiente etapa del proceso enseñanza-aprendizaje³.

En suma, la evaluación diagnóstica, formativa y sumativa son complementarias y a través de evaluar el desarrollo académico del estudiante y los resultados del proceso educativo, como las competencias, aportan datos útiles para tomar decisiones encaminadas a mejorar el aprendizaje, así como la enseñanza y el plan de estudios².

Competencias

En un análisis del concepto de competencias de varios organismos (Libro Blanco-España, AMFEM-México, ACGME-ABMS-EUA, Escuela de Medicina de Indiana, CanMeds-Canada, IOM-EUA)^{4,5,6,7,8,9} hay elementos comunes como son: el dominio de las bases científicas de la medicina, las habilidades clínicas para el diagnóstico, manejo y actividades preventivas, el profesionalismo y el comportamiento ético. Una de ellas conceptualiza las competencias como los roles del médico y otra describe los elementos de una buena práctica¹⁰.

Las competencias se definen como el conjunto de conocimientos, habilidades, actitudes y valores que interrelacionadas entre sí permiten tener un desempeño profesional eficiente, de conformidad con el estado del arte^{11,12,13}.

Una definición amplia de competencia profesional la considera como: el uso juicioso y habitual de la comunicación, los conocimientos, las habilidades técnicas,

el razonamiento clínico, las emociones, los valores, y la reflexión, en la práctica diaria en beneficio del individuo y de la comunidad a la que sirven¹⁴.

Las competencias se utilizan en cualquier tipo de estructura curricular ya sea por módulos, áreas y en los de asignaturas como por ejemplo en el Plan de Estudios 2010 que contempla en el perfil del egresado el cumplimiento de 8 competencias, dentro de éstas, la competencia 5 plantea las habilidades clínicas para el diagnóstico, pronóstico, tratamiento y rehabilitación.

Según Norman, la competencia clínica comprende un conjunto de atributos multidimensionales entre los que destacan:

- 1.- Habilidades clínicas: capacidad para adquirir información al interrogar y examinar pacientes e interpretar el significado de la información obtenida.
- 2.- Conocimientos y comprensión: capacidad para recordar conocimiento relevante acerca de condiciones clínicas que lleven a proveer atención médica efectiva y eficiente para los pacientes.
- 3.- Atributos interpersonales: expresión de aquellos aspectos de carácter profesional y personal del médico que son observables en las interacciones con pacientes.
- 4.- Solución de problemas y juicio clínico: aplicación del conocimiento relevante, habilidades clínicas y atributos interpersonales para el diagnóstico y manejo de los problemas de un paciente dado.
- 5.- Habilidades técnicas: capacidad para usar procedimientos y técnicas especiales en el estudio y manejo de pacientes¹⁵.

Los atributos anteriores y el amplio rango de sus componentes hacen compleja su evaluación con un solo instrumento. El desafío es entonces encontrar el grupo de instrumentos que nos permitan evaluar el espectro de componentes que más interesan.

Los instrumentos de evaluación del aprendizaje aceptables deben ser confiables, válidos y prácticos. De acuerdo con la pirámide de competencias de Miller (figura 1), hay distintos instrumentos que evalúan el “saber” y el “saber cómo”, entre ellos se encuentran: los exámenes orales, los ensayos y los reactivos de opción

múltiple que pueden aplicarse en contextos clínicos. Algunos instrumentos que son útiles para evaluar el “mostrar cómo” son: el caso largo o el examen ante paciente real, el Mini CEX, el Examen Clínico Objetivo Estructurado y el examen con pacientes estandarizados¹⁶. Hay todavía el reto para evaluar el “hacer”, es decir, el desempeño en el trabajo diario, con el desarrollo de mediciones confiables y que tengan validez predictiva de la competencia clínica subsecuente¹⁷.



Figura 1. Pirámide de las competencias de Miller

La finalidad de la evaluación del aprendizaje no es solamente la acreditación, sino que cada vez más se reconoce que su función principal debe ser esencialmente constructiva, de tal forma que permita elaborar un diagnóstico que conduzca a la identificación de los avances y deficiencias en la formación de los estudiantes y con base en esto se definan acciones que permitan consolidar logros y trazar planes de superación. Por ello, resulta necesario evaluar la práctica clínica en busca de explicaciones racionales para reorientar la enseñanza de la misma, reforzar la función tutorial y los programas académicos de los Planes de Estudios para enfocar las estrategias hacia los aspectos metodológicos y hacia un mayor desarrollo de las habilidades clínicas que se necesitan para el ejercicio profesional de la medicina general.

En el ámbito médico, la competencia clínica es evaluada por los profesores en exámenes ante pacientes reales y en evaluaciones intermitentes durante el proceso formativo, aunque este tipo de evaluación tiene varias desventajas, la principal es la escasa observación directa del desempeño clínico de los estudiantes por parte de los docentes¹⁸. Otras desventajas son el desempeño variable de los educandos entre un caso y otro, fenómeno conocido como especificidad de caso, y que la valoración con frecuencia es esporádica, oportunista, poco uniforme y subjetiva. Incluso profesores experimentados tienen diferencias cuando valoran un mismo evento, por lo que su validez y confiabilidad pueden ser limitadas^{19,20,21}. Esta forma de evaluación tradicional hace difícil valorar todos los componentes que integran la competencia clínica, para proporcionar información valiosa sobre el proceso formativo de los estudiantes^{6,22}. Los hallazgos acumulados en años anteriores sobre el poder predictivo de estos exámenes en los estudiantes de pregrado, y la ausencia de relación entre pruebas de conocimientos y otras mediciones llevaron a la conclusión de que, a pesar de que el conocimiento es un aspecto necesario de la competencia clínica, no es suficiente por sí mismo²³.

En relación con la evaluación en el caso largo, tiene como fortaleza que evalúa el desempeño del estudiante ante un paciente que representa un reto clínico real. El estudiante obtiene toda la información relevante del paciente a través del interrogatorio, la exploración física, el análisis de los estudios de laboratorio y gabinete, estructura el problema, sintetiza sus hallazgos, elabora el diagnóstico y formula un plan de manejo. La duración de este examen es de aproximadamente 2 horas. Sin embargo, la confiabilidad es baja debido a la baja reproducibilidad de este examen para todos los estudiantes¹⁹.

El MiniCEX es un tipo de examen para evaluar a los estudiantes ante pacientes reales, ya sea en consulta externa, en urgencias o en hospitalización, los profesores evalúan habilidades específicas y la competencia global finalizando con la realimentación al estudiante. Este tipo de examen tiene una duración de 5 a 30 minutos dependiendo de los objetivos del examen. Son necesarios 12 a 14 encuentros con pacientes para alcanzar una reproducibilidad de 0.80, suficiente

para un examen de altas consecuencias. El Mini CEX facilita observaciones múltiples en el tiempo por diferentes profesores. Esto mejora la validez y confiabilidad de las evaluaciones. El aspecto longitudinal es una de las principales fortalezas como herramienta y método de evaluación. Sin embargo, este se ha utilizado principalmente a nivel de residencias médicas en EUA²⁴.

Evaluación de la Competencia Clínica con el Examen Clínico Objetivo Estructurado

Se han hecho combinaciones de métodos múltiples para evaluar la competencia clínica, sin embargo un buen ejemplo de un solo evento con varios componentes es el Examen Clínico Objetivo Estructurado (ECOPE), el cual ha sido adaptado a una gran variedad de escenarios.

Harden et al. en 1975 iniciaron la observación directa con múltiples estaciones y lista de cotejo en el desarrollo del ECOPE, superando las desventajas del tradicional examen oral o caso largo, que demostró su baja confiabilidad y fue eliminado en EUA hace décadas²¹. Es necesario observar a los estudiantes interactuando con pacientes para evaluar sus habilidades clínicas, ya que entre más semejanza exista con la práctica clínica habitual, mayor valor tendrá la evaluación²⁵.

Por otra parte, el desempeño de un examinado en un caso o situación clínica no es un buen predictor de su desempeño en otro caso o situación. Esto se observa en el manejo de problemas de pacientes en pruebas escritas, simulaciones basadas en computadora, pacientes estandarizados y pacientes reales. La calidad del desempeño parece estar determinado más por el conocimiento específico y la experiencia que los estudiantes tienen en relación con cada caso o situación, que por sus habilidades generales para resolver problemas²¹. Estudios con médicos y estudiantes han encontrado incrementos en el desempeño con más educación y experiencia clínica, dando tiempo para la reflexión²⁶. Esto justifica un amplio muestreo de situaciones clínicas y el aumento de la amplitud de la prueba para alcanzar una adecuada confiabilidad.

Numerosos estudios en EUA y Canadá han demostrado que el ECOE es un examen práctico que tiene evidencia de validez y confiabilidad. Es uno de los métodos de evaluación de la competencia clínica de los estudiantes con mayor grado de objetividad^{27,28,29,30,31,32}.

Algunos de los componentes de la competencia clínica que se evalúan con este método son: interrogatorio, exploración física, interpretación de estudios de laboratorio y gabinete, diagnóstico, plan de manejo y comunicación en la relación médico-paciente³³; estos componentes se consideran en la evaluación que se realiza con esta metodología en la Facultad de Medicina de la UNAM

El ECOE consiste en que el estudiante, durante el examen, cambia y avanza por dieciocho estaciones, pasando seis minutos en cada una de ellas después de las cuales se trasladan a la siguiente estación. Cada estación presenta un problema diferente y hay un examinador con una lista de cotejo en cada una de ellas cuya utilización busca superar el problema de la inconsistencia de los evaluadores, calificando el desempeño de los examinados con formatos previamente validados por expertos con lo que aumenta la confiabilidad interevaluador. Al respecto se ha observado que el uso de listas de cotejo optimiza la calificación y disminuye el sesgo del evaluador; se argumenta que el uso de las escalas globales es más apropiado para evaluar estudiantes con niveles más altos de pericia³⁰.

En este trabajo se utilizan predominantemente las listas de cotejo en la evaluación de procedimientos debido a que los enunciados son más concretos y pueden juzgarse más objetivamente, aunado al proceso de validación y capacitación de los evaluadores.

Las escalas globales están sujetas a una mayor interpretación por parte del examinador, que lleva a una mayor tendencia a la subjetividad y menor confiabilidad intra e interevaluador³⁴. Al calificar con escalas globales hay varios errores o efectos que pueden cometer los evaluadores como son: el *efecto halo*, calificar más o menos igual a todos los examinados; el *efecto horn*, es la tendencia a generalizar la evaluación de una persona en un factor negativo a todos los aspectos del desempeño de esa persona; *la restricción de rango*, tendencia a calificar alrededor del centro y no utilizar toda la escala; *la inconsistencia*,

tendencia a aplicar escalas de calificación diferentes a las de otros evaluadores; la *leniencia*, los evaluadores tienden a comportarse muy blandos (paloma) o muy duros (Halcón). En general se dice que hay muy pocos halcones^{35,36}.

La confiabilidad interevaluador es el grado en que dos o más evaluadores concuerdan en la asignación de la calificación a un estudiante en una estación particular³⁵. Algunos estudios señalan que el uso de listas de cotejo con dos evaluadores por estación solo agrega una diferencia marginal a la confiabilidad por lo que se considera en el presente estudio que es suficiente contar con un evaluador en cada estación³⁸.

El ECOE, usado en combinación con los formatos escritos, puede incrementar la confiabilidad, si se incluye una cantidad suficiente de estaciones para lograr una muestra amplia de situaciones clínicas. Constituye una prueba flexible en la cual se pueden usar una variedad de métodos para obtener información amplia que permita evaluar las habilidades clínicas. El concepto de explorar varios componentes de la competencia clínica para mejorar la confiabilidad de la evaluación, es análogo a usar preguntas de opción múltiple para evaluar el conocimiento de los estudiantes³⁷.

Los criterios de evaluación son determinados por las competencias que deben alcanzar los estudiantes de acuerdo con los objetivos del plan de estudios y los programas académicos correspondientes¹⁵. Pueden utilizarse pacientes reales, estandarizados o maniqués según el problema a evaluar²⁷.

Las limitaciones del ECOE son:

Los conocimientos y destrezas del estudiante son examinados en compartimentos y no se valora la habilidad para ver a un paciente en forma integral; la naturaleza restrictiva del formato de calificación; la artificialidad de alguna de las estaciones; la naturaleza relajada de las tareas para los examinadores; el tiempo utilizado y el personal que participa en la elaboración y aplicación, el cual es mayor que en los exámenes tradicionales³⁷.

El ECOE es un instrumento de evaluación costoso debido a que se tiene que pagar a los pacientes, a los profesores que participan en la estructuración y en la aplicación así como la infraestructura que requiere. Estas actividades requieren un

proceso de capacitación de profesores continuo y amplio principalmente de las seis áreas de Internado para elaborar y validar estaciones que constituyan un banco que puede ser utilizado para estructurar varias versiones de examen de acuerdo con el número de estudiantes a evaluar.

La elaboración de estaciones debe ser continua y los cursos de capacitación de profesores son una oportunidad para solicitar la elaboración de nuevas estaciones que contribuyan al banco. Asimismo, los profesores deben tener una actitud favorable hacia el ECOE para participar en los exámenes de tipo formativo que no implican un pago.

En forma simultánea es necesario capacitar a los pacientes reales y estandarizados para participar en los exámenes, es decir; se les debe evaluar continuamente para que su participación sea acorde con los mejores estándares del ECOE.

Es importante considerar que ciertas habilidades técnicas no se pueden realizar en pacientes por cuestiones éticas o médicas, tales como tomar una muestra para estudio de Papanicolaou, aplicar un DIU, explorar la próstata, colocar una sonda, realizar una sutura, realizar maniobras de RCP, de tal manera que es necesario disponer de recursos financieros para adquirir modelos de varios tipos para estos procedimientos que también son necesarios evaluar. Todos estos retos se pueden superar cuando se tiene el financiamiento y capacitación adecuada de los formadores de recursos humanos en relación con el ECOE.

El ECOE a nivel nacional

El ECOE se inició en nuestro país en 1996 con el apoyo de un proyecto PAPIME para evaluar el desarrollo de la competencia clínica en un grupo del Núcleo de Calidad Eucativa (NUCE) de 30 estudiantes que estaban realizando el Internado en el Hospital "Dr. Manuel Gea Gonzalez"³⁹.

En el período comprendido entre 1998 y 2001 se aplicó el ECOE de tipo formativo en estudiantes durante el Internado Médico. En 2002 se comenzó a aplicar como parte de la Fase Práctica del Examen Profesional como alternativa al examen con un solo paciente. Del mismo modo, se aplicó de 2002 a 2005 durante 13 a 17

sábados consecutivos, en una sede, que se realizaban mientras los sustentantes ya estaban en el servicio social⁴⁰.

Como parte del proyecto para probar el ECOE a fin de asignar calificaciones en la práctica clínica, durante el año 2006 y 2007 se evaluaron, con esta metodología, 1347 estudiantes de la UNAM, así como a 355 estudiantes de diversas escuelas y Facultades de los estados, las cuales integraban el Proyecto Interinstitucional para la evaluación de las competencias al inicio del Internado en el proceso de formación de médicos en México. Estas acciones se llevaron a cabo durante 20 sábados en Unidades médicas de primer nivel de atención.

Hay estudios que comparan las calificaciones que da el profesor con el resultado del ECOE, los cuales sugieren mayor objetividad en éste para determinar la competencia real de nuestros alumnos y así poder garantizar el nivel de competencia clínica de los estudiantes⁴¹.

En junio de 2008 se aplicó en el examen profesional a un grupo de 40 estudiantes y en los años subsiguientes, 2009 y 2010, a 720 y 675 estudiantes en dos días utilizando seis sedes clínicas simultáneamente.

En 2009, el promedio del alfa de Cronbach fue de 0.64 ⁴², y en enero 2010, el promedio del alfa de Cronbach fue de 0.56 ⁴³.

Los estudiantes en la Facultad de Medicina aplican un examen teórico con duración de 10 horas en dos días y con un alfa de Cronbach de 0.92. Aprobar la fase teórica es requisito para pasar a la Fase Práctica con el Examen Clínico Objetivo Estructurado, obteniéndose de éste un promedio de alfa de Cronbach de 0.63 que consideramos se encuentra dentro de los criterios internacionales adecuados para un examen de 2 h. de duración^{26,30}.

Validez

La validez se refiere a la evidencia presentada que apoye o refute el significado o la interpretación asignada a los resultados de la evaluación. La validez es *sine qua non* de la evaluación, de tal manera que sin evidencia de validez, la evaluación en educación médica tiene poco o ningún significado.

En su conceptualización contemporánea la validez es un concepto unitario que mira hacia múltiples fuentes de evidencia y toda validez es de constructo⁴⁴. En la actualidad la evaluación en ciencias sociales incluyendo la educación médica, trata con los constructos- colecciones intangibles de conceptos abstractos y principios que se infieren de la conducta y son explicados por teorías educacionales o psicológicas. El desempeño educacional es un constructo, generalmente inferido del desempeño en las evaluaciones escritas, orales o del ECOE⁴⁵.

Las cinco fuentes de validez de constructo en evaluación educativa son: el contenido (las especificaciones de la prueba, el contenido temático); el proceso de respuesta (el control de calidad del examen, la familiaridad del estudiante con el formato de evaluación); la estructura interna (el análisis de reactivos, la confiabilidad del examen); su relación con otras variables (generalización, exámenes escritos, promedio de años previos, resultados en poblaciones de expertos en comparación con novatos) y las consecuencias (método de establecimiento del punto de corte para aprobar, consecuencias para el estudiante y para la sociedad)^{45,46}.

Los estudios que se han realizado sobre la competencia clínica parecen confirmar que los individuos con mayor entrenamiento clínico se desempeñan mejor en el ECOE que los principiantes^{30,31,32}.

Confiabilidad

La confiabilidad se refiere a la consistencia de los resultados de la evaluación, de tal manera que estos puedan ser reproducibles cuando se aplica en situaciones semejantes, es decir hasta qué punto el instrumento arroja los mismos resultados al repetir la prueba. Por ello, es un concepto estadístico que representa el grado en el cual las puntuaciones de los estudiantes serían similares si fueran examinados de nuevo, y en el cual el instrumento mide el fenómeno de manera constante en el tiempo^{46,47}.

En el modelo de la teoría clásica del test⁴⁸, la calificación observada es la suma de dos componentes: la calificación verdadera y el error. Desde la perspectiva de la

medición, el error es un aspecto inherente al proceso de la medición.

Cuadro 1. Modelo de la teoría clásica del test.

$*X = **V + ***E$
* X= Calificación observada **V= Calificación verdadera ***E= Error

El error estándar de la medición (EEM) es muy útil para describir la confianza en una puntuación.

El EEM proporciona la distribución del error asociado con la calificación. Pueden dar confianza los intervalos alrededor de la calificación usando el EEM. El error de medición surge de múltiples fuentes.

Para un ECOE, el error de la medición puede atribuirse a: a) La selección de un grupo particular de estaciones, b) los efectos de la portabilidad de los pacientes estandarizados, c) Los efectos ocasionales, d) Los efectos del evaluador, e) los escenarios (si el ECOE se aplica en múltiples sitios).

Se necesita identificar claramente estas fuentes y dirigir las *a priori*. Se pueden cuantificar estas fuentes de error usando la teoría G⁴⁹.

Spearman fue el primero que trabajó sobre el error de medición y señala tres supuestos:

El primero es definir la puntuación verdadera (V) como la esperanza matemática de la puntuación empírica, que formalmente puede escribirse así: $V = E(X)$. Lo que esto significa conceptualmente es que la puntuación verdadera de una persona en un examen es aquella puntuación que obtendría como media si se le pasase infinitas veces el examen. Se trata de una definición teórica, nadie va a pasar infinitas veces un examen a nadie, por razones obvias, pero es plausible pensar que si esto se hiciese, la puntuación media que esa persona obtenga en los exámenes sería su verdadera puntuación.

El segundo supuesto asume que no existe relación entre la cuantía de las puntuaciones verdaderas de las personas y el tamaño de los errores que afectan a

esas puntuaciones. En otras palabras, que el valor de la puntuación verdadera de una persona no tiene nada que ver con el error que afecta esa puntuación. Por tanto, puede haber puntuaciones verdaderas altas con errores bajos, o altos, no hay conexión entre el tamaño de la puntuación verdadera y el tamaño de los errores. De nuevo se trata de un supuesto en principio razonable, que formalmente puede expresarse así: $r(v, e) = 0$.

El tercer supuesto establece que los errores de medición de las personas en un examen no están relacionados con los errores de medición en otro examen distinto. Es decir, no hay ninguna razón para pensar que los errores cometidos en una ocasión vayan a covariar sistemáticamente con los cometidos en otra ocasión. Formalmente este supuesto puede expresarse así: $r(e_j, e_k) = 0$. Estos supuestos parecen razonables y sensatos, pero no se pueden comprobar empíricamente de forma directa, serán las deducciones que luego se hagan a partir de ellas las que permitan confirmarlas o falsearlas⁴⁸.

El alfa de Cronbach es un ejemplo de coeficiente de confiabilidad basado en la teoría clásica del test, dado que, nos da una estimación de varianza de puntuaciones observadas de estación a estación a través de un examen que tiene varianza de puntuaciones verdadera.

El Alfa de Cronbach es una medición de la consistencia interna, por lo cual en una evaluación los mejores estudiantes deberían desempeñarse relativamente bien a través del examen (por Ej. en la puntuación de las listas de cotejo de cada estación)^{50,35}.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum S_i^2}{S_{sum}^2} \right)$$

k = número de enunciados de la prueba

S_i^2 = varianza de los enunciados

S_{sum}^2 = varianza de la prueba total

El coeficiente mide la confiabilidad del test en función de dos términos: el número de enunciados (o longitud de la prueba) y la proporción de varianza total de la prueba debida a la covarianza entre sus partes (enunciados). Ello significa que la confiabilidad depende de la longitud de la prueba y de la covarianza entre sus enunciados. En nuestro caso va a depender de las 18 estaciones, la covarianza entre éstas y de las 2 horas de duración de la misma.

Se pueden calcular dos tipos de alfa, la no estandarizada y la estandarizada, comúnmente la segunda siempre es más alta. Esta es una medición de la media de intercorrelación calculada entre varianzas y da como resultado el mismo valor que la teoría G para un simple modelo cruzado de enunciados con los estudiantes. El valor del alfa va de 0 a 1 y estima el error existente, a medida que se aleja de 1 aumentan los errores de medición. El valor total del alfa, que se considera aceptable en los exámenes de altas consecuencias, donde participan pacientes reales y estandarizados y la métrica de la estación individual, es arriba de 0.70⁵¹.

Puede considerarse que el alfa de Cronbach tiene limitaciones para estimar la confiabilidad del ECOE debido a que se enfoca en una faceta simple de error de medición y toma en cuenta el orden por categorías de los estudiantes más que su desviación absoluta de un estándar⁵².

Dentro de la teoría clásica del test no existe un sistema de notación común para describir múltiples fuentes de error o definiciones alternativas de calificación verdadera y error. La teoría de la generalizabilidad (teoría G) proporciona un sistema de notación, distinciones conceptuales críticas y procedimientos computacionales relacionados. En lugar de la calificación verdadera, la teoría G define una calificación universal, que es el valor esperado, sobre el universo de generalización de las calificaciones observadas para un único objeto de medición.

Teoría de la Generalizabilidad

La teoría de la generalizabilidad se puede ver como una extensión de la confiabilidad en la teoría clásica del test a través de la aplicación de ciertos procedimientos del análisis de varianza (ANOVA) para tópicos de medición.

La teoría G fue introducida por Cronbach y cols. en respuesta a las limitaciones del popular modelo de respuesta verdadera de la teoría clásica del test⁵³.

Cronbach et al. (1972) definieron el coeficiente de generalizabilidad, Ep^2 análogo al coeficiente de confiabilidad de la teoría clásica del test, como la relación de la varianza del universo de calificaciones con el valor esperado de la varianza de las calificaciones observadas^{47,50}.

En la teoría de la Generalizabilidad se utilizan dos coeficientes, el G o de Generalizabilidad y el D o índice de dependabilidad.

El coeficiente de generalizabilidad Ep^2 es la proporción de la varianza del universo de puntajes entre el mismo más la varianza del error relativo⁵².

$$Ep^2 = \frac{\sigma^2 (r)}{\sigma^2 (r) + \sigma^2 (\delta)}$$

Ep^2 = coeficiente de generalizabilidad

$\sigma^2 (r)$ = proporción de la varianza del universo de puntajes de los estudiantes

$\sigma^2 (\delta)$ = varianza del error relativo

El índice de dependabilidad Φ es la proporción de la varianza del universo de puntajes entre el mismo más la varianza del error absoluto⁵⁴.

$$\Phi = \frac{\sigma^2 (r)}{\sigma^2 (r) + \sigma^2 (\Delta)}$$

Φ = índice de dependabilidad

$\sigma^2 (r)$ = proporción de la varianza del universo de puntajes

$\sigma^2 (\Delta)$ = varianza del error absoluto + la varianza de otras fuentes de variación

Debido a que $\sigma^2 (\Delta)$ es mayor que $\sigma^2 (\delta)$ por consiguiente Φ es menor que Ep^2 .

El índice de dependabilidad es apropiado cuando se dan interpretaciones de las puntuaciones absolutas, como en situaciones referidas al dominio o referidas al criterio.

Para profundizar en el estudio de la teoría G es necesario definir algunos conceptos que nos permitan entender las relaciones entre los componentes.

Coficiente de generalizabilidad e índice de dependabilidad.

Faceta

Facetas anidadas

Universo de generalización

Universo de calificaciones

Población

Diseños cruzados y anidados

Error relativo y absoluto

Estudios G de generalizabilidad y Estudios D de decisión

Faceta. Es una fuente mayor de variación o de error de medición que puede restar mérito a la confiabilidad de las puntuaciones del ECOE, en nuestro caso, las estaciones, los estudiantes, las sedes y las versiones.

Facetas anidadas

Dos o más niveles de la faceta anidada aparecen con un nivel de otra faceta. En nuestro estudio tenemos la faceta estudiante anidada dentro de otras facetas.

Universo de generalización

Una medición es una muestra del universo de observaciones admisibles o generalización, que deseamos tratar como intercambiables (todas las estaciones, todas las versiones y sedes).

Similarmente, si intentamos generalizar los resultados de la muestra de estaciones a todas las estaciones posibles del ECOE en Medicina, entonces el universo de estaciones se definiría como todas las posibles estaciones del ECOE en Medicina. El error de medición siempre está presente cuando decidimos hacer estas generalizaciones.

Si todas las estaciones en el universo fueran igualmente difíciles y todos los posibles evaluadores fueran igualmente estrictos, entonces una calificación de un estudiante debería ser el mismo a través de las estaciones, versiones y sedes.

En realidad, la dificultad de la estación y lo estricto de los evaluadores varía por lo que la generalización a través de las estaciones y los evaluadores contiene errores de muestreo.

Universo de calificación

Con nuestro ECOE, una calificación del universo de estudiantes corresponde a la calificación media que él o ella obtendrían a través de todas las posibles estaciones del ECOE y todas las posibles sedes y versiones. Idealmente, queremos estimar cuál es la calificación del universo de estudiantes y cuáles son sus competencias clínicas y no solo con respecto a las estaciones específicas, sedes y versiones que hemos seleccionado para el examen.

Debido a que necesitamos estimar el universo de calificación, es necesario evaluar que tan confiablemente se pueden generalizar las calificaciones de los estudiantes desde estas estaciones y evaluadores al universo de todas las estaciones y evaluadores admisibles de Medicina.

El concepto **población** se reserva a todos los estudiantes que puede haber en la Facultad de Medicina, en nuestro caso los 278 estudiantes que constituyen la muestra y fueron seleccionados de forma aleatoria de todas las sedes del Internado del DF para el ECOE. La variabilidad entre éstos reflejan las diferencias en conocimientos y habilidades, etc. La varianza del universo de calificación es análogo a la varianza de la calificación verdadera en la teoría clásica del test⁴⁹.

En el presente estudio, **el universo** se reserva para las condiciones de medición o facetas: las estaciones, los estudiantes, las sedes y las versiones. Las estaciones, los evaluadores y los pacientes estandarizados fueron considerados como una faceta única de análisis.

Diseños cruzados

Si se acepta como admisible a cualquier estudiante de la población para ver cualquier caso con cualquier estación entonces decimos que la población está cruzada con el universo de observaciones admisibles. En un diseño cruzado todos los niveles de cada faceta son observados con el objeto de la medición.

Cuadro 2. Sistema de notación con teoría G.

${}^1p * : {}^2s {}^3v **x {}^4s$
¹ p= persona (estudiante u objeto de la medición) ² s= sede ³ v= versión ⁴ e= estación * “.” significa “anidado dentro de” ** “x”, significa “cruzado con”

En nuestro ECOE el diseño **p : s v x e** significa que los estudiantes(p) están anidados dentro de sede(s) y de versión(v) y cada estudiante está cruzado con cada estación(e). Las versiones están cruzadas con las sedes. Las estaciones están cruzadas con las sedes y con las versiones. Se utilizó el software JMP versión 11, que fue desarrollado principalmente para el análisis de los modelos estadísticos lineales tanto con efectos fijos como efectos aleatorios.

Estudios G de generalizabilidad y Estudios D de decisión

El diseño, recolección de datos y el análisis para estimar los componentes de varianza se conocen como estudios G o de generalizabilidad

El uso de los resultados del estudio G para estudiar diseños alternativos que ayuden a resolver preguntas específicas y tomar decisiones son llamados estudios D o de decisión.

Componentes de varianza en la teoría G

Los componentes de varianza para las facetas en nuestro diseño nos señalan la cantidad relativa de error de muestreo con que contribuye cada faceta. En la Teoría de la Generalizabilidad, los componentes de varianza asumen una importancia central. Para las estaciones, el componente de varianza produce un estimado de que tan diferente se espera que sea la media de calificaciones (a través de las personas) para una estación seleccionada aleatoriamente de todas las estaciones en el universo. Para los estudiantes nos indican la magnitud del error comprendido en la generalización de la calificación obtenida por un estudiante a su universo de calificación (promedio a través de todas las estaciones, sedes y versiones en el universo).

Los componentes de varianza se estiman, de forma preferente, a través del análisis de varianza⁵⁴.

EFFECTOS EN EL ESTUDIO

$$\begin{aligned}
 X_{pevs} &= \mu \\
 &+ \mu_{pvs} - \mu_{vs} \\
 &+ \mu_e - \mu \\
 &+ \mu_v - \mu \\
 &+ \mu_s - \mu \\
 &+ \mu_{pe} - \mu_p - \mu_e + \mu \\
 &+ \mu_{pv} - \mu_p - \mu_v + \mu \\
 &+ \mu_{ps} - \mu_p - \mu_s + \mu
 \end{aligned}$$

Variación entre efectos

(gran media)
 efecto persona[version, sede]
 efecto estación
 efecto versión
 efecto sede
 efecto persona x estación
 efecto persona : versión
 efecto persona : sede

$+\mu_{ev} - \mu_e - \mu_v + \mu$	efecto estación x versión
$+\mu_{es} - \mu_e - \mu_s + \mu$	efecto estación x sede
$+\mu_{vs} - \mu_v - \mu_s + \mu$	efecto versión x sede
+	el efecto residual ⁵⁵

Los errores relativo y absoluto varían como función del diseño del estudio G y de la forma en que son tratadas las facetas (fijas o aleatorias).

Esta es una aproximación útil para monitorear el ECOE como un todo, que identifica cuáles facetas de la evaluación (p.ej. las estaciones, las sedes, las versiones o los estudiantes) son la mayor fuente de error de medición. Sin embargo, no especifica cuál elemento individual dentro de una faceta es la causante del error. De esta manera los estudios con la teoría G pueden sugerir cómo mejorar la calidad a través de las intervenciones grupales (p.ej, aumentando el número de estaciones, de versiones o de sedes)³⁶.

La teoría clásica del test y la teoría G descomponen la varianza observada en calificación verdadera y componentes de varianza. Sin embargo, la fortaleza de la teoría G reside en que se pueden estimar múltiples fuentes de error en una medición, en forma separada, en un solo análisis y permite al tomador de decisiones determinar el número de ocasiones de examen, los formatos de prueba y las estaciones, versiones y sedes que se necesitan para obtener calificaciones confiables⁵⁵. En este proceso se obtiene un coeficiente de generalizabilidad que refleja el nivel de confiabilidad, considerándose adecuado un nivel de 0.80 para un examen de altas consecuencias. Sin embargo es difícil alcanzar este nivel en un examen de 2 hs por lo que 0.60 se considera aceptable para esta duración de un ECOE⁵⁶.

Uno de los propósitos de la teoría G es evaluar las mayores fuentes de variación de tal manera que la no deseada puede reducirse al recolectar futuros datos. En esta aplicación estamos pensando acerca del rediseño de la medición (por ejemplo, la longitud del examen), más que evaluar las mediciones ya realizadas. Para el diseño de una faceta, hay que estimar e interpretar la magnitud de los componentes de la varianza para los enunciados y residuales comparados con el universo de las calificaciones⁵⁵.

Lawson (2006) realizó un estudio con la teoría de la generalizabilidad en un ECOE de altas consecuencias en la licenciatura, donde concluye que la teoría G puede contribuir a la comprensión de fuentes de varianza y dar dirección al mejoramiento de las estaciones individuales⁵⁷. Brailovsky (2000) realizó un estudio de generalizabilidad para evaluar el impacto del sitio de examen (4), circuitos (14) e idiomas (Inglés y Francés) sobre los resultados de los candidatos en el examen de licenciatura con el ECOE que tiene una duración de 5.5 hs conformado de 26 casos clínicos de 7, 10 o 15 minutos de duración, donde confirma la calidad del examen, asegurando su continua mejoría y apoyando las inferencias hechas de los resultados del mismo⁵⁸.

En una revisión sistemática de confiabilidad del ECOE se encontró una media no ponderada de Alfa de Cronbach de 0.62 y un coeficiente de generalizabilidad no ponderado de 0.49⁵².

Los estudios con la teoría G se llevan a cabo para investigar la relación entre una calificación observada y la calificación del universo para obtener información de las fuentes de variabilidad (i.e., facetas del universo) que influyen en la generalización de las observaciones. Debido a que su medición depende de su propósito y del universo de observaciones admisibles que representan, se pueden necesitar diferentes estudios G en función de la aplicación de la medición propuesta³⁶.

Un ejemplo de éste es el reporte de la prueba piloto para el examen de licenciatura del Consejo de Canadá de 1992 en donde el análisis de generalizabilidad del efecto de las versiones y el sitio para 120 candidatos aleatorizados mostró una varianza de acuerdo al sitio de 0.16 comparada con la varianza total que fue de 280.86 indicando que virtualmente no había diferencia en las calificaciones de los candidatos atribuible al sitio asignado³¹.

En el aspecto psicométrico, la teoría de la generalizabilidad puede ayudarnos a diseñar evaluaciones que permitan hacer un uso eficiente de los recursos, tales como el tiempo de los profesores, manteniendo la confiabilidad de los resultados¹⁰.

Los estudios G se usan para estimar la varianza de los componentes. Su propósito es anticipar los múltiples usos de las mediciones y proporcionar tanta información

como sea posible acerca de las fuentes de variación de la medición. De este modo, un estudio G debería intentar identificar e incorporar en sus diseños tantas fuentes de variación como sea posible. O de otra forma, debería definir el universo de observaciones admisible tan ampliamente como sea posible.

Por otra parte, los estudios D hacen uso de la información proporcionada por los estudios G para diseñar las mejores aplicaciones posibles de medición en ciencias sociales y tomar decisiones para un propósito particular. En la planeación de un estudio D, el tomador de decisiones hace lo siguiente: a) define un universo de generalización -número y amplitud de las facetas sobre las cuales se va a generalizar; b) especifica la interpretación propuesta de la medición -relativa o absoluta; la interpretación propuesta define el error de medición e identifica las fuentes de error de mayor importancia; y c) usa la información del estudio G acerca de la magnitud de varias fuentes de error de medición para evaluar la efectividad de diseños alternativos para minimizar el error y maximizar la confiabilidad. Esta evaluación se hace de manera análoga a la profecía de Spearman-Brown en la teoría clásica del test. Se incrementa el número de condiciones de una faceta en una medición (p.ej. aumentando el número de observadores que califican cada conducta), el error con el que contribuye cada faceta puede disminuirse, en gran medida añadiendo ítems al test con lo que disminuye el error (y aumenta la confiabilidad) en la teoría clásica del test. De esta forma se puede desarrollar un diseño costo-beneficio para la aplicación de mediciones en ciencias sociales⁵⁵.

En el presente estudio, tuvimos en cada una de las seis sedes de examen, un grupo de 20 estudiantes, otro grupo de 18 profesores, 18 estaciones con casos con diferente nivel de dificultad, trece de las cuales tuvieron pacientes estandarizados y cinco fueron estaciones de interpretación de estudios de laboratorio y gabinete; además cuatro sesiones de examen con versiones equivalentes. Cada sede clínica tuvo tres grupos de estudiantes, pacientes y profesores por día, sin embargo, el profesor y el paciente se encontraban ligados a cada estación. En este caso tuvimos un diseño cruzado con cuatro facetas, los

estudiantes, las estaciones, las sedes y las versiones. Por otra parte, los estudiantes estaban anidados en las cuatro versiones y las seis sedes del ECOE.

Los estudiantes en el Plan de Estudios de la Facultad de Medicina de la UNAM, desde sus años clínicos (tercero y cuarto año), tienen contacto con pacientes y analizan las enfermedades con sus profesores, desarrollan su competencia clínica en mayor o menor grado de acuerdo a su interés y la exposición a experiencias repetidas y reflexivas⁵⁹.

El Internado Médico de pregrado

El término Internado, apareció oficialmente en el plan de estudios publicado en el Diario Oficial de la Federación de México, en enero de 1912, en el cual se decretó que los alumnos del sexto año de la carrera de Medicina deberían permanecer dentro del hospital para atender personalmente y bajo su responsabilidad los casos médicos, quirúrgicos y obstétricos señalados por las disposiciones reglamentarias.

A partir de 1934 se estableció que la práctica en hospitales debería realizarse durante un año, en jornadas de 10 horas diarias, ofreciendo a los estudiantes la oportunidad de elegir la adscripción a los servicios que convenían a sus intereses, y se hizo obligatorio en 1952, también por decreto presidencial.

Con base en el estudio realizado en 1956 por la Comisión Nacional de Internados y Residencias, se asentó que las prácticas hospitalarias del sexto año de la carrera, se realizarían a tiempo exclusivo bajo el nombre de Internado Rotatorio de Pregrado, exigiéndose la permanencia y rotación en los servicios asistenciales básicos.

En 1983 se creó la Comisión Interinstitucional para la Formación de Recursos Humanos para la Salud (CIFRHS) y el Comité de Enseñanza de Pregrado y Servicio Social, que emitió las normas académicas y administrativas para el desarrollo del Internado. En 1986 el Internado se constituyó como una etapa eminentemente práctica, diseñada para que los alumnos, dediquen el mayor número de horas a realizar actividades que les permitan adquirir las competencias

básicas que pondrán en práctica al egreso⁶⁰.

Actualmente, el Internado Médico es una asignatura teórico-práctica que corresponde al quinto año de la carrera del Plan de estudios de la Facultad de Medicina. Tiene una duración de doce meses con 4067 horas teórico-prácticas y le corresponden 82 créditos. El estudiante realiza el Internado Médico en doce meses con un sistema de rotación, en el que pasa la mayor parte del tiempo en unidades médicas de primer y segundo nivel de atención. Existen rotaciones bimestrales por las seis áreas disciplinares que son: Cirugía, Ginecobstetricia, Medicina Familiar, Medicina Interna, Pediatría y Urgencias. Al término de cada una de éstas se les aplica un examen teórico con casos clínicos, con 50 preguntas y 5 opciones de respuesta. Al finalizar el año de Internado se promedian las 6 calificaciones para dar el promedio global del mismo.

Durante este año académico, el estudiante se incorpora a tiempo exclusivo a las unidades médicas de las instituciones de salud y mediante asesoría del profesor y del personal de la institución, colabora activamente en tareas asistenciales, acción que le permite integrar y aplicar los conocimientos adquiridos previamente, de ciencias básicas y clínicas, así como ejercitar las habilidades y destrezas médico-quirúrgicas necesarias para la prevención, diagnóstico, tratamiento y rehabilitación de los padecimientos más frecuentes en el país. Se trata en esencia, de una etapa de la carrera eminentemente práctica, dedicada especialmente a consolidar y fortalecer los aspectos psicomotrices y afectivos.

El Objetivo general es formar un médico capaz de ejercer la práctica de la medicina general con los conocimientos, la calidad técnica, la ética profesional y el humanismo que requieren el cuidado y la promoción de la salud.

El programa para cada una de las seis rotaciones por las áreas está conformado por tres apartados:

- 1.- Objetivo. Desarrollar en el alumno las capacidades para integrar y reconocer el diagnóstico clínico; aplicar las medidas preventivas y terapéuticas en el primer contacto de las patologías más frecuentes en nuestro medio de acuerdo con la morbimortalidad en el país y que competen al médico general e identificar la patología que requiera ser referida al nivel de atención correspondiente.

2.- Subprograma de aprendizaje autodirigido

a) Temas de autoaprendizaje. Conformado por los temas del área que el alumno debe estudiar de acuerdo con las necesidades detectadas a través de la evaluación diagnóstica.

b) Estudio de casos clínicos. Conformado por los problemas más frecuentes a los que se enfrenta el médico general.

3.- Subprograma de práctica clínica, en el que se desglosan las habilidades clínicas a desarrollar.

El programa académico está organizado y se lleva a cabo en todas las sedes de las cinco instituciones de salud que son: IMSS, ISSSTE, Secretaría de Salud, Gobierno del Distrito Federal y privados para lo cual existe un compromiso de cumplimiento, supervisiones continuas por parte de la Facultad de Medicina así como reuniones bimestrales con los jefes de grupo de estudiantes y con los tutores coordinadores para realimentar el cumplimiento. En el año de 2010 había convenios con 37 hospitales de las cinco instituciones en el país que dan cabida a 915 estudiantes, 25 de los hospitales están ubicados en el Distrito Federal con 708 estudiantes y el cupo de cada una de las sedes varía de 10 a 60 estudiantes⁶¹. El proceso de selección de la plaza para realizar el Internado médico se realiza en un evento público, de acuerdo con el promedio obtenido en los 4 años previos comenzando por el más alto.

Es indispensable considerar que se requieren de sistemas de evaluación más objetivos que ofrezcan información para la toma de decisiones en la mejora del proceso enseñanza-aprendizaje y el logro de las competencias planteadas⁶².

III. Justificación

Los esfuerzos para evaluar la competencia clínica no han sido sistemáticos ni completos y se requiere valorar la competencia clínica que nos permita fortalecerla. En este sentido, el Examen Clínico Objetivo Estructurado se considera el estándar de oro para evaluar la competencia clínica lo que nos permitió sistematizar ésta y con ello contar con un instrumento que forme parte del sistema de evaluación.

El quinto año de la carrera o Internado médico es crucial tanto para la enseñanza clínica como para evaluar las competencias ya que es un año donde todavía se puede dar realimentación al estudiante para fortalecer sus habilidades y mejorar su desempeño clínico de acuerdo con el perfil de egreso.

Es importante que los estudiantes sean evaluados formativamente con el ECOE de forma sistemática, porque se utiliza como una modalidad en la Fase Práctica del Examen Profesional, como examen de altas consecuencias debe tener la validez y confiabilidad necesarias.

Este trabajo además es relevante debido a que, no existen estudios que valoren la confiabilidad con la teoría de la generalizabilidad del ECOE en nuestro país.

IV. Planteamiento del Problema

La competencia clínica es una actividad esencial y compleja, ya que comprende un conjunto de atributos multidimensionales como son: los conocimientos, las habilidades clínicas, técnicas y de comunicación interpersonal, solución de problemas y el juicio clínico. En este sentido, el Examen Clínico Objetivo Estructurado se considera el estándar de oro para evaluar la competencia clínica, además de que ha demostrado que tiene validez y confiabilidad, es práctico, y tiene un mayor grado de objetividad en comparación con otras estrategias y no es utilizado como evaluación formativa en el Internado.

El alfa de Cronbach, tiene limitaciones para estimar la confiabilidad del ECOE debido a que se enfoca a una faceta simple de error de medición y considera el orden por categorías de los estudiantes más que su desviación absoluta de un estándar. Una alternativa para valorar la confiabilidad es la teoría de la generalizabilidad que es una extensión de la teoría clásica del test a través de la aplicación de ciertos procedimientos del análisis de varianza (ANOVA) para temas de medición.

El Internado es la etapa de la licenciatura donde existen las condiciones para el desarrollo de la competencia clínica por parte de los estudiantes, ya que tienen una mayor cercanía y responsabilidad con los pacientes; se encuentran la mayor parte del año en hospitales generales bajo la supervisión de sus profesores; pero están sujetos a un proceso de evaluación que no es uniforme ni está sistematizado en el área clínica a pesar de ponderarse con un 60% en la evaluación global. En contraste, los exámenes departamentales escritos bimestrales con casos clínicos y aunque se pretende que sean homogéneos para todos los estudiantes, por razones de seguridad no son los mismos para todos ellos a lo largo del año y tienen una ponderación del 40%.

Por lo anterior, se observa que no hay una evaluación válida y confiable de los estudiantes de Internado que valore el desarrollo de la competencia clínica objetivamente y los realmente en su desempeño, así como a los profesores y a la propia Facultad de Medicina de la UNAM.

Las preguntas de investigación son las siguientes:

- ¿Cuál es el impacto del programa de Internado médico en el desarrollo de la competencia clínica de los estudiantes?
- ¿Cuál es la validez y confiabilidad del ECOE aplicado en estudiantes de Internado de la Facultad de Medicina de la UNAM?

V. Objetivos Generales

- Evaluar el impacto del Internado médico en el desarrollo de la competencia clínica de los estudiantes.
- Valorar la validez y confiabilidad del Examen Clínico Objetivo Estructurado mediante el alfa de Cronbach y la teoría de la generalizabilidad.

Objetivos específicos

- Valorar el nivel de competencia clínica de los estudiantes al terminar el 4º año que pertenece al área clínica.
- Identificar los niveles de competencia clínica de los estudiantes al término del Internado según Hospital e Institución sede.
- Valorar el nivel predictivo de las calificaciones del Internado, en la fase práctica del examen profesional con el ECOE.
- Identificar el nivel predictivo de las calificaciones del examen profesional fase teórica sobre el desempeño en la fase práctica con el ECOE.
- Identificar las diferencias en el nivel de competencia clínica de las 6 áreas de Internado al inicio y al final del mismo
- Identificar las diferencias en los componentes de la competencia clínica que se evalúan.

HIPÓTESIS nulas:

1. No hay diferencia en la competencia clínica entre el inicio y el final del Internado médico de pregrado.
2. La institución del sector salud en la que se realiza el internado médico de pregrado no tiene efecto diferencial en el desarrollo de la competencia clínica de los estudiantes.

HIPÓTESIS de Trabajo:

1. El Internado médico de pregrado tiene un efecto positivo en el desarrollo de la competencia clínica de los estudiantes.
2. Hay diferencia en el nivel de competencia clínica entre los estudiantes de las distintas instituciones del sector salud.
3. La confiabilidad con el estudio de generalizabilidad es mayor de 0.60.

HIPÓTESIS nulas:

1. $\mu \text{ CCEI | pre-test} = \mu \text{ CCEI | pos-test}$
2. $\mu \text{ CCEI IS | pre-test} = \mu \text{ CCEI IS | pos-test}$

HIPÓTESIS DIRECCIONAL

1. $\mu \text{ CCEI | pre-test} < \mu \text{ CCEI | pos-test}$
2. $\mu \text{ CCEI IS | pre-test} < \mu \text{ CCEI IS | pos-test}$
3. El coeficiente de confiabilidad del ECOE con el estudio de generalizabilidad es > 0.60

μ = media

CCEI = Competencia clínica estudiantes de Internado

IS = Instituciones de salud

VI. Metodología

1. Tipo de estudio

De cohorte prospectivo⁶³. Pre-test y pos-test de un solo grupo⁶⁴.

2. Población

Estudiantes de Internado médico de pregrado

Criterios de inclusión:

Estudiantes que se encuentren inscritos en el Internado Médico en el ciclo 2010, de instituciones del sector salud en la zona metropolitana.

Criterios de exclusión

- Estudiantes que no aceptaron participar.

3. Tamaño de la muestra:

La población total de estudiantes de Internado médico, de la Facultad de Medicina de la UNAM, en toda la República Mexicana era de 903, de los cuales en el Distrito Federal eran 708 y a partir de este número se calculó el tamaño mínimo de la muestra aleatoria, que fuera representativa:

$$n = Z^2 \alpha \frac{N \cdot p \cdot q}{i^2(N - 1) + Z^2 \alpha \cdot p \cdot q}$$

n Tamaño muestral

N Tamaño de la población, número de sujetos.

Z Valor correspondiente a la distribución de Gauss 1.96 para $\alpha = 0.05$ y 2.58 para $\alpha = 0.01$.

P Prevalencia esperada del parámetro a evaluar. En caso de desconocerse, aplicar la opción más desfavorable ($p=0.5$), que hace mayor el tamaño muestral.

Q $1-p$ (Si $p=30\%$, $q=70\%$)

I Error que se prevé cometer. Por ejemplo, para un error del 10%, introduciremos en la fórmula el valor 0.1. Así, con un error del 10%, si el parámetro estimado resulta del 80%, tendríamos una seguridad del 95% (para $\alpha = 0.05$) de que el parámetro real se sitúa entre el 70% y el 90%. Vemos, por tanto, que la amplitud total del intervalo es el doble del error que introducimos en la fórmula.

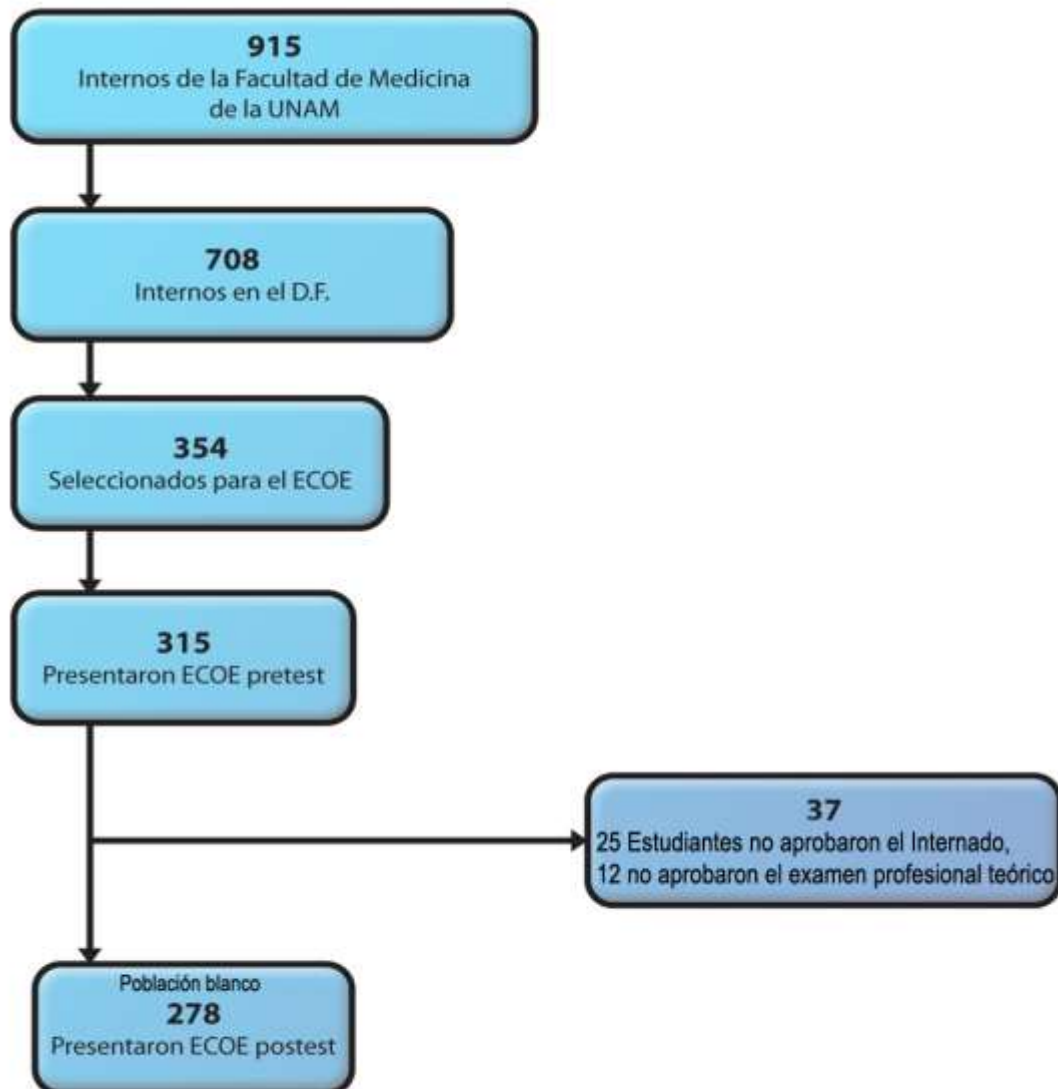
$$n = 1.96^2 \frac{708 \cdot 0.5 \cdot 1 - 0.5}{0.1^2(708 - 1) + 1.96^2 \cdot 0.5 \cdot 1 - 0.5}$$

$$n = 3.8416 \frac{708 \cdot 0.25}{7.07 + 3.8416 \cdot 0.25}$$

$$n = 3.8416 \frac{177}{7.07 + 0.9604}$$

$$n = 3.8416 \times 22.04 = 84.68.$$

La muestra mínima era de 85 estudiantes.



Sin embargo, para que estuvieran representadas todas las sedes hospitalarias de Internado, los estudiantes y las instituciones, se obtuvo una muestra estratificada del 50% del total de estudiantes de las sedes hospitalarias del Distrito Federal, lo que nos dio un número de 354 estudiantes. Para este fin se utilizó el programa SPSS versión 15.

En este estudio se utilizó el diseño de cohortes ya que es el más factible y pertinente para conocer el impacto del Internado sobre el desarrollo de la competencia clínica. El diseño prospectivo de cohortes es una estrategia potente para definir la incidencia e investigar las posibles causas de un efecto determinado. Como la medición de los posibles factores causales se efectúa antes de que se produzca el desenlace, un estudio de cohortes puede establecer que dichos factores precedieron al desenlace. Esta secuencia temporal refuerza la inferencia de que el factor pueda ser la causa del resultado⁶³. Este diseño tiene limitaciones en la atribución única de una mayor competencia clínica como efecto del Internado médico, ya que el tiempo entre el pre y el pos-test permite que haya maduración de los estudiantes. Sin embargo, se utilizaron varias estrategias para reducir el impacto de las variables confusoras: 1) se obtuvo una muestra aleatoria estratificada de los estudiantes, 2) se aplicó un examen equivalente en el pos-test para disminuir el efecto de realización de pruebas⁶⁴. Otra variable de confusión fue el de la instrumentación, entre el pre-test y el pos-test los evaluadores pueden cambiar, ser más hábiles o más indiferentes, por lo que los evaluadores son capacitados y previo al examen se hizo hincapié en las conductas esperadas. Cabe señalar que, no era factible tener un grupo control al cual se le aplicara el pre-test, que no fueran expuestos a la variable independiente, el Internado médico, y al final aplicarles el pos-test.

Clasificación Metodológica de Variables

Variables independientes: Las sedes clínicas de las instituciones de salud

Promedio de los 4 años previos

Puntuaciones de los exámenes departamentales por áreas troncales del Internado

Promedio global del Internado médico (exámenes departamentales)

Puntuación del Examen profesional fase teórica

Variable dependiente: La competencia clínica de los estudiantes

Clasificación Estadística y Operacionalización de Variables

Variable	Tipo	Escala / Modalidad
Institución	Cualitativa nominal	IMSS ISSSTE Privados SSGDF SSA
Hospitales	Cualitativa nominal	IMSS HGZ con UMF No. 26 Las Américas HGZ con UMF No. 8, Dr. Gilberto Flores Izquierdo HGZ No. 1-A Venados HGZ No. 27 Dr. Alfredo Badallo García HGZ No. 2-A Francisco Del Paso Y Troncoso HGZ No. 30 Iztacalco HGZ No. 32 Villa Coapa HGZ No. 47 Vicente Guerrero HGR No.1 Dr. Carlos Mcgregor Sánchez

		<p style="text-align: center;">ISSSTE</p> <p>Hospital General "Dr. Darío Fernandez Fierro"</p> <p>"Hospital General "Dr. Fernando Quiroz Gutiérrez</p> <p>Hospital Regional "Lic. Adolfo López Mateos"</p> <p style="text-align: center;">Privados</p> <p>Hospital Angeles de Las Lomas</p> <p>Hospital de Jesús</p> <p>Hospital Inglés ABC</p> <p>Hospital Médica Sur</p> <p style="text-align: center;">SS GDF</p> <p>Hospital General Balbuena</p> <p>Hospital General Iztapalapa</p> <p>Hospital General La Villa</p> <p>Hospital General Xoco</p> <p style="text-align: center;">SSA</p> <p>Hospital General "Dr. Manuel Gea González"</p> <p>Hospital General de México</p> <p>Hospital Juárez de México</p> <p>Instituto Nacional de Ciencias Médicas y de la Nutrición "Salvador Zubirán"</p>
--	--	---

IMSS Instituto Mexicano del Seguro Social

ISSSTE Instituto de Seguridad y Servicios Sociales para los Trabajadores del Estado

SSGDF Secretaría de Salud del Gobierno del Distrito Federal

SSA Secretaría de Salubridad y Asistencia

Privados

Variable	Tipo	Escala / Modalidad
Exámenes departamentales por áreas troncales (puntuación)	Cuantitativa continua	0-100

Variable	Tipo	Escala / Modalidad
Examen profesional, fase Teórica (puntuación)*	Cuantitativa continua	0-100

Variable	Tipo	Escala / Modalidad
Competencia clínica (puntuación del pre-test y pos-test)	Cuantitativa continua	0-100

* El examen profesional teórico estuvo constituido por 174 casos clínicos con 330 preguntas de opción múltiple, el 27.7% fueron de ciencias básicas y sociomédicas y 72.73% de las seis áreas clínicas troncales.

4. Instrumentos de recolección de información

El instrumento de evaluación fue un ECOE de 18 estaciones, que se conformaron y validaron por un comité de expertos profesores de la Facultad de Medicina, con al menos 10 años de experiencia clínica, pertenecientes a una especialidad troncal

(Cirugía, Ginecoobstetricia, Medicina Familiar, Medicina Interna, Pediatría y Urgencias), que habían acreditado un curso-taller sobre el ECOE y participado en al menos tres exámenes con esta metodología. Un desempeño superior en medicina está estrechamente relacionado con el compromiso en la práctica y con la realimentación durante el entrenamiento médico. Se considera que los médicos especialistas, mejoran su desempeño para alcanzar los niveles más altos a los 10 años de práctica clínica deliberada y se mantienen en ese nivel con la práctica continuada⁶⁵. Estos 18 profesores elaboraron las estaciones, que fueron validadas además por un grupo de 6 médicos familiares en las siguientes etapas:

Etapas en la conformación de los instrumentos:

Primera etapa: Se seleccionaron las habilidades a evaluar tomando como base el programa académico de Internado Médico del Plan de Estudios de la Facultad de Medicina, obteniéndose el perfil de los problemas más frecuentes a los que se enfrenta el médico general.

Segunda etapa: Una vez que se tuvo este perfil, los miembros del Comité procedieron a determinar el número de estaciones, de acuerdo a la habilidad clínica que se consideró más representativa en cada enfermedad y que además fuera posible evaluar en los seis minutos de cada estación. Con base en la tabla de especificaciones y los componentes de la competencia clínica a evaluar, se elaboraron las siguientes estaciones: diez de interrogatorio, tres de exploración física, una de diagnóstico y manejo, dos de interpretación radiográfica, una de interpretación de estudios de laboratorio y una de análisis de un trabajo de investigación.

Tercera: Se elaboró el material de apoyo para cada estación: resúmenes clínicos de casos reales, material para el alumno, listas de cotejo (Ver Anexos 1 y 2) y estudios de laboratorio. Para asegurar la confiabilidad interevaluador y registrar el desempeño clínico relevante en cada una de las estaciones, las listas de cotejo utilizadas fueron elaboradas por médicos especialistas en cada una de las áreas, mismas que fueron revisadas por el Comité y validadas por otro grupo de médicos familiares con el fin de asegurar el enfoque del médico general.

Cuarta: Se seleccionaron pacientes reales y personas no enfermas que se capacitaron para que representaran un problema médico en forma consistente como pacientes estandarizados, siguiendo un libreto tomado del caso clínico real (Ver Anexo 3). Los pacientes seleccionados participaron en un curso-taller de capacitación cuyo propósito fue prepararlos para representar un padecimiento determinado y con un nivel de consistencia que nos permitiera diferenciar un buen o mal desempeño de los estudiantes. Al final del curso, participaron en un examen y un médico examinador calificó su desempeño e indicó si la persona era apta como paciente estandarizado. Posteriormente tuvo que participar en un mínimo de 2 exámenes más para poder participar como paciente en estos dos exámenes. Se conformaron estaciones de dos tipos:

- a) Estaciones de procedimientos donde se evaluó la capacidad del estudiante para interrogar o examinar a pacientes reales o estandarizados, formular diagnósticos presuncionales y en algunos casos establecer el plan de manejo inicial.
- b) Estaciones de interpretación de exámenes de laboratorio y/o gabinete en donde se evaluó además la integración diagnóstica y la capacidad para instalar un plan de manejo integral.

Esta metodología es utilizada en la elaboración de cada estación, lo que asegura que se evalúen las competencias que debe poseer un médico general de acuerdo al perfil del egresado de la Facultad de Medicina.

Con base en la tabla de especificaciones se elaboraron las estaciones de las seis áreas de la medicina que sirvió de matriz para estructurarlas (Ver anexos 4 y 5). Por ejemplo, se diseñaron varias estaciones con contenido similar (depresión) pero con diferentes libretos, (depresión posparto en una mujer joven, depresión en un adulto mayor, entre otras). Se solicitó a los autores de estaciones su diseño con base en casos reales para elaborar el resumen con todos los datos que serían utilizados en la estación, tales como el formato de instrucciones del estudiante, la lista de cotejo, el libreto del paciente y se fue estructurando de tal manera que

existiera congruencia entre las tres partes elaboradas con el fin de darle la validez ⁶⁶.

En relación con las características generales de los casos utilizados, contienen una presentación con un escenario clínico y un formato con instrucciones para que el estudiante se enfoque al problema y siga las indicaciones; las listas de cotejo contienen enunciados relacionados con la actividad que tiene que realizar, ya sea interrogatorio, exploración física, dar consejería, integrar uno o varios diagnósticos, proponer indicaciones de manejo, interpretar las alteraciones de estudios de laboratorio y de gabinete y analizar artículos de investigación. En las estaciones con pacientes, el examinador, además de calificar en la lista de cotejo, emite una calificación global sobre las habilidades de comunicación interpersonal con una escala global de 1 a 9, cuyas equivalencias son: 1 a 3=Insatisfactorio, 4 a 6=Satisfactorio y 7 a 9=Superior⁹.

4.1 Número de enunciados

El número de enunciados en las estaciones de procedimientos con pacientes varía entre 6 y 28.

Un alto porcentaje en el pre-test y el pos-test tienen hasta 20 enunciados.

En el pre-test: 61% tienen menos de 20 enunciados
39% de 21 a 28 enunciados
En el pos-test: 74% tienen menos de 20 enunciados
18% de 22 a 28 enunciados
8% de 31 a 32 enunciados

Las actividades clínicas que tienen que realizar los estudiantes en 6 minutos son alcanzables y el número de enunciados están dentro de lo razonable para lograrlo.

En la hoja de lector óptico (Ver Anexo 6) se aplicó el estándar de pase para todas y cada una de las estaciones. Hay varios métodos para calcular el estándar de pase, dentro de los cuales están el de Angoff, Ebel, Hofstee, sin embargo, para este estudio se seleccionó el de *Grupos en el Límite*⁶⁷. En las estaciones con pacientes, el examinador, además de calificar en la lista de cotejo, otorga dos

calificaciones globales, una es el estándar de pase y la segunda corresponde a las habilidades de comunicación interpersonal. El estándar de pase tiene tres posibilidades: Aprobado, No aprobado o Suficiente para aprobar^{67,30}. El estándar de pase se obtiene al promediar la calificación de todos los estudiantes que fueron juzgados “en el límite” para aprobar. Éste nos permite conocer el nivel a partir del cual los estudiantes aprueban esa estación. Este método fue seleccionado debido a su validez y confiabilidad además de la facilidad en su uso por parte de los evaluadores^{67,35}.

La segunda calificación global es la de habilidades de comunicación interpersonal con una escala de 1 a 9 correspondiendo del 1 al 3 a Insatisfactorio, de 4 a 6, Satisfactorio y de 7 a 9, Superior. En las categorías de Insatisfactorio y Superior existen indicadores para guiar al profesor, quien ha recibido la capacitación y de acuerdo a su experiencia clínica y como evaluador marcó su calificación en la hoja de lector óptico⁶⁷.

4.2 Lineamientos para calificar los enunciados de las listas de cotejo

- Cada enunciado fue calificado como 0 o 1
- Un enunciado realizado recibió un punto.
- Un enunciado no realizado o realizado en forma incorrecta recibió 0 puntos.
- La calificación de una estación se obtuvo al sumar el resultado de dos o tres componentes que la constituían. En el caso de las estaciones de procedimientos existen tres componentes: 1) interacción con pacientes (interrogatorio, exploración física o manejo) 2) formulación de diagnósticos y 3) habilidades de comunicación interpersonal. En las estaciones de interpretación de estudios de laboratorio o gabinete, existen dos componentes: 1) datos normales y anormales que se interpretan y 2) formulación de un diagnóstico.
- La calificación final del examen resultó del promedio de puntajes de todas las estaciones, la escala fue de 0 a 100 por ciento de aciertos.

4.3 Formato de los enunciados de las listas de cotejo

Cada estación tenía un formato con los enunciados de las listas de cotejo que se emparejaba con la hoja de lector óptico para que el profesor marcara con lápiz si el examinado realizó lo especificado en el enunciado o lo dejara en blanco si no lo realizaba o lo hacía inadecuadamente.

4.4 Realimentación a los estudiantes e Instituciones

A los estudiantes se les dio realimentación personalizada y por escrito acerca de los resultados del pre-test mediante la Secretaría de Enseñanza Clínica, quien les hizo llegar a través de los tutores coordinadores de cada sede hospitalaria.

En relación al pos-test, que es el Examen Profesional de altas consecuencias, en caso de solicitarlo el sustentante, se les dio la realimentación de manera verbal. Los resultados se les entregaron a los profesores de las instituciones de salud a través de la Secretaría de Enseñanza Clínica para fines de realimentación en forma personal y confidencial a los estudiantes.

4.5 Fuentes de recolección de la información:

Se hizo de la observación directa a los estudiantes que iniciaban el Internado médico y sustentantes del examen profesional al término del quinto año.

Para ello se utilizaron las hojas de lector óptico mencionadas anteriormente en el apartado 4.1 y los evaluadores que participaron en el ECOE fueron seleccionados por haber cursado el taller de capacitación con el propósito de que conocieran los fundamentos del ECOE, la forma de elaborar las estaciones, la validación y la ponderación así como todas las responsabilidades de ser un examinador en un examen de tipo formativo o profesional. El curso-taller, obligatorio para los profesores del examen profesional, incluyó asistir a un examen en calidad de observadores a través de todas las estaciones del mismo y realizar como trabajo final una estación para el examen, basada en un caso real, con el objetivo de profundizar en el conocimiento de éste. Posteriormente de acuerdo a su participación en el mismo curso y la elaboración de su estación, se les invitó a participar como examinador (supervisado) en tres ECOE, para finalmente ser

evaluador en el examen profesional. Estas medidas de control contribuyeron a mejorar la confiabilidad interevaluador³⁴.

Las ganancias de la tecnología en el uso de hojas de lector óptico son:

Ahorro en tiempo por la rapidez en el procesamiento, exactitud del mismo y se posibilita, además la entrega oportuna y realimentación de los estudiantes de forma confiable, precisa y rápida. Se ahorró papel y otros insumos. Es factible a través de esta tecnología, resguardar los resultados en bases y bancos de datos electrónicos, con la posibilidad de administrar en forma cómoda y exacta el procesamiento estadístico de esta información.

4.6 Plan de análisis de resultados

Se realizaron análisis estadísticos con medidas de tendencia central y de dispersión.

-Prueba “t” de Student para dos muestras correlacionadas, entre los resultados del pre-test y el pos-test.

-Prueba ANOVA para 3 o más muestras.

-Se calculó la *d* de Cohen como medida del tamaño del efecto en la puntuación global del ECOE, usando la siguiente fórmula⁶⁸.

$$d \text{ de Cohen} = M_1 - M_2 / \sigma_{\text{combinada}}$$

Donde: M = media de cada grupo

σ = desviación estándar

$$\sigma_{\text{combinada}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$$

-Prueba ANOVA anidada entre Instituciones y hospitales.

-Prueba ANOVA entre las 6 áreas de Internado.

-Coeficiente de correlación de Pearson:

-Análisis de regresión lineal.

-Análisis de regresión múltiple.

-Pruebas de confiabilidad del examen mediante el Alfa de Cronbach y

-La Teoría de la Generalizabilidad

Se realizó el estudio de generalizabilidad con los resultados de la rutina “model fit” de JMP del análisis de varianza, versión 11, que es adecuado y además cuenta con la herramienta MVR (Máxima verosimilitud restringida). El objeto de la medición, también llamado faceta de diferenciación se le etiquetó como “p” que significa “persona”. Se empleó un modelo de efectos aleatorios para identificar las principales fuentes de variación. Con este diseño se obtuvieron los estimados de los componentes de varianza para cada faceta, éstos son: los de las estaciones, las sedes, las versiones y todas sus interacciones, se incluyó la interacción estación x estudiantes anidados en versión y sede. Además se estimó el componente de varianza del residuo que es la variabilidad no explicada por ninguna de las facetas.

VII. Organización general del estudio

Se realizaron simultáneamente los exámenes en seis unidades de primer nivel de atención médica. Las Clínicas que participaron fueron: la CMF Dr. Ignacio Chávez y CMF Narvarte del ISSSTE; las UMF 7, 10, 21 y 28 del IMSS.

Para evaluar el desarrollo de la competencia clínica de los estudiantes de Internado Médico, se utilizó el Examen Clínico Objetivo Estructurado con 18 estaciones, 2 a 3 de cada área de conocimiento del Programa de Internado: Cirugía, Ginecoobstetricia, Medicina Familiar, Medicina Interna, Pediatría y Urgencias, debido a que son integradoras.

Cada examen tuvo una duración de dos horas y cada una de las estaciones fue de 6 minutos incluyendo dos estaciones de descanso e incluyeron diez estaciones de interrogatorio, tres de exploración física, una de diagnóstico y manejo, una de interpretación radiográfica, una de interpretación de estudios de laboratorio, una de interpretación de un electrocardiograma y una de análisis de un trabajo de investigación.

En cada estación los casos clínicos contenían una presentación en un escenario clínico y un formato con instrucciones para que el estudiante realizara únicamente lo que se le solicitaba. Por otra parte, los evaluadores poseían las listas de cotejo que utilizaron con los enunciados relacionados con la actividad que tenía que realizar el estudiante, ya sea interrogatorio, exploración física, dar consejería, formular diagnósticos, plantear indicaciones de manejo, anotar las alteraciones de estudios de laboratorio, de los estudios radiográficos, del electrocardiograma y los datos relevantes del trabajo de investigación.

En relación con los dos ECOEs, el pre-test se realizó en marzo de 2010 y el post-test en enero de 2011.

En el pre-test se evaluaron a 315 estudiantes de Internado médico. Los estudiantes se dividieron en 3 subgrupos de 20 cada uno, sesenta por sede, por lo que fueron evaluados 315 estudiantes en un día, en seis sedes, en forma

simultánea. Éste fue un examen de tipo formativo y voluntario en el que se les dio realimentación personalizada de sus resultados. Por ser formativo, el desempeño de los estudiantes se esperaba menor al deseable ya que no tenía un impacto en la calificación del Internado ni en la continuación de sus estudios. Antes de iniciar el examen en cada sede se les dieron instrucciones y se les solicitó realizar su mejor desempeño.

El pos-test se aplicó en dos días en enero de 2011 y los estudiantes estuvieron distribuidos aleatoriamente en las cuatro versiones del examen profesional. Éste fue un examen sumativo y obligatorio de altas consecuencias porque para realizar el servicio social y obtener el título de licenciatura es imprescindible aprobarlo. Por ser sumativo se esperaba que el desempeño fuera mayor por la trascendencia para cada uno de los estudiantes.

Para evitar el efecto de pruebas se puso especial cuidado en la utilización de estaciones equivalentes diferentes a las utilizadas en el pre-test y hubo periodos de descanso adecuados entre cada turno, con rotación de personal en donde fue posible.

El pos-test fue parte del examen profesional y requirió cuatro versiones de examen para que los estudiantes tuvieran las mismas oportunidades y no hubiera ventaja de unos sobre otros, lo que planteó una dificultad para tener una equivalencia completa de estación a estación. Para la equivalencia se trató de que fueran casos parecidos de las 6 áreas troncales del programa del Internado y se buscó que las estaciones tuvieran el mismo grado de dificultad, con base en los resultados de los procesos de validación. Se seleccionaron los casos que tuvieron un pronóstico cercano al 60% en promedio, lo que nos indica que el nivel de dificultad general fue similar.

VIII. Aspectos éticos.

Se realizaron reuniones informativas con los estudiantes, a quienes se les invitó a participar voluntariamente y se mantuvo la confidencialidad de sus resultados. El protocolo de este estudio fue aprobado por el Comité de Investigación y Ética de la Facultad de Medicina (Número 124-2009) y recibió financiamiento del Programa

de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza (PAPIME) de la UNAM, con el número de proyecto PE207410.

Debido a que no se publicaron los nombres de los estudiantes ni sus resultados y fueron 354 estudiantes que participaron voluntariamente se consideró que al darles la información y aceptarla no era necesario que firmaran una carta de consentimiento por escrito.

Se publicaron en la página de la Secretaría de Educación Médica los mapas de la ubicación de las seis sedes participantes

IX. Recursos

1. Humanos.

Se contó con:

120 profesores evaluadores

122 pacientes estandarizados

Personal capacitado en informática

2. Materiales

Impresión de hojas de respuesta para lector óptico 13,000 Ver anexo 3

Dos computadoras

Dos impresoras láser.

Dos memorias USB.

Hojas blancas para manuales de cursos 20,000

Acetatos. 100

Plumines 20

Reproducción del material de apoyo y bibliográfico para los cursos.

Lector óptico de marcas OMR insigth 4 OMR/22

3. Financieros

El proyecto requirió recursos financieros para pago de profesores, pacientes estandarizados, personas de apoyo, servicios de transporte, lunch y materiales. Esto dio un promedio de costo por alumno de \$ 700.00.

La evaluación a 354 alumnos del pre-test tuvo un costo total aproximado de \$ 200,000.00. (Ver Anexo 7)

Esta investigación se inscribió dentro de un proyecto PAPIME que se presentó ante la Comisión de Investigación de la Facultad de Medicina quien lo aprobó con el No. 124-2009. Sin embargo, estos recursos, \$200,000.00 pesos, no llegaron a tiempo lo que hizo que el pre-test se aplicara en marzo de tal manera que ya tenían 2 meses y medio de avanzado el Internado.

Los sustentantes pagaron la cantidad de \$2,700.00 por concepto de examen profesional, teórico y práctico, de tal manera que los recursos financieros se obtuvieron de esta partida para el pos-test que se aplicó dentro de la fase práctica del Examen Profesional con el Examen Clínico Objetivo Estructurado.

Resultados

Se aplicó el ECOE para evaluar la competencia clínica con 18 estaciones a 278 estudiantes del Internado Médico en marzo de 2010, lo que constituyó el pre-test; y se realizó un ECOE equivalente en enero de 2011, siendo éste el pos-test. Este último se estructuró con el mismo formato que el pre-test, fue elaborado por el mismo grupo de profesores, considerando la tabla de especificaciones y los componentes similares de la competencia clínica. La equivalencia se confirmó por consenso de expertos y por el comportamiento de las estaciones. En la aplicación del examen participaron 108 profesores evaluadores.

La media global del ECOE pre-test al principio del Internado fue de 55.6 ± 6.6 y la media de la medición pos-test al final del Internado fue de 63.2 ± 5.7 , con una diferencia absoluta de 7.6% ($p < 0.001$) (Figura 1). El índice “*d*” de Cohen⁶⁸, como medición del tamaño del efecto en la puntuación global del ECOE, fue de 1.2

En la figura 1 se observa que el área de Medicina Familiar tuvo la mayor diferencia de medias pre-post, que fue de 12.1, seguida por el área de Cirugía. En contraste, el área de menor diferencia fue la de Medicina Interna, aunque todas las diferencias fueron significativas.

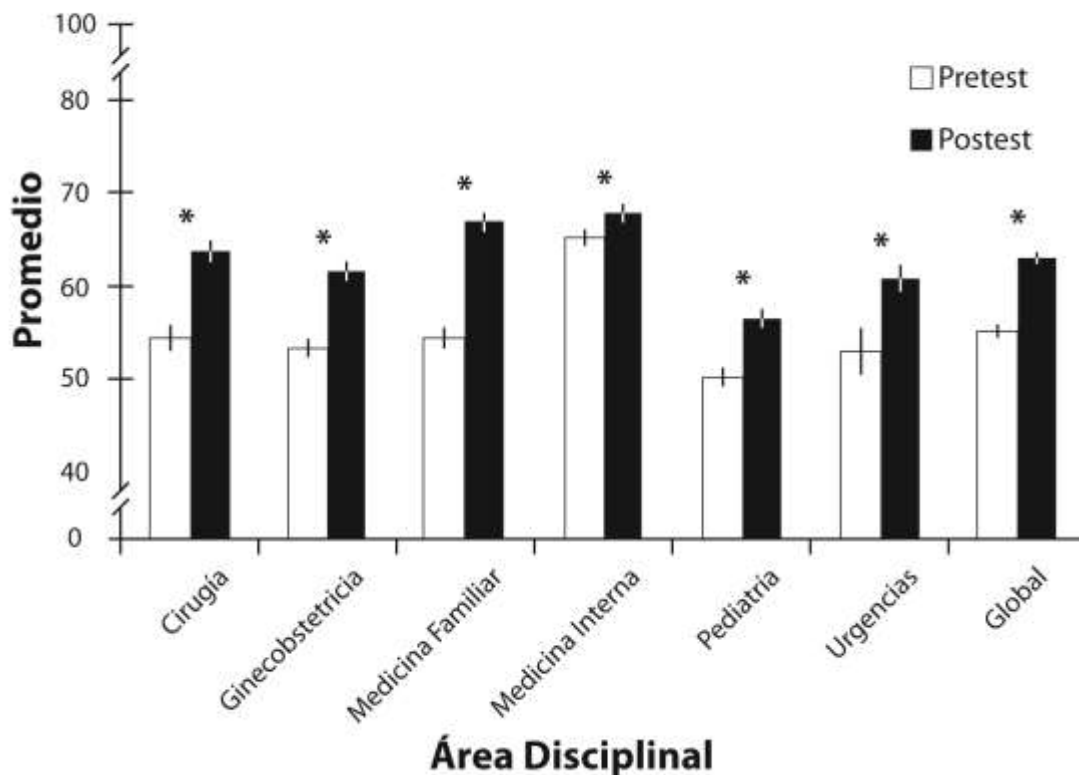


Figura 1. Resultados obtenidos por los estudiantes en el ECOE (n=278); las abscisas indican el área disciplinal y el eje de las “Y” el promedio de las puntuaciones. Las barras representan el promedio obtenido en cada área disciplinal en el pre-test y el pos-test. Las líneas verticales de las barras representan los intervalos de confianza al 95%; la diferencia entre los promedios fue estadísticamente significativa (*=p<0.0001).

El análisis de las puntuaciones de las habilidades clínicas entre el pre-test y el pos-test muestra que en lo general fueron mayores en este último con una diferencia estadísticamente significativa ($p < 0.001$). Al analizar las puntuaciones de las habilidades clínicas entre el pre-test y el pos-test según los componentes de la competencia clínica, se observa que el componente con mayor diferencia en puntuación fue el de interpretación radiológica, seguido de la exploración física. En contraste, el componente con menor diferencia fue el de interpretación de estudios de laboratorio (Figura 2).

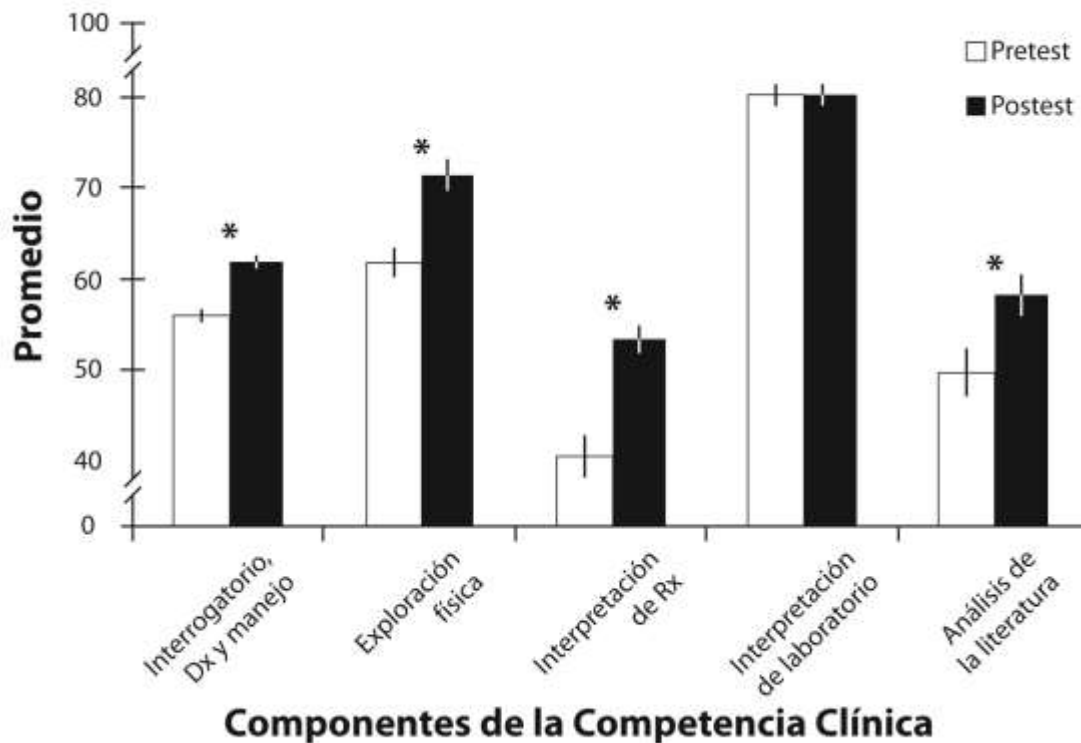


Figura 2. Resultados obtenidos por los estudiantes en el ECOE (n=278); las abscisas indican los componentes de la competencia clínica y el eje de las “Y” el promedio de las puntuaciones. Las barras representan el promedio obtenido en cada componente en el pre-test y el pos-test. Las líneas verticales de las barras representan los intervalos de confianza al 95%; la diferencia entre los promedios fue estadísticamente significativa (*= $p < 0.0001$) con excepción del componente de interpretación de laboratorio.

Para las puntuaciones obtenidas por Institución del Sistema de Salud, se observa que en lo general las diferencias entre el pre-test y pos-test fueron estadísticamente significativas (Tabla 1). La institución que tuvo la mayor diferencia en la puntuación de la competencia clínica, que se refleja en la mayor diferencia de medias, fue el IMSS, en contraste con la menor diferencia de medias que fue la SS GDF. Debe señalarse que la SSA obtuvo tanto en el pre-test como en el pos-test las medias más altas (Instituto Nacional de Ciencias Médicas y de la

Nutrición Salvador Zubirán, Hospital General de México y Hospital Gea González) en comparación con las otras instituciones. Estas diferencias, sin embargo, al comparar las instituciones de salud mediante ANOVA no fueron significativas.

Tabla 1. Resultados obtenidos por los estudiantes en el ECOE (n=278) según Institución de Salud de procedencia. Se observa el promedio y la Desviación Estándar (DE) en el pre-test y pos-test, y en todas ellas la diferencia fue significativa (p<0.0001).

Institución de Salud	N	Media y DE Pre-test	Media y DE Pos-test	Diferencia de medias	p
SSA	60	57.4 ± 11.5	64.5 ± 6	7.1	<0.0001
ISSSTE	18	54.5 ± 5.8	62 ± 6.2	7.5	<0.0001
IMSS	158	55.3 ± 6.6	63.1 ± 5.9	7.8	<0.0001
SS GDF	31	55.6 ± 5.4	62.1 ± 5.7	6.5	<0.0001
Promedio global	278*	55.6 ± 6.6	63.2 ± 5.7	7.6	<0.0001

***= 11 estudiantes pertenecían a diversas instituciones particulares que no entraron en esta comparación por el tamaño de muestra pequeño.**

SSA= Secretaría de Salud; ISSSTE= Instituto de Seguridad y Servicios Sociales para los Trabajadores del Estado; IMSS= Instituto Mexicano del Seguro Social; SSGDF= Secretaría de Salud del Gobierno del Distrito Federal.

Resultados de la Correlación de Pearson

Se realizó la correlación de Pearson entre los ECOEs pre-test y pos-test con: los exámenes departamentales por áreas durante el Internado Médico, con el promedio de los exámenes de Internado, con el promedio de los 4 años previos al Internado y con el examen profesional teórico.

Hubo correlaciones bajas entre el ECOE pre-test global con los exámenes departamentales por áreas durante el Internado. **Tabla 2**

Cuadro 2. Correlación de Pearson entre el ECOE pre-test global con los exámenes Departamentales (ED) por áreas durante el Internado.

ECOEs pre-test: Cirugía (ED)	ECOEs pre-test: Ginecología (ED)	ECOEs pre-test: Pediatria (ED)	ECOEs pre-test: Medicina Interna (ED)	ECOEs pre-test: Medicina Familiar (ED)	ECOEs pre-test: Urgencias (ED)
r= 0.15 p= 0.01	r= 0.15 p= 0.01	r= 0.18 p= 0.002	r= 0.31 p= <0.0001	r= 0.11 p= 0.06	r= 0.22 p= 0.0001

* < .35 ligera correlación

.36 - .64 moderada correlación

> .65 alta correlación

Las correlaciones fueron bajas entre los exámenes departamentales por áreas del Internado con el ECOE pos-test por áreas. **Tabla 3**

Cuadro 3. Correlación de Pearson entre los exámenes departamentales (ED) por áreas durante el Internado con el ECOE pos-test por áreas.

Cirugía (ED): ECOEs Cirugía	Ginecología (ED): ECOEs Ginecología	Pediatria (ED): ECOEs Pediatría	Medicina Interna (ED): ECOEs Medicina Interna	Medicina Familiar (ED): ECOEs Medicina Familiar	Urgencias (ED): ECOEs Urgencias:
r=0.10 p= 0.07	r=0.10 p= 0.07	r= -0.08 p= 0.15	r=0.12 p= 0.03	r=0.12 p= 0.03	r=0.21 p= 0.0004

Hubo correlaciones moderadas entre el ECOE pos-test con el promedio por áreas del mismo ECOE pos-test: Cirugía 0.67, Ginecobstetricia 0.55, Pediatría 0.53, Medicina Interna 0.52, Medicina Familiar 0.50, Urgencias 0.49, todas fueron significativas (<.0001).

Las correlaciones fueron bajas entre el ECOE pos-test con las puntuaciones de los exámenes departamentales de cada una de las seis áreas troncales del Internado. **Tabla 4.**

Cuadro 4. Correlación de Pearson entre el ECOE pos-test global y los exámenes departamentales (ED) por área durante el Internado.

Cirugía (ED): ECOE pos-test	Ginecobstetricia (ED): ECOE pos-test	Pediatría (ED): ECOE pos-test	Medicina Interna (ED): ECOE pos-test:	Medicina Familiar (ED): ECOE pos-test	Urgencias (ED): ECOE pos-test
r=0.17 p= 0.003	r=0.11 p= 0.05	r=0.11 p= 0.05	r=0.23 p= 0.0001	r=0.08 p= 0.14	r=0.21 p= 0.0004

Hubo correlaciones bajas (tendientes a moderadas) entre el examen profesional teórico global con el ECOE pos-test por áreas. **Tabla 5.**

Cuadro 5. Correlación de Pearson entre el examen profesional teórico global con el ECOE pos-test por áreas.

Examen profesional teórico: Cirugía ECOE	Examen profesional teórico: Ginecobstetricia ECOE	Examen profesional teórico: Pediatría ECOE	Examen profesional teórico: Medicina Interna ECOE	Examen profesional teórico: Medicina Familiar ECOE	Examen profesional teórico: Urgencias ECOE
r=0.22 p= 0.0002	r=0.25 p= <.0001	r=0.13 p= <.0001	r=0.32 p= <.0001	r=0.32 p= <.0001	r=0.25 p= <.0001

Se obtuvieron correlaciones moderadas entre el promedio final del Internado con las puntuaciones de los exámenes teóricos de cada una de las 6 áreas: Pediatría

0.62, Ginecobstetricia 0.53, Urgencias 0.53, Medicina Interna 0.52, Medicina Familiar 0.50 y Cirugía 0.50, todas fueron significativas (<.0001).

Se obtuvieron moderadas correlaciones pero significativas entre el ECOE del Examen profesional (pos-test) con la fase teórica del examen profesional ($r=0.44$, $p<.0001$), con el promedio del ECOE pre-test de ($r=0.44$, $p<.0001$) y con el promedio de los 4 años previos al Internado ($r=0.39$, $p<.0001$) y baja correlación con el promedio de los exámenes del Internado ($r=0.25$, $p<.0001$). **Tabla 6**

Al realizar el análisis de regresión lineal para valorar el nivel predictivo de su desempeño en el ECOE pos-test a partir de las variables, los coeficientes de determinación (R^2) fueron bajos aunque todos fueron significativos con el examen profesional teórico, con el ECOE pre-test, con el promedio de los 4 años previos al Internado y con el promedio de los exámenes del Internado. **Tabla 6**

Cuadro 6. Coeficiente de correlación (r) y determinación (R^2) entre el promedio de los 4 años previos, el promedio de los exámenes del Internado y el ECOE pre-test con el ECOE pos-test.

Promedio de los 4 años previos: ECOE pos-test	ECOE pre-test: ECOE pos-test	Promedio de los exámenes del Internado: ECOE pos-test	Examen profesional teórico: ECOE pos-test
$r =0.39$ <.0001	$r =0.44$ <.0001	$r =0.21$ <.0001	$r =0.44$ <.0001
$R^2 =0.15$	$R^2 =0.19$	$R^2 =0.04$	$R^2 =0.20$

El análisis de regresión múltiple de estas mismas variables para el ECOE pos-test R^2 fue de 0.30 que se considera bajo. **Tabla 7**

Cuadro 7. Análisis de regresión múltiple para evaluar la validez predictiva del ECOE pre-test, el promedio de los 4 años previos, los exámenes escritos durante el Internado y el examen profesional teórico con el ECOE pos-test.

<p style="text-align: center;">ECOE pre-test + Promedio de los 4 años previos + Exámenes escritos durante el Internado+ Examen profesional teórico vs ECOE pos-test</p>
<p style="text-align: center;">$R^2=0.30$ $p<0.001$</p>

Resultados de la confiabilidad

La confiabilidad del estudio con el Alfa de Cronbach fue de 0.62 en el pre-test y de 0.64 en el pos-test. Por otro lado, el coeficiente G del pre-test fue de 0.79 y el pos-test fue de 0.93.

En el pre-test se aplicó la misma versión para los 278 estudiantes. En el pos-test la distribución de estudiantes en las versiones del examen fue: 97 en la versión uno, 53 en la dos, 98 en la tres y 30 en la cuatro.

El estudio de generalizabilidad reveló que la mayor fuente de varianza (Cuadro 8) fue el error residual (59.4% y 51%) en el pre y el pos-test respectivamente. Otra fuente de varianza fueron las estaciones en el pre y el pos-test (19.6% y 11.4%). No se encontró fuente importante de varianza en las sedes ni en las versiones en el pos-test (0 y 0.2 %). También se observa que la varianza total del pos-test fue menor que la del pre-test aunque la primera tenía una fuente más de variación que eran las versiones.

También se observa en el cuadro 8 que los porcentajes de los estimados de los componentes de varianza en el pre y el pos-test conservan el mismo orden aunque éstos son menores en el segundo.

Cuadro 8. Estimados de los Componentes de Varianza obtenidos al realizar el análisis del ECOE (n=278) en el pre y el pos-test.

Componentes	PRE-TEST		POS-TEST	
	Estimados componentes de Varianza	Por ciento del total	Estimados componentes de Varianza	Por ciento del total
Estación	87.58	19.6	42.157	11.4
Versión*	**	**	0.737	0.2
Versión x estación	**	**	38.968	10.6
Sede	-3.35**	0	-2.057**	0
Sede x estación	67.43	15.1	46.631	12.7
Sede x estación x versión	**	**	34.692	9.4
Estudiante [versión,sede]	28.83****	6.4****	17.652	4.8
Estudiante x estación [versión,sede]***	264.86****	59.4****	187.374	51
Sede x versión	**	**	0.867	0.2
Total	445.36	100	367.024	100

These estimates based on equating Mean Squares to Expected Value.

[] =Anidado x = Interacción

*En el pre-test no hay estimados de componentes de varianza donde aparece la versión debido a que fue una versión solamente.

** Estos componentes fueron negativos y muy pequeños, por lo que se ajustaron a cero para propósitos del reporte y para calcular el coeficiente G.

*** Corresponde al error residual

****En el pre-test, los estudiantes solo estaban anidados en sede

Para determinar el coeficiente G del pre y el pos-test, se calculó la varianza del error relativo en el que se consideran los componentes de varianza que tienen interacción con las personas o estudiantes y se divide entre las estaciones, las sedes y en el pos-test, además las versiones y los productos de estas interacciones. Se utilizaron las siguientes fórmulas:

$$g = \frac{\sigma p^2}{\sigma p^2 + \frac{\sigma_{pe}^2}{n_e} + \frac{\sigma_{pes}^2}{n_e \times n_s}}$$

$$g = \frac{28.83}{28.83 + [(0/18) + (28.83/18) + (264.86/18 \times 6)]}$$

$$g = \frac{28.83}{28.83 + 1.60 + 2.45}$$

$$g = \frac{28.83}{32.88} = 0.87$$

Coefficiente G= 0.87 Error Estándar de la Medición Relativo = 4.05

$$g = \frac{\sigma p^2}{\sigma p^2 + \frac{\sigma_{pvs}^2}{n_v \times n_s} + \frac{\sigma_{pevs}^2}{n_e \times n_v \times n_s}}$$

$$g = \frac{17.65}{17.65 + [(17.652/ 4 \times 6) + (187.374 / 432)]}$$

$$g = \frac{17.65}{17.65 + 0.735 + 0.433} = \frac{17.65}{18.81} = 0.93$$

Coefficiente G = 0.93 Error Estándar de la Medición Relativo = 1.16

σp^2 = varianza entre estudiantes
 σs = varianza de sedes
 n_s = número de sedes
 σv = varianza de versiones
 n_v = número de versiones
 σe = varianza de estaciones
 n_e = número de estaciones

Para determinar el coeficiente D (Φ) del pre y el pos-test, se calculó la varianza del error absoluto en el que se consideran todos los componentes de varianza con excepción de los que tienen personas o estudiantes y se divide entre las estaciones, las sedes y en el pos-test las versiones y los productos de estas interacciones. Se utilizaron las siguientes fórmulas:

$$\Phi = \frac{\sigma p^2}{\sigma p^2 + \frac{\sigma_s^2}{n_s} + \frac{\sigma_e^2}{n_e} + \frac{\sigma_{es}^2}{n_e \times n_s}}$$

$$\Phi = \frac{28.83}{28.83 + [(0/6) + (87.58/18) + (67.43/18 \times 6)]}$$

$$\Phi = \frac{28.83}{28.83 + 4.86 + 0.624}$$

$$\Phi = \frac{28.83}{34.31} = 0.84 \quad EEM = 4.86 + 0.624 = 5.48$$

Coefficiente D= 0.84 Error Estándar de la Medición Absoluto = 5.48

$$\Phi = \frac{\sigma p^2}{\sigma p^2 + \frac{\sigma_s^2}{n_s} + \frac{\sigma_v^2}{n_v} + \frac{\sigma_{sv}^2}{n_s \times n_v} + \frac{\sigma_e^2}{n_e} + \frac{\sigma_{ev}^2}{n_e \times n_v} + \frac{\sigma_{es}^2}{n_e \times n_s} + \frac{\sigma_{evs}^2}{n_e \times n_v \times n_s}}$$

$$\Phi = \frac{17.65}{17.65 + [(0/6) + (0/4) + (0.867/6 \times 4) + (42.16/18) + (38.96/18 \times 4) + (46.63/18 \times 6) + (34.69/18 \times 4 \times 6)]}$$

$$\Phi = \frac{17.65}{17.65 + [0 + 0 + 0.036 + (42.16/18) + (38.96/72) + (46.63/108) + (34.69/432)]}$$

$$\Phi = \frac{17.65}{17.65 + [0 + 0 + 0.036 + 2.34 + 0.54 + 0.432 + 0.08]} = \frac{17.65}{21.07} = 0.83$$

Coefficiente D= 0.83 Error Estándar de la Medición Absoluto= 3.42

$$Y_{ijklm} = M + S_j + V_i + E_l + (SV)_{ij} + ES_{(ij)} + (EV)_{ij} + (ESV)_{ij} + Pm_{(ij)} + (PE)_{lm(ij)} + E_{(ijlm)}$$

i= sede
j = versión
l= estación
m= estudiante

P= persona (estudiante)
E= Estación
M= Media
S= Sede
V= Versión

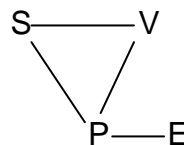


Figura 2. Modelo de análisis estadístico, para obtener los estimados de los componentes de varianza que sirvieron para calcular el coeficiente G, también se representa de forma gráfica.

Los coeficientes G (Generalizabilidad) y D (Dependabilidad) obtenidos en los análisis del pre y pos-test se muestran en el cuadro 9, junto con sus respectivos errores estándar de medición(SEM).

Cuadro 9. Coeficientes G y D del pre y pos-test		
	Pre-test	Pos-test
Coeficiente G	0.87	0.93
SEM relativo	4.05	1.16
Coeficiente D	0.84	0.83
SEM absoluto	5.48	3.42

Discusión

El médico general en nuestro país es una figura indispensable en la prestación de los servicios de salud, ya que atiende la gran mayoría de consultas que se ofrecen diariamente y se estima que resuelve entre el 80 y 90% de los problemas de salud de la población⁶⁹. La evaluación de la competencia clínica de los médicos generales es una prioridad en el contexto nacional e internacional, por lo que es necesario identificar sus áreas de fortaleza, con el fin de consolidarlas y las áreas de oportunidad, con el objeto de aumentar la calidad del proceso educativo en la formación del médico general y lograr el perfil del egresado a cabalidad.

La adquisición y el desarrollo de la competencia clínica en los estudiantes de la Facultad de Medicina de la UNAM se lleva a cabo principalmente en las instituciones del Sistema Nacional de Salud, donde la asistencia médica y la relación médico-paciente desempeñan un papel esencial en la adquisición de los conocimientos, habilidades, destrezas, actitudes y valores que el médico debe poseer. El Internado médico es un año crucial en la adquisición de la competencia clínica del médico general y debe ser evaluado en todos sus componentes entre los que destaca el desarrollo de la competencia clínica en los estudiantes⁷⁰.

En nuestro país son escasos los trabajos que exploran la evaluación de la competencia clínica en estudiantes de medicina, utilizando instrumentos que posean evidencia de validez y confiabilidad^{39,62}. En ese sentido, el presente estudio es el primer trabajo en nuestro medio que, con un diseño de investigación pre y pos-test y un tamaño de la muestra satisfactorio, evaluó la competencia clínica en estudiantes de Internado al inicio y final del mismo. En nuestro estudio se documenta que, después del Internado médico, hay un incremento estadísticamente significativo en la competencia clínica determinada por el ECOE, de acuerdo con los resultados obtenidos por los estudiantes en las dos mediciones realizadas.

En la literatura hay algunos estudios realizados en Internado, y utilizan al ECOE; también identificaron incremento significativo en la competencia clínica en el pos-test^{71,72}. En uno de ellos, con 8 estaciones de 12 minutos, 106 estudiantes, hubo mejoría significativa en las 6 competencias nucleares del Consejo de Acreditación

de Educación Médica de Posgrado de EUA, en 7 de las 8 estaciones, la diferencia de medias de las estaciones en el pre y pos-test fue de 11.5⁷¹. Otro estudio, con 8 estaciones de 10 minutos, encontró una diferencia pre-pos-test de 30.7 (mínima-máxima de 53.4 a 84.1), en 28 estudiantes en la práctica de Medicina General⁷². En estos artículos la dirección y magnitud del cambio en el nivel de las competencias evaluadas con el ECOE son similares a las de nuestro estudio, lo que sugiere que el Internado es una experiencia educativa que incrementa la competencia clínica.

En nuestro estudio el cambio absoluto de competencia clínica medido con el ECOE fue de 7.6%, y al expresarlo en forma de tamaño del efecto con el índice *d* de Cohen fue de 1.2. Esta magnitud de diferencia observada, en términos porcentuales, no parece ser muy elevada; sin embargo el tamaño del efecto se puede interpretar como grande, de acuerdo con la clasificación de Cohen (>0.8)⁶⁸. El uso de índices de tamaño del efecto como la *d* de Cohen en publicaciones de investigación en educación, podría ayudar a visualizar de una manera más estandarizada la magnitud de las diferencias entre los grupos y entre diversos estudios, así como promover la reflexión sobre el potencial significado educativo de los resultados^{73,74}.

El Internado les permite a los estudiantes desarrollar su competencia clínica de forma más integral, debido a varios factores como son: colaborar con sus profesores en el manejo de los pacientes, llevar un programa académico homogéneo con discusión de casos clínicos, las experiencias clínicas repetidas y reflexivas y las actividades clínicas complementarias. Llama la atención que el área disciplinal de Medicina Familiar tuvo el mayor incremento en la puntuación del ECOE, esto puede ser debido a que en lo general, el desempeño docente es guiado por el perfil del egresado que expresa el Plan de Estudios. Este resultado sugiere que los profesores, en el proceso educativo, le dan la importancia debida al perfil del egresado en la Facultad de Medicina y sugieren la conveniencia de ampliar la rotación de los internos en esta área de especialidad, que podría ser un contexto apropiado para la formación del médico general. Los resultados de este estudio, en el pre-test, son semejantes a lo reportado anteriormente en nuestra

Institución en un examen formativo de Medicina Familiar, en el que se encontró una media de 53.9 en las estaciones de interrogatorio y de 62.3 en la exploración física, lo que constituye evidencia de validez agregada a los hallazgos del presente estudio⁷⁵.

Aunque el menor cambio en la puntuación pre-post fue en el área de Medicina Interna, se observa que dicha disciplina fue la más alta en el pre y pos-test, lo que puede deberse a que los años clínicos de 3º y 4º se desarrollan a nivel hospitalario y las asignaturas corresponden en su mayoría a subespecialidades de Medicina Interna^{76,77}. La desviación estándar en general en todas las áreas disciplinares disminuyó en el pos-test, excepto en Medicina Interna, lo que sugiere que al terminar el Internado los estudiantes son una población más homogénea en lo académico y en sus habilidades clínicas. Estos resultados globales son semejantes a los obtenidos en el examen profesional de la Facultad de Medicina con el ECOE, en los años 2002 a 2004, lo que también contribuye a agregar evidencia de validez al estudio⁴⁰.

En el área de Pediatría se observaron los promedios más bajos tanto en el pre-test como en el pos-test. La confiabilidad de cada estación de pediatría tuvo un mayor nivel que el general. Consideramos que estas puntuaciones menores en relación con las otras áreas, se pueden deber a las características propias de este grupo de estudiantes egresados del 2011 y a mayores índices de dificultad.

En relación con los componentes de la competencia clínica, merece especial atención la interpretación radiológica, que obtuvo las puntuaciones más bajas en relación con los otros componentes. Esto sugiere que se tiene que realizar una intervención con el fin de fortalecer este componente clínico. El componente de interpretación de estudios de laboratorio no mostró diferencia en el pre y pos-test, una posible explicación de este hecho es que estas habilidades fueron aprendidas durante el tercer y cuarto año, con un menor impacto en el proceso educativo del Internado en el desarrollo de este componente. Por otra parte, las puntuaciones son altas en comparación con los otros componentes que evalúan otras habilidades.

Las correlaciones de Pearson del ECOE pre-test con los exámenes departamentales en general son bajas aunque significativas ($r=0.11-0.31$, $p<0.05$). La correlación moderada del ECOE pre-test con el examen departamental de medicina interna ($r=0.31$), en parte pueden explicarse porque, esta área obtuvo los mayores puntajes en el ECOE en el pre y el pos-test, tuvo la menor diferencia de medias y es un área que tiene un gran peso curricular desde el tercer año de la carrera. Tabla 2

Las correlaciones entre estos mismos exámenes y el pos-test por áreas fueron bajas pero significativas con tres áreas (Medicina Familiar, Medicina Interna y Urgencias) y no significativas con las otras tres ($r=-0.08-0.10$). En este resultado pueden influir muchos factores: entre estos, la variabilidad de las puntuaciones de los exámenes departamentales porque no son los mismos exámenes bimestrales para todos los estudiantes por motivos de seguridad, los conocimientos y las habilidades clínicas que van adquiriendo los estudiantes son procesos dinámicos y la competencia clínica no se logra de forma homogénea para todos ellos debido a variables individuales.

Las correlaciones bajas-moderadas entre las áreas del ECOE pos-test con el examen profesional teórico ($r=0.13-0.32$) pueden explicarse por la diferencia de aspectos que se evalúan en cada uno de ellos (aspectos parciales versus aspectos globales) Tabla 5

Las moderadas correlaciones y significativas entre la fase teórica y el ECOE del examen profesional ($r=0.44$, $r=0.57$ desatenuada) y con el ECOE pre-test ($r=0.41$, $r=0.54$ desatenuada) apoyan el argumento de validez de las puntuaciones de estos exámenes. En estos casos, como en el de las correlaciones entre el ECOE pre-test con el pos-test ($r=0.44$, $r=0.70$ desatenuada) son moderadas y se explica porque las tres son globales y evalúan la competencia clínica. Las correlaciones fueron de bajas a moderadas entre el promedio de los exámenes del Internado y el ECOE pos-test ($r=0.25$, $r=0.39$ desatenuada) que pueden explicarse por la diferencia de aspectos que se evalúan en cada uno de ellos.

De los estudios realizados en estudiantes de pregrado, Muller encontró correlaciones de bajas a moderadas ($r=0.16-0.38$) entre un examen a gran escala (7,230 estudiantes) de habilidades clínicas (CSE) y el USMLE Step 2 lo que sugirió que el primero evalúa competencias distintas de aquellas evaluadas por los componentes del segundo⁷⁸, similares a nuestras correlaciones de exámenes departamentales y el ECOE pre y pos-test.

Simon encontró que el desempeño en el USMLE Step 1 de los estudiantes de segundo año reportó una modesta correlación con el ECOE ($r=0.41$) aplicado en la escuela de medicina de Harvard. Esta correlación moderada puede interpretarse como reflejo de las similitudes y diferencias entre el ECOE y el USMLE⁷⁹.

Petrusa (1987) en un estudio de 204 estudiantes de pregrado encontró correlación del ECOE con las puntuaciones de sus profesores ($r=0.43$) y con el examen del NBME parte 2 ($r=0.43$) similares a nuestro estudio. Las correlaciones entre las evaluaciones de la escuela de medicina y la puntuación de los supervisores están alrededor de $r=0.10 - 0.30$ y algunas veces incluso son negativas. Gran parte de esta falta de correlación la explican por dificultades metodológicas⁸⁰. Langford encontró en alumnos de quinto año de terapéutica farmacológica correlaciones débiles entre el examen con reactivos de opción múltiple y el ECOE, $r=0.2$ ⁸¹.

Wilkinson encontró correlaciones moderadas y altas, significativas ($r=0.63, 0.44$ y 0.59) entre el ECOE y los exámenes de habilidades clínicas, humanística global y global total. En general existe una pobre relación entre los resultados de las evaluaciones en la escuela de medicina y las mediciones del desempeño después de la graduación. El tradicional caso largo tiene una relación inversa con el desempeño subsecuente⁸².

Efthekar encontró correlaciones de moderadas a altas entre el ECOE y los exámenes que evalúan principalmente conocimientos, juzgados por promedio de calificaciones y un examen nacional de preinternado. El ECOE puede proporcionar información que no es evaluada con los exámenes escritos que evalúan los conocimientos⁸³.

Towsend encontró falta de correlación entre el ECOE pre-test y el ECOE pos-test, pero una correlación moderada y una alta, ambas significativas entre las calificaciones del ECOE pos-test con un examen combinado, teórico y práctico, mediante el ECOE, ($r=0.57$) y con la puntuación total de la práctica general, ($r=0.74$) explicado probablemente por el reconocimiento y enfoque de las deficiencias durante los 10 meses de la rotación⁷².

Smith encontró correlaciones del ECOE en estudiantes de pregrado con el examen final del primer año de la residencia de $r=0.28$ y 0.23 en los rubros de recolección de datos y habilidades de comunicación y no encontró correlación alguna con los exámenes del NBME⁸⁴.

En una revisión sistemática de Hamdy acerca de las correlaciones entre el ECOE y la puntuación de los evaluadores obtuvo un promedio estimado de $r=0.30$. Las correlaciones bajas, a pesar de ser estadísticamente significativas, obtenidas en varios estudios para obtener el valor predictivo de los constructos evaluados con el ECOE como la recolección de datos, las habilidades interpersonales y en la exploración física con las puntuaciones de sus supervisores pueden explicarse sobre la base de que las evaluaciones de los residentes no evalúan objetivamente el mismo constructo que evalúa el ECOE en el programa de pregrado⁸⁵.

Para los estudios realizados en estudiantes de posgrado encontramos que Dupras con 229 residentes de pediatría obtuvo correlaciones bajas y una moderada entre el ECOE con un examen con reactivos de opción múltiple ($r=.22-.34$), con la evaluación del curso ($r=.24-.32$), con el desempeño global ($r=.46-.52$) y con otros exámenes comparables ($r=.28-.36$). Petrusa encontró moderadas y altas correlaciones en posgrado, $r=0.35$ a 0.65 entre el ECOE y las puntuaciones de trabajo clínico diario⁸⁶.

Curtis estimó el valor predictivo de un ECOE basado en conocimientos con un ECOE basado en desempeño y obtuvo un $R^2=0.045$ que se interpreta como muy bajo⁸⁷.

Lawson encontró un valor predictivo de $R^2=0.40$ entre el examen escrito de 2º año con el ECOE de 3er año en el examen del Consejo Canadiense de Medicina Quiropráctica⁸⁸.

Russel encontró en el análisis de regresión lineal simple que la combinación de las puntuaciones del curso de habilidades clínicas y las puntuaciones de los estudiantes del ECOE inicial sugieren una moderada predicción con el desempeño en el ECOE final $R^2 = 0.38^{89}$.

En nuestro estudio el modelo de regresión múltiple para relacionar el pos-test con las cuatro variables: el promedio de los 4 años previos al Internado, el ECOE pre-test, el promedio de los exámenes de Internado y el examen profesional teórico resultó bajo, $R^2 = 0.30$ y en forma separada se observaron tres valores predictivos bajos para el resultado del pos-test, $R^2 = 0.15$, 0.19 y 0.20 y uno muy bajo $R^2 = 0.04$ con el promedio de los exámenes de Internado que resultó una variable muy inestable. Tabla 6

Nuestros resultados sugieren que ni el ECOE ni los exámenes escritos con reactivos de opción múltiple pueden sustituirse unos a otros. Estos exámenes son complementarios y deben combinarse con otros métodos de evaluación que abarquen todos los atributos de la competencia clínica de manera que predigan mejor el desempeño profesional futuro.

El ECOE en este estudio tuvo una adecuada confiabilidad global medida mediante el alfa de Cronbach, de 0.62 y 0.64 en el pre y pos-test, respectivamente, que se acercan a lo establecido por Streiner y Norman de 0.7 para un examen de altas consecuencias⁵¹, sin embargo, fue posible valorar y complementar la confiabilidad del instrumento con la Teoría de la Generalizabilidad.

En las diversas revisiones narrativas y sistemáticas sobre el tema del ECOE también se ha explorado la confiabilidad o consistencia de los juicios realizados por los evaluadores/examinadores. Lo anterior es de gran importancia ya que debido al número de estaciones en un ECOE típico, las estaciones individuales residen en el juicio de un solo examinador. Generalmente esta variable se ha explorado en trabajos de investigación con métodos para medir la confiabilidad interobservador como el coeficiente de correlación intraclase, la estadística kappa o con la teoría de generalizabilidad, y el peso de la evidencia sugiere que la confiabilidad inter-observador en la mayoría de los estudios publicados es

aceptable. En la revisión sistemática más extensa del tema, se documentó que la confiabilidad inter-evaluador en el ECOE va de 0.50 a 0.72, cifras consideradas como aceptables por la comunidad de investigadores en el área, y que han llevado a la decisión de utilizar sólo un evaluador por estación en la mayor parte de las instituciones que utilizan el ECOE como instrumento formativo y/o sumativo^{90,91}. En el caso de la Facultad de Medicina de la UNAM, con base en la información citada se ha decidido utilizar solamente un evaluador por estación. Por otra parte, la participación de dos evaluadores por estación en nuestro medio sería prohibitiva desde los puntos de vista logístico, financiero y de disponibilidad de recursos humanos. La capacitación de los evaluadores y el uso de listas de cotejo contribuyen a disminuir la variabilidad interobservador³⁸.

La Teoría de la Generalizabilidad se aplica en exámenes formativos y sumativos de altas consecuencias y se recomienda para investigar las fuentes de error y el número de observaciones requeridas para un nivel de confiabilidad determinado^{51,92}.

El análisis de las fuentes de error de las mediciones en exámenes sumativos de altas consecuencias es un mecanismo de control de calidad con el fin de asegurar inferencias válidas de los resultados obtenidos. Además se debe considerar que los estudiantes que presentaron el pre-test no tenían experiencia con este tipo de examen, en contraste, en el pos-test ya todos tenían cuando menos la experiencia en el pre-test.

La varianza total del pos-test fue menor que la del pre-test y, aunque la primera tenía una fuente más de variación que fueron las versiones, se puede interpretar como que al final del Internado y debido al proceso educativo similar al que se somete a los estudiantes hay mayor homogeneidad de los mismos.

También se observa en el cuadro 8 que los porcentajes de los estimados de varianza en el pre y el pos-test conservan el mismo orden aunque los componentes son menores en el segundo. Esto se debe a que a los estudiantes tienen un mejor desempeño en las diferentes estaciones y tienen menor variabilidad en el pos-test que en el pre-test.

El estudio de generalizabilidad reveló que la mayor fuente de varianza (Cuadro 8) fue el error residual (59.4% y 51%) en el pre y el pos-test respectivamente lo que indica que una gran proporción de la varianza se debió a la interacción de las estaciones con los estudiantes anidados dentro de las versiones y las sedes.

En relación con los estimados de los componentes de varianza de los estudiantes, 6.4% en el pre-test y 4.8% en el pos-test muestra que tanto difieren uno de otro en su desempeño. El componente de varianza de las estaciones (19.6% y 11.4%) en el pre y pos-test respectivamente, refleja la varianza de los errores constantes asociados con los niveles de dificultad en el universo de las estaciones y la posición relativa de los estudiantes difiere de estación a estación. El menor porcentaje en el pos-test nos habla de una mayor homogeneidad de los estudiantes y un menor grado de variación en la dificultad de las estaciones después de haber realizado el Internado. La variación en la puntuación de las estaciones pudo deberse al contenido de éstas en el ECOE: interrogatorio, exploración física, diagnóstico y manejo, interpretación radiográfica y de estudios de laboratorio y análisis de un trabajo de investigación⁹³.

La varianza en el efecto de la interacción entre los estudiantes y las estaciones indica que hay diferencia en el manejo de los casos por los estudiantes. Por ejemplo, un estudiante podría encontrar algunas de las estaciones fáciles de manejar y otras estaciones difíciles. Estos resultados esperados reflejan la variación en la dificultad de las estaciones.

La interacción entre las versiones y las estaciones (10.6%) en el pos-test indican que existen diferencias en las puntuaciones de las estaciones de acuerdo a la versión. La interacción entre las sedes, versiones y las estaciones del pos-test (9.4%) se explica por las diferencias en las puntuaciones de las estaciones de acuerdo a las sedes y las versiones.

En relación con el cálculo de los coeficientes G y D se tiene la ventaja de un acercamiento basado en la teoría de la generalizabilidad (TG) que permite examinar las implicaciones de aumentar o disminuir el tamaño de la muestra de estaciones para valorar su impacto en el coeficiente G. El estudio de seguimiento es el estudio D, que deriva de los resultados del estudio G, y predice lo que

pasaría si se incrementa el número de estaciones^{54,55}. En nuestro estudio, el incremento a 22 estaciones, con una duración de dos horas 24 minutos incrementaría el coeficiente G del pre-test de 0.87 a 0.89 y el del pos-test sería de 0.93 a 0.94. Las versiones y las sedes tienen un componente de varianza muy cercano a 0% por lo que el aumento de la confiabilidad sería muy marginal en caso de modificar éstas.

Para calcular el coeficiente D (Dependabilidad) se consideraron todas las varianzas con excepción de aquellas donde aparece persona o estudiante y el nivel del coeficiente D fue de 0.84 en el pre-test y de 0.83 en el pos-test^{54,55}.

En el periodo comprendido entre el pre-test y el pos-test los estudiantes desarrollan su competencia clínica y esto hace que la variabilidad en sus componentes se refleje en el valor más pequeño del error estándar de medición (SEM) absoluto y relativo del pos-test que fueron más homogéneos, en comparación con el pre-test. (Cuadro 12).

Con respecto del nivel de confiabilidad y el alfa de Cronbach nuestros resultados fueron de 0.64 y 0.62 en el pre y pos-test respectivamente. Brailovsky obtuvo un alfa de 0.63 a 0.80 en un examen de 5.5 hs⁵⁸, el estudio de Lawson alcanzó un alfa de 0.68 el primer día y de 0.74 el segundo día⁵⁷, semejante a lo encontrado en un metaanálisis⁵² donde el promedio no ponderado del coeficiente alfa fue de 0.62. Por otra parte, Wass aplicó un ECOE con 10 casos de 20 minutos y 3.5 h de duración obteniendo un alfa de Cronbach de 0.73⁹⁴. Los coeficientes alfa consideran solamente diferencias relativas para ser significativas en tanto que los coeficientes de generalizabilidad consideran diferencias absolutas para ser significativas.

El coeficiente G mide la proporción de la variación total que se debe a la variación en conocimientos y habilidades de los estudiantes⁹⁵. Al ser más alto el valor de G implica que las otras fuentes de variación son de menor importancia comparadas con la variación entre los estudiantes. Además otra fuente de variación entre pre y post es que uno fue formativo y el otro sumativo.

En relación con el análisis del coeficiente G del pre-test y el pos-test mostró una diferencia desfavorable al primero, 0.87, con una versión y el segundo, 0.93, con

4 versiones que implica una fuente de variación adicional, estos valores se considera que representan una confiabilidad aceptable para un examen de altas consecuencias con duración de dos horas. En un metaanálisis se encontró que el promedio no ponderado del coeficiente de generalizabilidad fue de 0.49, como sería de esperar debido a los significados y los procesos computacionales que se realizan⁵².

Lawson que realizó un ECOE con 18 estaciones a 182 estudiantes, obtuvo un coeficiente G de 0.65 el primer día y de 0.42 el segundo día⁵⁷. Hatala realizó otro ECOE con 2 escenarios que constaban de 3 secciones de 10 minutos cada uno en 20 estudiantes de cuarto año en Medicina Interna y el coeficiente G fue de 0.68⁹⁶, ambos estudios fueron de altas consecuencias y estos resultados fueron similares a nuestro pre-test. Auerawarakul realizó un ECOE en estudiantes de cuarto año de un programa de 6 años con un ECOE de 2 horas después de pasar por Medicina Interna y obtuvo un coeficiente G de 0.62⁹⁷. Estos valores fueron inferiores a los nuestros en el pre-test y pos-test lo que fortalece nuestros argumentos de validez. Boulet realizó un ECOE a 86 estudiantes de cuarto año de la carrera con 12 estaciones de 15 minutos cada uno, en total de 3 horas, obteniendo un coeficiente G de 0.76⁹⁸. Vallevand realizó un ECOE de altas consecuencias para estudiantes internacionales y su estudio tuvo un coeficiente G que varió de 0.73 a 0.84 con 14 estaciones⁹⁵. Baig también realizó un ECOE con 10 estaciones, de altas consecuencias en estudiantes internacionales y obtuvo un coeficiente G que varió de 0.51 a 0.78⁹⁹. Donnon realizó un ECOE a 29 estudiantes con 3 estaciones de 15 a 20 minutos y dos jueces y obtuvo un coeficiente G de 0.70¹⁰⁰. En contraste con estos estudios que tuvieron una duración en tiempo mayor, su coeficiente G fue igual o menor que el nuestro.

Para ECOEs formativos se ha reportado una gran dispersión respecto de los coeficientes G de 0.12–0.86^{37,101,102}. Esta variabilidad en los valores del coeficiente G pueden ser explicados porque para los exámenes formativos los estudiantes suelen no prepararse a profundidad en contraste con los exámenes sumativos de altas consecuencias.

Limitaciones

Se reconoce que la teoría de la generalizabilidad es compleja y no está al alcance de todos, incluso hay poca experiencia entre los estadísticos¹⁰³.

Existen varios softwares para realizar estudios de generalizabilidad dentro de los que se encuentran el urGENOVA⁵⁵, SAS, "G_String"¹⁰³ así como el SPSS¹⁰³.

La teoría G plantea un modelo lineal de efectos aleatorios que es procesado por las rutinas del JMP que cuenta con las herramientas necesarias para calcular los estimados de los componentes de varianza de estudios con diseños cruzados y anidados motivo por el cual se utilizó éste.

Solo tuvimos un evaluador por estación que no nos permitió determinar la confiabilidad interevaluador. Para superar esta desventaja utilizamos la teoría G (TG) para producir una medición de la confiabilidad similar en calidad.

El estudio de la confiabilidad con la teoría G nos permitió identificar diversas fuentes de variación que están implicadas en los ECOEs y hacer una predicción para incrementar la confiabilidad del ECOE, aumentando el número de estaciones.

Usar la TG puede superar algunos problemas inherentes a la confiabilidad interevaluador, tales como la sobreestimación de la confiabilidad como lo señala Fan¹⁰⁴. Hay un gran número de profesores comprometidos con nuestro ECOE, e hicimos el supuesto razonable de que cualquier error debido a diferencias de evaluadores está distribuido aleatoriamente. El error de varianza como consecuencia de las diferencias entre todos los evaluadores suponemos que sea pequeño debido a que participaron médicos clínicos como evaluadores que fueron capacitados ampliamente y estandarizados en su desempeño en el ECOE¹⁰⁵.

Debido a que el ECOE es un instrumento que evalúa las habilidades clínicas con estaciones relativamente breves, de manera fragmentada y en escenarios simulados, pero al fin de cuenta artificiales, puede ser que no evalúe de manera realista la atención integral de los pacientes, con posible alejamiento de la realidad clínica cotidiana⁹⁰. La evaluación estandarizada en ambientes artificiales puede ser problemática para inferir de ella el desempeño clínico en situaciones reales. Se ha documentado en varios estudios que existen diferencias, de magnitud variable, entre lo que los médicos realizan en situaciones controladas en ambientes

simulados, con lo que hacen realmente en la práctica¹⁰⁶. Esto implica que las competencias complejas necesarias para que un médico se desempeñe efectivamente en la vida profesional no están medidas a cabalidad por las puntuaciones obtenidas a través de la evaluación de estaciones fragmentadas que no evalúan todo el espectro de habilidades y capacidades de los médicos practicantes.

Algunos expertos proponen que se hable de evaluación de **competencia** cuando ésta se realice en representaciones controladas (ambientes artificiales, simulaciones) de la práctica profesional, y de evaluación del **desempeño** cuando se realicen las observaciones en ambientes clínicos reales de la práctica profesional (como la evaluación de 360°, la evaluación basada en el trabajo, entre otras)¹⁰⁷.

Otro aspecto importante en cuanto a la validez externa del estudio, es que el trabajo se realizó en una sola escuela de medicina, que tiene su contexto y características particulares, lo que pudiera limitar la generalización de los resultados a otras instituciones nacionales e internacionales. Sin embargo puede argumentarse que la Facultad de Medicina de la UNAM es la escuela de medicina más grande del país, es de carácter público, sus estudiantes representan una muestra amplia de los estratos económicos y sociales de la población mexicana, lo que sugiere que nuestros resultados pudieran ser aplicables a la mayoría de las escuelas de medicina nacionales.

El examen pre-test se realizó a poco más de dos meses del inicio del Internado, lo que podría incidir en la medición pre-test ya que los estudiantes contaban con algo de experiencia en el Internado, sin embargo la medición inicial fue substancialmente diferente de la final, lo que documenta que este efecto probablemente no existió o que fue menor; un aspecto importante del estudio es que el pre-test fue una evaluación formativa, mientras que el pos-test fue parte del Examen Profesional de la carrera de médico cirujano, con carácter sumativo, lo que podría contribuir a que el desempeño en la medición final fuera en parte explicado por el carácter de “altas consecuencias” del citado examen. No es posible separar el potencial impacto del carácter sumativo de la evaluación pos-

test en los resultados del estudio, sin embargo ambas mediciones se realizaron en las mismas condiciones físicas, de tiempos y estructurales, con la misma disciplina por parte de los evaluadores y estudiantes, y con similar dificultad de las estaciones determinada en la integración de los instrumentos, por lo que consideramos que las diferencias encontradas sí reflejan un cambio de la competencia clínica al final del Internado. Otras potenciales amenazas a la validez interna del estudio son la maduración de los sujetos, la aplicación del pre-test que puede mejorar por sí mismo el desempeño en el pos-test, y el efecto Hawthorne¹⁰⁶. El efecto del pre-test se minimizó al aplicar un pos-test con el mismo formato pero estaciones diferentes, equivalentes, y la participación de los mismos evaluadores. Además el tiempo transcurrido entre el pre y pos-test (10 meses) es suficiente para que esta amenaza a la validez sea controlada.

Una estrategia de diseño de investigación experimental, como un ensayo controlado con asignación al azar, pudiera resolver algunas de las amenazas a la validez interna citadas (como la ausencia de grupo control concurrente, entre otras), pero en la práctica este diseño es muy difícil de aplicar en educación médica, por las condiciones inherentes al proceso educativo, los recursos disponibles en nuestras instituciones, los aspectos éticos de no proporcionar la intervención educativa al grupo control, así como la logística y calendarios de los exámenes sumativos de la carrera, como es el Examen Profesional.

Futuras vertientes de investigación

Las características del estudiante como son el género o nivel de capacitación, como el Programa de Alta Exigencia Académica (PAEA), aunque son también fuentes potenciales de error, quedaron fuera del objetivo de este estudio por lo que se recomienda seguir con esta línea de investigación y considerar estas variables en estudios posteriores.

Otra línea de investigación podría ser analizar en profundidad las estaciones que son la base del ECOE para mejorar los resultados.

Conclusiones

El presente trabajo representa el único estudio de su tipo en investigación médica en nuestro país, presenta información original que debe interpretarse en su justa dimensión y consideramos debe aprovecharse para diseñar estudios posteriores y mejorar la calidad de la educación médica en nuestros hospitales.

También es importante considerar que este tipo de examen es un reto logístico para las instituciones, ya que se requiere coordinar múltiples sedes que trabajan en forma simultánea, en las que participan un número amplio de profesores examinadores en cada una de ellas y un grupo de más de 100 pacientes estandarizados, que deben conocer sus libretos para desempeñarse de acuerdo a los mejores estándares. Esta metodología requiere una considerable inversión de recursos humanos, financieros y materiales, pero consideramos que la importancia del objetivo, que es el formar y evaluar a los médicos que atienden los pacientes en nuestro país, justifica los gastos.

Los estudiantes de Internado se encuentran sometidos a una gran carga de trabajo en nuestras instituciones de salud y como sucede en otras partes del mundo, no se tiene la supervisión y realimentación constantes y de calidad deseadas^{18,20}. Además, aunque el programa académico de Internado considera una serie de prácticas, hacen falta una serie de experiencias más estructuradas con simuladores y realimentación formal con seguimiento.

Todavía existe en nuestro país desconocimiento sobre lo que es el ECOE, sus virtudes y limitaciones, aunque se han impartido cursos-taller en varias entidades de la República como Guanajuato, Chihuahua, Estado de México. Ha sido difícil la difusión e implementación a nivel nacional, como parte de la evaluación en educación médica, ya que únicamente en la UNAM y en la Universidad Autónoma de Guadalajara, hasta donde tenemos conocimiento, se utilizan como fase práctica del examen profesional.

El asumir el reto de mejorar la evaluación en educación médica permite avanzar en este campo, ya que contribuye a la formación profesional de los médicos proporcionando información y realimentación fundamentada, con la intención de ayudar a los estudiantes y profesores a incrementar la calidad del proceso y de los

resultados. Asimismo, los resultados de la evaluación orientan los cambios en los programas académicos y la toma de decisiones en dirección a las competencias que se busca obtener.

Por todo lo anterior, se debe consolidar la formación continua del interno de pregrado que colabora en un Sistema Nacional de Salud, fundamentada en el primer nivel de atención médica para atender la gran mayoría de los problemas de salud de la población, y valorar la ampliación del tiempo de rotación de los internos por unidades de Medicina Familiar, donde las actividades se realizan de acuerdo al perfil de egreso, lo que podría mejorar su desempeño clínico en el ECOE y en su futura práctica como Médicos Generales.

Este estudio contribuye a aumentar la cultura de la evaluación de la competencia clínica en nuestro medio, y proporciona resultados obtenidos con metodología rigurosa, mediante el uso del ECOE, que es uno de los instrumentos más objetivos que hay para valorar la competencia clínica^{25,107,108,109}.

La evaluación de la competencia clínica con el ECOE es un método que ha demostrado su utilidad. Una buena confiabilidad de este examen es de gran importancia para los estudiantes, profesores e instituciones que utilizan ECOEs de tipo formativo pero sobre todo en los sumativos de altas consecuencias.

Para fines de la seguridad del examen es comprensible tener cuatro versiones del examen que sean equivalentes y una buena planeación nos dará resultados aceptables en cuanto a la confiabilidad con el alfa de Cronbach y en forma más amplia y completa al aplicar la teoría de la Generalizabilidad.

Los ECOEs formativo y sumativo de altas consecuencias muestran evidencia de confiabilidad de acuerdo a los dos análisis realizados, el Alfa de Cronbach y la Teoría de la Generalizabilidad (TG).

Una contribución importante de este estudio inicial con la teoría G es que nos permite, utilizar los estimados de los componentes de varianza, para planear un estudio D y aumentar el coeficiente G, sin embargo observamos varias limitantes, una en el aspecto logístico como son los tiempos y el número de profesores y la segunda y más importante en el aspecto financiero porque requiere de mayores recursos económicos.

Asegurar una alta calidad en las evaluaciones de la competencia clínica de los estudiantes es una responsabilidad de los profesores. Esta responsabilidad también es importante para las autoridades académicas que tienen que tomar decisiones de altas consecuencias para otorgar la licenciatura. La responsabilidad institucional es mayor considerando a los estudiantes en sí mismos pero, más aún, a la población en general y principalmente a los pacientes que serán atendidos por estos futuros profesionales, por lo que se debe garantizar que los egresados alcanzaron la competencia clínica expresada en el perfil profesional.

X. Referencias bibliográficas

1. Martínez A, López J, Herrera P, Ocampo J, Petra, I, Uribe G, García MC, Morales S. Modelo de competencias del profesor de Medicina. *Educ. Med.* 2008; 11(3):157-167.
2. Stufflebeam D. Shinkfield E. (1993). *Evaluación sistemática*. Barcelona, España: Paidós.
3. Durante I, Lozano R, Martínez A, Morales S, Sánchez M. Evaluación de competencias en Ciencias de la Salud. México D.F. Editorial Panamericana. 2011.
4. Libro blanco de España. Competencias específicas del médico. Informe de la Comisión de Evaluación del Diseño del Título de Grado en Medicina. U. Coordinadora: Universidad de Granada, abril 2005 pag. 88 <http://www.med.uva.es/documentos/libro-blanco.pdf>
Fecha de consulta 12 de septiembre de 2014.
5. Perfil por Competencias del Médico General Mexicano Asociación Mexicana de Facultades y Escuelas de Medicina, A.C. México, 2008.
<http://www.amfem.edu.mx/intranet/descargas/competencias.pdf> Fecha de consulta 12 de septiembre de 2014.
6. The Accreditation Council for Graduate Medical Education (ACGME)
<http://www.acgme.org/acgmeweb/> Fecha de consulta 12 de septiembre de 2014.
7. Indiana University School of Medicine, <http://medicine.iu.edu/> Fecha de consulta 12 de septiembre de 2014.
8. Bandiera G, Sherbino J, Frank JR. The CanMEDS assessment tool handbook. An introductory guide to assessment methods for the CanMEDS competencies. Ottawa: The Royal College of Physicians and Surgeons of Canadá. 2006.
9. Institute of Medicine of the National Academies. <http://www.iom.edu/> Fecha de consulta 12 de septiembre de 2014.
10. Holmboe E, Hawkins R. and National Board of Medical Examiners. Practical Guide to the Evaluation of Clinical Competence. Philadelphia. Mosby Elsevier. 2008.
11. Plan de Estudios 2010. Facultad de Medicina, UNAM.
12. Albanese MA et al. Defining characteristics of educational competencies. *Med Educ* 2008; 42: 248-255.
13. Goñi J. *El Espacio Europeo de Educación Superior, un reto para la universidad*. Barcelona: Octaedro / ICE Universidad de Barcelona, 2005.
14. Elstein RM, Hundert E. Defining and Assessing Professional Competence. *JAMA*. 2002; 287: 226-235.

15. Norman G. *Defining Competence: A Methodological Review*. (15-37) En: Neufeld VR, Norman GR. (Eds.) *Assessing Clinical Competence*. New York: Springer Publishing Company; 1985. 15-37.
16. Miller G. The assessment of Clinical Skills/Competence/Performance. *Acad Med*. 1990; 65(9): S63-S67.
17. Van Der Vleuten C, Schuwirth L. Assessing Professional competence: from methods to programmes. *Med Educ*. 2005; 39: 309-317.
18. Holmboe E. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med*. 2004; 79: 16-22.
19. Norcini J. The death of the long case? *Br Med J*. 2002; 324: 408-9.
20. Williams R, Klamen D, MacGaghie W. Cognitive, social and environmental sources biases in clinical performance ratings. *Teach Learn Med*. 2003; 15(4): 28.
21. Elstein A, Shulman L, Sprafka S. *Medical Problem Solving*. Cambridge, Massachusetts: Harvard University Press. 1978. 292-4.
22. Boulet D, De Champlain A. Setting defensible performance standards on OSCEs and standardized patients examinations *Med Teach*. 2003; 25(3): 245-9
23. Walton J. *The place and purpose of clinical skills in the years after 2000*. En: *Medical Training in the European Community*. Commission of the European Communities. The Advisory Committee in Medical Training. Bruselas. Springer -Verlag. 1987. 19-26.
24. Norcini J, Blank L, Arnold G et al. The miniCEX (clinical evaluation exercise): A preliminary investigation. *Ann Intern Med*. 1995; 123: 795-799
25. Harden R, Stevenson W, Downie W, Wilson G. Assessment of Clinical Competence Using an Objective Structured Examination (OSCE). *Brit Med J*. 1975; 1: 447-451
26. Petrusa E. *Clinical performance assessments*. En: Norman GR, van der Vleuten CPM & Newble D. (eds) *International Handbook for Research in Medical Education*. 673-709. Kluwer Academic Publishers. 2002.
27. Collins J, Harden R. Real patients, simulated patients and simulators in clinical examinations. *Med Teach*. 1998; 20(6): 508-21.
28. Eva K, Neville A, Norman G. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Acad Med*. 1998; 73(10): S1-S5.
29. Hodges B, Turnbull J. *Communication skills. A resource handbook. Educating future physicians of Ontario (EFPO) Project*. 1995; 87-94
30. Newble D. Techniques for measuring clinical competence: Objective Structured Clinical Examinations. *Med Educ*. 2004; 38: 199-203.

31. Reznick R, et al. An objective structured clinical examination for the licenciate: Report of the pilot project of the Medical Council of Canada. *Acad Med.* 1992; 67(8): 487-494.
32. Sloan D, Donnelly M, Schwartz R, Strodel W. The objective structured clinical examination. The new gold standard for evaluating postgraduate clinical performance. *Ann Surg.* 1995; 222(6): 735-42.
33. Neufeld V. *Historical Perspectives on clinical competence.* En: Norman G Assessing Clinical Competence. New York. Springer Publishing Company. 1985.
34. Friedman M *Principles of Assessment.* En: Harden R. A practical guide for medical teachers. United Kingdom. Elsevier. 2005.
35. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: A review of metrics – AMEE guide no. 49. *Med Teach.* 2010; 32: 802–811
36. Iramaneerat Ch, Yudkowsky R, Myford C, Downing S. Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. *Adv Health Scie Educ.* 2008; 13: 479–493.
37. Newble D, Swanson D. Psychometric characteristics of the Objective Structured Clinical Examination. *Med Educ.* 1988; 22: 325-334.
38. Van der Vleuten C. Swanson DB. Assessment of clinical skills with standardized patients: State of the Art. *Teach Learn Med.* 1990; 2(2): 58-76
39. Larios H, Trejo J, et al. Desarrollo de la Competencia Clínica durante el Internado Médico. *Rev Educ Sup.* 2000; Vol. XXIX (3): 115-128.
40. Larios H, Trejo J, et al. Examen Profesional con el Examen Clínico Objetivo Estructurado. *Rev Educ Sup.* 2005; 34(4): 7-17.
41. Pineda H, Peña J, Mesina A, Trejo A, Mejía H. Comparación de las calificaciones de práctica clínica asignadas por los profesores con las obtenidas en el ECOE. Trabajo presentado en la Conferencia AMEE 2006, realizado en Génova, Italia.
42. Informe Técnico del Examen Clínico Objetivo Estructurado. Secretaría de Educación Médica. Facultad de Medicina. Enero 2009. UNAM.
43. Informe Técnico del Examen Clínico Objetivo Estructurado. Secretaría de Educación Médica. Facultad de Medicina. Enero 2010. UNAM.
44. Kane M. *Validation (17-64)* En: Brennan R. Educational Measurement. National Council on Measurement in Education and American Council on Education. 4th Edition. USA. Praeger Publishers 2006.
45. Downing S. Validity: on the meaningful interpretation of assessment data. *Med Educ.* 2003; 37: 830-837.

46. Durante I. Evaluación en Educación médica. En: Graue E, Sánchez M, Durante I, Rivero O, Educación en las residencias médicas. México. ETM Editores de Textos mexicanos.
47. Haertel E, *Reliability* en: Brennan R. Educational Measurement. National Council on Measurement in Education and American Council on Education. 4th. Edition. USA. Praeger Publishers. 2006. p.65-110.
48. Muñiz J. Las Teorías de los Tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*. 2010; Vol. 31(1): 57-66
49. Gessaroli M, De Champlain A: Use of Generalizability Theory in Designing and Analyzing assessments in the Health Professions. National Board of Medical Examiners. Workshop presented in the 2009 AMEE Meeting. Málaga, Spain. August 30th.
50. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16: 292-334.
51. Streiner D, Norman G. Reliability. En Health measurement scales: A practical guide to their development and use. D. L. Streiner & G. Norman (Eds.), Oxford: Oxford University Press, 2003, pp. 126–152
52. Brannick M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ*. 2011: 1- 9.
53. Matt G. Teoría G, International Encyclopedia of social and educational Psychology. Department of Psychology, San Diego State University, San Diego CA. 2001.
54. Brennan R. *Generalizability Theory. Statistics for Social Science and Public Policy*. New York. Springer-Verlag. 2010.
55. Shavelson J, Webb N. MMSS Generalizability Theory. A primer. California, USA. SAGE Publications. 1991.
56. Swanson, D. Use of generalizability theory for the design and analysis of the OSCE and other Performance –based Testing Methods. National Board of Medical Examiners. 2011.
57. Lawson D. Applying generalizability theory to high-stakes objective structured clinical examinations in a naturalistic environment. *J Manipulative Physiol Ther*. 2006; Jul-Aug 29(6): 463-7.

58. Brailovsky C, Grand'Maison P. Using Evidence to Improve Evaluation: A comprehensive Psychometric Assessment of a SP-Based OSCE Licensing Examination. *Adv Health Sci Educ Theory Pract.* 2000; 5(3): 207-219.
59. De la Fuente J, Piña E, Gutiérrez H. La formación del médico del siglo XXI y el Plan Único de Estudios de la Facultad de Medicina de la Universidad Nacional Autónoma de México. *Educación Médica y Salud.* 1994; 28: 331-40.
60. Manual del Interno. Universidad Michoacana de San Nicolás de Hidalgo: UMSNH; 2003-2004.
61. Programa Académico de Internado Médico, Plan Único de Estudios de la Facultad de Medicina. 2011.
62. Larios H, Trejo J, Cortés T. Evaluación de la Competencia Clínica. *Rev Med IMSS.* 1998; 36(1): 79-84.
63. Cummings S. Diseño de un nuevo estudio: II Estudios de cohortes. En: Diseño de la investigación científica. Un enfoque epidemiológico. Hulley SB, Cummings SR (Eds). Editorial Doyma, Barcelona. 1993. p. 69-81
64. Campbell D, Stanley J. Diseños experimentales y cuasi experimentales en la investigación social. Amorrortu editores, Argentina, 1996, p.20-28.
65. Ericsson K. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Acad Med.* 2004; 79(10): S70-81
66. Hodges B, Hanson M, Mcaughton N, Regehr G. Creating, monitoring and improving a Psychiatry OSCE. *Acad Psych.* 2002; 26(3): 134-161
67. Downing S, Tekian A, Yudkowski R. Procedures for Establishing defensible absolute passing scores on performance examinations in health professions education. *Teach Learn Med.* 2006; 18(1): 50–57
68. Cohen J. Chapter 1.4. The effect size. En: Cohen J: *Statistical power analysis for the behavioral sciences.* Nueva Jersey, EUA: Lawrence Earlbaum Associates, 2nd Ed. 1988. pp 8-14.
69. Sistema Nacional de Información en Salud. Secretaría de Salud, en <http://www.sinais.salud.gob.mx/estadisticasportema.html>, 2007. Fecha de consulta 12 de septiembre de 2014.
70. Sánchez M, Durante I, Morales S, Lozano R, Martínez A, Graue E. Plan de Estudios 2010 de la Facultad de Medicina de la Universidad Nacional Autónoma de México. *Gac Med Mex.* 2011; 147(2): 152-158.

71. Short M, Jorgensen J, Edwards J, Blakenship R, Roth B. Assessing Intern Core Competencies with an Objective Structured Clinical Examination. *J Grad Med Educ.* 2009; 1(1):30-36.
72. Townsend A, McIlvenny S, Miller C, Dunn E. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Med Educ.* 2001; 35: 841-46.
73. Sullivan G, Feinn R. Using effect size - or why the *P* value is not enough. *J Grad Med Educ.* 2012; 4(3):279-82.
74. Norman G: The effectiveness and effects of effect sizes. *Adv Health Sci Educ Theory Pract.* 2003; 8(3):183-7.
75. Trejo J, Larios H, Cortés T, Martínez N, Gracia O. Competencia Clínica en el área de Medicina Familiar. *Arch Med Fam.* 2000; 2(3): 65-74.
76. Manual Departamental Medicina General I. Programa Académico teórico-práctico. Departamento de Tercer año, SECISS, Facultad de Medicina, Universidad Nacional Autónoma de México, Cd. Universitaria, México. Enero 2012.
77. Manual Departamental Medicina General II. Programa Académico teórico-práctico. Departamento de Cuarto año, SECISS, Facultad de Medicina, Universidad Nacional Autónoma de México, Cd. Universitaria, México. Enero 2012.
78. Muller E, Harik P, Margolis M, Clauser B, Mckinley D, Boulet J. An Examination of the Relationship between Clinical Skills Examination Performance and Performance on USMLE Step 2. *Acad Med.* 78(10) Suppl. 2000: S27-S29
79. Simon S, Volkan K, Hamann C, Duffey C, Fletcher S The relationship between second-year medical students' OSCE scores and USMLE Step 1 scores. *Med Teach.* 2002; 24(5): 535–539
80. Petrusa E, Blackwell T, Rogers L, Saydjari C, Parcel S, Guckian J. An objective Measure of Clinical Performance. *Am J Med.* 1987; 83: 34-42
81. Langford N, Landray M, Martin U, Kendall M, Ferner R. Testing the practical aspects of therapeutics by objective structured clinical examination. *J Clin Pharm Ther.* 2004; 29(3): 263–266.
82. Wilkinson T, Frampton C. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Med Educ.* 2004; 38: 1111–1116

83. Eftekhari H, Labaf A, Anvari P, Jamali A, Sheybaee-Moghaddam F. Association of the pre-internship objective structured clinical examination in final year medical students with comprehensive written examinations. *Med Educ Online*. 2012, 17: 15958 - <http://dx.doi.org/10.3402/meo.v17i0.15958> Fecha de consulta 12 de septiembre de 2014.
84. Smith LJ, Price DA, Houston OB. Objective structured clinical examination compared with other forms of students assessment. *Arch Dis Child*. 1984; 59: 1173-76.
85. Hamdy H, Prasad K, Anderson B, Scherpbier A, Williams R, Zwierstra R, Cuddihy H. BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice? *Med Teach*. 2006; 28(2): 103–116
86. Petrusa E, Blackwell T, Ainsworth M. Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. *Arch Intern Med*. 1990; 150: 573-577.
87. Curtis D, Lind S, Brear S, Finzen F. The correlation of student performance in preclinical and clinical prosthodontic assessments. *J Dent Educ*. 2007; 71(3): 365–372.
88. Lawson D, Till H. Predictors of performance of students from the Canadian Memorial Chiropractic College on the licensure examinations of the Canadian Chiropractic Examining Board. *J Manipulative Physiol Ther*. 2006; 29(7): 566–569.
89. Russell B, Hoiriis K, Guagliardo J. Life University Correlation Between Student Performances on Course Level Integrated Clinical Skills Examinations and Objective Structured Clinical Examinations in a Chiropractic College Program. *J Chiropr Educ*. 2012; 26(2): 138-145 U.S.A. doi: 10.7899/JCE-10-026
90. Harden R, Gleeson F. Assessment of Clinical Competence Using an Objective Structured Clinical Examination (OSCE). *Med Educ*. 1979; 13: 41-54.
91. Patricio M. A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies. Tesis Doctoral, Facultad de Medicina, Universidad de Lisboa. 2012. <http://tinyurl.com/c5c7cbs> Fecha de consulta 12 de septiembre de 2014.
92. Cronbach L, Shavelson R. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas*. 2004; 64(3): 391–418.
93. Trejo J, Martínez A, Méndez I, Morales S, Ruíz L, Sánchez M. Evaluación de la Competencia Clínica con el Examen Clínico Objetivo Estructurado (ECO) en el Internado Médico de la UNAM. *Gac Med Mex*. 150: 8-17

94. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of Clinical competence. *Lancet*. 2001; 357: 945–49
95. Vallevand A, Violato C. A Predictive and Construct Validity Study of a High-Stakes Objective Clinical Examination for Assessing the Clinical Competence of International Medical Graduates. *Teach Learn Med*. 2012; 24(2): 168–176
96. Hatala R, Marr S, Cuncic C, Bacchus C. Modification of an OSCE format to enhance patient continuity in a high-stakes assessment of clinical performance. *BMC Med Educ*. 2011. May 24; 11: 23.
97. Auewarakul Ch, Downing S, Jaturatamrong U, Praditsuwan R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment method. *Med Educ*. 2005; 39: 276–283s
98. Boulet J, McKinley D, Whelan G, Hambleton R. Quality assurance methods for performance-based assessments. *Adv Health Sci Educ*. 2003; 8: 27–47.
99. Baig L, Violato C. Temporal stability of objective structured clinical exams: a longitudinal study employing item response theory. *BMC Medical Education*. 2012; 12:121.
100. Donnon T, Paolucci E. A generalizability study of the medical judgment vignettes interview to assess students' noncognitive attributes for medical school. *BMC Medical Education*. 2008; 8: 58 doi:10.1186/1472-6920-8-58 Fecha de consulta 12 de septiembre de 2014.
101. Hull A, Hodder S, Berger B, Ginsberg D, Lindheim N, Quan J, Kleinhenz M. Validity of three clinical performance assessments of internal medicine clerks. *Acad Med*. 1995; 70: 517–22.
102. Wilkinson T, Newble D, Wilson P, Cater J, Helms R. Development of a three-centre simultaneous objective structured clinical examination. *Med Educ*. 2000; 34: 798–807
103. Bloch R, Norman G: Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012; 34: 960–992
104. Fan X, Chen M: Published studies of inter-rater reliability often overestimate reliability: Computing the correct coefficient. *Educ Psych Meas*. 2000; 60: 532-542
105. Hamann C, Volkan K, Fishman M, Silvestri R, Simon S, Fletcher S. How well do second-year students learn physical diagnosis? Observational study of an objective structured clinical examination(OSCE) *BMC Med Educ*. 2002;2(1): 1-11
106. Rethans J, Norcini J, Barón-Maldonado M, et al. The relationship between competence and performance: implications for assessing practice performance. *Med Educ*. 2002; 36: 901-9.

107. Van der Vleuten C. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ.* 1996; 1: 41-67.
108. Mavis B, Henry R. Between a rock and a hard place: finding a place for the OSCE in medical education. *Med Educ.* 2002; 36: 408-9.
109. Hill D, Guinea A, McCarthy W. Formative assessment: a student perspective. *Med Educ.* 1994; 28: 394-9.

ANEXO 1 MATERIAL PARA EL ALUMNO

INSTRUCCIONES

Lea cuidadosamente la presentación clínica y siga las indicaciones que se le dan a continuación. Cuenta usted con 6 minutos.

PRESENTACIÓN CLÍNICA

Se trata de femenino que acude a su consulta por presentar cefalea.

- 1. Realice el interrogatorio dirigido al padecimiento actual.**
- 2. Exprese a la paciente en voz alta su diagnóstico presuntivo.**

ANEXO 2 MATERIAL PARA EL EVALUADOR

LISTA DE COTEJO

El estudiante interrogó:

1. Inicio del padecimiento actual
2. Primer episodio o repetición
3. Localización del dolor
4. Tipo
5. Irradiación
6. Intensidad
7. Duración
8. Frecuencia
9. Factores que lo desencadenan: (Exceso de trabajo, crisis emocionales)
10. Factores que lo exacerban
11. Factores que lo disminuyen
12. *Náusea* **Síntomas acompañantes**
13. Vómito
14. Fotofobia
15. Fonofobia o Acúfenos
16. Ansiedad (Preocupación constante, irritabilidad)
17. Grado de incapacidad que le produce
18. Terapéutica empleada
19. Su diagnóstico presuntivo fue: **Cefalea Tensional**

HABILIDADES DE COMUNICACIÓN INTERPERSONAL

1	2	3	4	5	6	7	8	9
INSATISFACTORIO			SATISFACTORIO			SUPERIOR		
Desatento al saludar, presentarse y preguntar el nombre a la paciente.						Saludó, se presentó y le preguntó su nombre a la paciente.		
Muestra frialdad y falta de respeto durante la entrevista.						Muestra calidez y respeto durante la entrevista.		
Poco atento a informar lo que va a realizar.						Informó que le iba a realizar unas preguntas.		
Su lenguaje es inadecuado a nivel de la paciente.						Utilizó un lenguaje claro y comprensible.		
Muestra poca atención a lo que dice la paciente.						Escucha con atención.		
Evita el contacto visual.						Mantiene contacto visual.		
Se muestra distante a los sentimientos de la paciente						Comprensión intelectual de los sentimientos de la paciente.		
Plantea preguntas sin relación al padecimiento						Plantea las preguntas relacionadas con el padecimiento.		

ANEXO 3 MATERIAL PARA EL PACIENTE ESTANDARIZADO

LIBRETO CEFALEA TENSIONAL

Usted acude por dolor de cabeza, es ama de casa, con problemas económicos y familiares; por lo que esta muy estresada. Estará pálida y con ojeras. Está un poco nerviosa durante la consulta. **Contestar a los alumnos solo lo que le pregunten y lo mismo, si le preguntan algo que no esté en el libreto, conteste no sé y si le dicen que más tiene usted diga como qué, doctor.**

¿Cuál es su nombre?.....Contesta con su nombre
¿Qué edad tiene?.....Conteste su edad
¿A qué se dedica?.....Trabajo en mi casa doctor, soy ama de casa.
¿Porqué acude a consulta?.....Es que me duele la cabeza Dr. (a) Usted señala la nuca)
¿Es la primera vez que le sucede esto?.....No. Dr. (Dra)
¿Desde cuándo le sucede esto?.....Ya tiene como un mes
¿El dolor que presenta el día de hoy es similar a los que se han presentado desde un mes?.....Sí, igual
¿Cuándo inicio este dolor?.....Desde hace dos horas que lo tengo
¿Cómo me describiría Usted su dolor, es decir como si le explotara la cabeza, como si se la jalarán, como si se la oprimieran?.....Cómo si me la apretarán muy fuerte.
¿Se le corre hacia algún lado?.....Sí doctor, en un inicio estaba ahí en la nuca pero después se extendió a la frente como una cinta.
Si le dijera que el dolor más fuerte es 10 y el más leve es de 1 ¿qué calificación le pondría usted a su dolor?.....de 8 Doctor
¿Cuánto tiempo le dura?.....Es variable, pero en promedio como de 5 horas.
¿Se presenta en alguna hora del día en especial?.....Sí, empieza como a medio día y se va poniendo peor a lo largo del día.
¿Con qué aumenta ese dolor?.....Pues no sé, doctor, pero en las tardes me empeora
Ese dolor de cabeza ¿se agrava con el movimiento?.....No. doctor (a)
¿Qué la llegue a despertar por las noches?.....No
¿Con qué disminuye el dolor?.....Cuando me tranquilizo un poco o me tomo las pastillas
Entonces ¿ha tomado algún medicamento para el dolor?.....Sí
¿Qué tomó?.....Dos aspirinas
¿Hace cuánto tiempo?.....Hace una hora
Además de este dolor ¿ha tenido náuseas o vómitos?.....No
¿Le molesta la luz?.....No
¿Le molesta el ruido?.....Pues, un poco
¿Se siente más cansada de lo normal?.....Sí
¿Qué no tenga ganas de hacer algo?.....Sí
¿Este dolor se presenta diariamente?.....Sí, doctor
Cuándo le da el dolor ¿tiene que dejar lo que está haciendo para acostarse, es decir interfiere con sus actividades?.....Generalmente no
¿Usted ¿a qué le achaca ese dolor?.....No sé.

¿Ha tenido algún golpe en la cabeza?.....No
 ¿Ha estado estresada por algo?.....Pues sí, a mi esposo le ha ido mal en los negocios y ya no me alcanza para el gasto.
 ¿Hay algún enfermo en la familia?.....sí, mi mamá está muy enferma y la tengo que cuidar

ANEXO 4 TABLA DE ESPECIFICACIONES DEL ECOE

Marzo 2011

Área	Interrogatorio	Exploración Física	Interpretación laboratorio o gabinete	Consejería	Dx y manejo	Análisis artículo
Cirugía	1		1			
Gineco-obstetricia	3					
Medicina Familiar	2	1			1	
Medicina Interna	2	1	1			
Pediatría	2			1		
Urgencias			1			
Medicina basada en evidencias						1

ANEXO 5 MATRIZ DE UN EXAMEN ECOE

Marzo 2011

ESTACION	SISTEMA	PROBLEMA
1.- Interrogatorio, Dx y Manejo	Cirugía	Apendicitis
2.- Interrogatorio FR y Dx	Ginecobstetricia	Embarazo de alto riesgo
3.- Interrogatorio, Dx y Manejo	Pediatría	Gastroenteritis viral
4.- Interpretación Rx y Dx	Cirugía	Oclusión intestinal
5.- Exploración Física y Dx	Medicina Interna	Infección de vías aéreas
6.- Habilidad técnica	Medicina Familiar	Aplicación de DIU
7.- Interrogatorio FR, Dx y manejo	Ginecobstetricia	FR Ca Mama
8.- Consejería	Pediatría	Alimentación infantil
9.- Dx y manejo	Medicina Familiar	Hipertensión arterial
10.- Interrogatorio y Dx	Medicina Interna	Angina estable
11.- Interrogatorio, Dx y manejo	Medicina Familiar	Pielonefritis aguda
12.- Interrogatorio y Dx	Pediatría	Control del niño
13.- Interrogatorio, Dx y manejo	Medicina Interna	Úlcera duodenal
14.- Interpretación Lab y Dx	Medicina Interna	Cetoacidosis diabética
15.- Interrogatorio, Dx y manejo	Ginecobstetricia	Placenta previa
16.- Exploración física, Dx y manejo	Medicina Familiar	Colitis amibiana
17.- Interpretación EKG y Dx	Urgencias	Infarto de Miocardio
18.- Análisis estudio investigación	Medicina Basada Evidencias	Influenza

ANEXO 6 HOJA DE LECTOR OPTICO



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE MEDICINA
SECRETARÍA DE EDUCACIÓN MÉDICA
EXAMEN CLÍNICO OBJETIVO ESTRUCTURADO
CÉDULA DE CAPTACIÓN



FOLIO
NOMBRE DEL ALUMNO
NUM. CTA.
GRUPO
SEDE
ESTACIÓN
FECHA

FOLIO			
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9

SEDE
1
2
3
4
5
6
7
8
9
0

ESTACION			
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9

**FIRMA Y NOMBRE
 DEL PROFESOR**

HCI
0
1
2
3
4
5
6
7
8
9

ESTACION
0
1
2
3
4
5
6
7
8
9

PE
0
1
2
3
4
5
6
7
8
9

NÚMERO DE ÍTEM
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

Edición 2010. El Examen Médico S.A. de C.V. Tel: 55 5224 0101 y 5224 0102

L-044-01-1002011-01

ANEXO 7 PRESUPUESTO POR EXAMEN

Cada examen para 60 alumnos requirió:

45 pacientes estandarizados (300.00 x T)	\$ 13,500.00
17 profesores examinadores (400.00 x T)	\$ 20,400.00
Servicio de Café y lunch	\$ 6,000.00
Lectores ópticos	\$
Tiempo extra	\$
Transporte y gasolina	\$
Vigilancia y limpieza	\$ 250.00
Total	\$ 40,150.00

Para evaluar a 360 alumnos de Internado médico en marzo 2010 se necesita un día en seis sedes.

Total seis sedes, tres turnos \$ 240,900.00 por día