

Evaluación en educación médica

GUILLERMO ZEPEDA F.¹, CONSTANZA COGHLAN A.², CAROLINA HERESI V.³,
RIGOBERTO MARÍN C.⁴, ÁLVARO HERRERA A.⁵, THELMA SUAU C.⁶

Assessment in medical education

Medical education is the university pedagogical process that leads to the training in competencies of students of health sciences, both undergraduate and postgraduate, and of professionals in continuous training. Within this process, probably one of the most complex edges is the evaluation of each student, due to the inherent principles of this action, such as justice, asymmetry of power between student and teacher, academic responsibility, among others, and because of the desirable characteristics of an instrument such as validity, reliability, acceptability, feasibility and educational impact. The evaluation of clinical competencies is a complex process, because it requires accounting for the integration of knowledge, skills, and attitudes, mobilized in a context close to professional performance. In recent decades, evidence has accumulated on the usefulness of various evaluation instruments in the field of health sciences education, highlighting the role of timely feedback in educational impact. For all the above, this article updates important aspects of the evaluation of a student's teaching-learning process by listing the types of instruments most used, their main characteristics, and in which situations they can be used.

Keywords: Education, Medical; Clinical competence; Reproducibility of results; Students; Universities.

Resumen

La educación médica es el proceso pedagógico universitario que conlleva a la formación en competencias de los estudiantes de las ciencias de la salud, tanto en pregrado como en postgrado, y de profesionales en formación continua. Dentro de este proceso probablemente una de las aristas más complejas es la evaluación de cada estudiante, debido a los principios inherentes de esta acción, como son la justicia, asimetría de poder entre estudiante y docente, responsabilidad académica, entre otros, y por las características deseables de un instrumento de evaluación como son la validez, confiabilidad, aceptabilidad, factibilidad y el impacto educacional. La evaluación de competencias clínicas

¹ Neumólogo Pediatra, Magister en Educación en Ciencias de la Salud, Profesor Asociado, Departamento de Pediatría y Cirugía Infantil Campus Norte, Escuela de Medicina y Facultad de Medicina. Universidad de Chile, Hospital de niños Dr. Roberto del Río. ORCID: 0000-0001-5073-5011.

² Estudiante 4° año Carrera de Medicina, Facultad de Medicina, Universidad de Chile.

³ Neuróloga Pediatra, Magister en Educación en Ciencias de la Salud, Profesora Asistente, Departamento de Pediatría y Cirugía Infantil Campus Norte, Escuela de Medicina y Facultad de Medicina, Universidad de Chile, Hospital de niños Dr. Roberto del Río. ORCID: 0000-0002-0689-5859.

⁴ Profesor de Biología y PhD en Educación, Profesor Asociado, Departamento de Educación en Ciencias de la Salud y Escuela de Medicina, Facultad de Medicina, Universidad de Chile. ORCID: 0000-0001-5484-7926.

⁵ Obstetra-ginecólogo, Magister en Educación, Doctorando en Educación y Sociedad, Profesor Asistente, Facultad de Medicina de la Universidad San Sebastián. Departamento de Obstetricia y Ginecología Campus Occidente, Facultad de Medicina, Universidad de Chile, Hospital San Juan de Dios. ORCID: 0009-0007-4861-2144.

⁶ Pediatra, Magister en Educación en Ciencias de la Salud, Profesora Asociada, Directora Escuela de Medicina. Universidad de Chile. Departamento de Pediatría y Cirugía Infantil Campus Sur, Facultad de Medicina. Universidad de Chile, Hospital Dr. Exequiel González Cortés. ORCID: 0009-0005-2985-8956.

es un proceso complejo, porque requiere dar cuenta de la integración de conocimientos, habilidades y actitudes, movilizadas en un contexto cercano al desempeño profesional. En las últimas décadas se ha acumulado evidencia sobre la utilidad de diversos instrumentos de evaluación en el ámbito de la educación en ciencias de la salud, destacando el rol de la retroalimentación oportuna en el impacto educacional. Por todo lo anterior, este artículo actualiza aspectos importantes de la evaluación del proceso enseñanza aprendizaje de un estudiante enumerando los tipos de instrumentos más utilizados, sus principales características, y en qué situaciones se pueden usar.

Palabras clave: Educación Médica; Competencia Clínica; Reproducibilidad de los resultados; Estudiantes; Universidades.

Introducción

La educación médica (EM) es un proceso dinámico de enseñanza-aprendizaje para la adquisición y desarrollo de competencias del ámbito médico, tanto de estudiantes de pregrado y postgrado, como de profesionales en la etapa de continua formación. Su principal objetivo es preparar a los futuros y actuales médicos para el manejo integral del paciente basado en los avances tecnológicos y nuevos descubrimientos científicos¹.

De esta manera, la finalidad de la EM universitaria es asegurar que los estudiantes cuenten con los conocimientos, experiencias, habilidades y conductas que los prepare en el ejercicio de la profesión, tanto desde la atención a pacientes como desde la investigación, liderazgo, gestión y la misma docencia médica¹⁻⁵.

Por este motivo, la EM es de vital importancia para el aseguramiento de una atención de calidad hacia los pacientes, por lo que adquiere gran relevancia considerar los métodos que se usan para evaluar el desempeño de los estudiantes, midiendo la forma en la que adquieren conocimientos y su capacidad para ponerlos en práctica, siendo de esta manera la evaluación de la EM un proceso longitudinal, progresivo y multidimensional. En este proceso, los estudiantes deben tener un compromiso activo con su propio aprendizaje para el buen desarrollo de sus habilidades clínicas y actitudinales^{2,3}.

El proceso de evaluación de competencias clínicas es complejo, porque implica dar cuenta de los conocimientos y habilidades de cada estudiante, en escenarios lo más reales posibles, y con posibilidad de retroalimentación oportuna que impacte adecuadamente el aprendizaje. La acumulación de evidencia sobre la utilidad de múltiples instrumentos de evaluación en los últimos años obliga a realizar una actualización sobre sus características principales, destacando los escenarios donde cada uno puede tener una mayor utilidad.

Evaluación de la educación médica

Para garantizar que los resultados de aprendizajes se logren y sean cumplidos se requiere una comprobación de las competencias que se buscan desarrollar, siendo la evaluación de los conocimientos teóricos y prácticos una parte integral del proceso formativo de los estudiantes y profesionales. Esta evaluación se puede realizar mediante distintos sistemas, los cuales se clasifican básicamente en diagnósticos, formativos y sumativos⁶.

Las evaluaciones diagnósticas se realizan previo a iniciar el proceso formativo y sirven para determinar el estado basal o inicial del estudiante, programar actividades de nivelación y establecer una secuencia curricular adecuada al nivel de formación del estudiante.

Las evaluaciones formativas tienen como objetivo generar y entregar retroalimentación sobre el aprendizaje del evaluado, así como también generar un espacio de reflexión en torno a sus propias capacidades utilizadas en el proceso y generar puntos de referencia para orientarlo en éste.

Las evaluaciones sumativas se centran en los resultados, en los cuales se basan para generar una toma de decisión con respecto a la competencia profesional^{2,6,7}.

Una vez decidida la forma que tendrá la evaluación, se debe escoger el instrumento más adecuado para llegar al propósito de ésta, dentro de los cuales encontramos distintas herramientas según la competencia que se busca evaluar, abarcando desde pruebas escritas, entrega de productos hasta observaciones con pacientes simulados o reales. Se debe considerar que estos sistemas de evaluación deben tener características que certifiquen su habilidad para determinar la adquisición y consiguiente aplicación de las competencias médicas⁶.

Estas características son descritas por el modelo de utilidad de van der Vleuten, el cual busca evaluar los instrumentos utilizados para el proceso de evaluación de la EM mediante la apli-

cación de cinco criterios: validez, confiabilidad, aceptabilidad, factibilidad o costos y el impacto educacional⁸. En este modelo se plantea una ecuación conceptual donde cada una de estas variables contribuye con diferente peso a la utilidad del instrumento. Si una de estas variables tiene valor cero, el instrumento es inútil (Tabla 1). El peso de cada una de estas variables puede variar en función de los objetivos de la evaluación. Por ejemplo, en una evaluación formativa va a ser muy relevante el impacto del instrumento sobre el aprendizaje, mientras que, en una evaluación sumativa, donde se toman decisiones importantes como aprobar o reprobar al estudiante, los pesos de las variables confiabilidad y validez deben ser muy altos. Si nos ponemos en un escenario donde es necesario evaluar en forma simultánea a un grupo grande de estudiantes, los costos y factibilidad de un instrumento pesan diferente a una situación donde se requiere evaluar a grupo pequeño de estudiantes en forma simultánea.

Por lo tanto, el método que se utilice para evaluar las competencias debe seleccionarse en base a los objetivos de aprendizaje y a que su diseño cumpla con las características anteriormente mencionadas, entendiendo que en cada situación pueden ser más apropiados instrumentos diferentes⁹.

Independientemente del método de evaluación elegido, debe tener algunas características para lograr que este proceso sea exitoso en un amplio sentido y que son: múltiples métodos de evaluación, múltiples evaluadores, selección y capacitación de los evaluadores, una reconceptualización de la evaluación psicométrica y el reconocimiento del trabajo grupal en alcanzar decisiones críticas acerca de la competencia evaluada¹⁰.

Pirámide de Miller

Con el propósito de establecer conceptos esenciales para la EM y guiar a los educadores para mejorar la experiencia educativa y en el desarrollo del diseño curricular, George Miller, creó en 1990 una esquematización de las bases de la evaluación de la EM, en forma de una pirámide, la cual sigue vigente actualmente, debido a su capacidad de asistencia que orienta el diseño curricular, los métodos de enseñanza y de evaluación, además de tener un sustento teórico, generando su uso transversal a diferentes niveles educativos. Además, de aportar a la comprensión de la evaluación como un proceso jerárquico que conduce al aprendizaje^{11,12}.

La pirámide diseñada por Miller se divide

Tabla 1: Modelo de Utilidad de van der Vleuten⁸

Utilidad = Validez x Confiabilidad x Aceptabilidad x Impacto Educativo x 1/Costos U= V · C · A · IE · 1/C		
El modelo se plantea como una ecuación conceptual, en la que, si uno de los elementos es pobre, el instrumento tiene baja utilidad.		
V	Validez	La validez consiste en el grado con el cual el instrumento mide lo que se busca medir, asegurando la coherencia entre el objetivo y el instrumento evaluativo. Esta misma característica se divide en validez de contenido, validez de criterio y validez de constructo. La validez del instrumento de evaluación depende más del formato en el que se presenta la pregunta, que del formato de respuesta, lo que es importante de considerar en el diseño de los instrumentos.
C	Confiabilidad	La confiabilidad es el grado en que el instrumento mide de forma consistente lo que se pretende medir, con resultados reproducibles en el tiempo y al ser usados por diferentes observadores. Puede ser afectada por muchos factores, como el juicio de los evaluadores, los casos usados, el nerviosismo del evaluado y las condiciones del examen utilizado.
A	Aceptabilidad	La aceptabilidad hace referencia al grado de aceptación del instrumento, tanto por evaluadores como por estudiantes, considerando así adecuado el instrumento para cumplir el objetivo de la evaluación. En esta variable se incluyen aspectos subjetivos, como la percepción de justicia del instrumento utilizado.
IE	Impacto Educativo	El impacto educacional condiciona el aprendizaje y la práctica futura, ya que el tipo de evaluación determina la forma en la cual el estudiante o profesional va a dirigir su estudio. En este aspecto influyen los contenidos que son priorizados en la evaluación, las características de los instrumentos utilizados, la frecuencia en que se realiza la evaluación y la presencia o no de retroalimentación, entre otros aspectos. El impacto sobre el aprendizaje a veces es difícil de predecir, y por esta razón debe ser monitoreado permanentemente
C	Costos	Los costos y la factibilidad se relacionan con la posibilidad de realizar la evaluación habiendo considerado todos los costos y recursos necesarios para ella.

en 4 niveles de aprendizaje, donde cada uno se relaciona con un aspecto esencial, reflejando conocimiento (*knowledge*), competencia (*competence*), desempeño (*performance*) y actuar médico (*action*), desde la base hasta el ápice, siendo los dos primeros niveles enfocados en el componente cognitivo de la EM, mientras que los dos niveles más altos en el componente conductual¹¹. De esta forma, la base de la pirámide hace referencia a los conocimientos en abstracto, mediante el término “saber”; el segundo nivel busca los conocimientos contextualizados, incluyendo la toma de decisiones y razonamiento clínico mediante el “saber cómo”, el tercer nivel busca “demostrar cómo” incluyendo el comportamiento y habilidades en la competencia clínica, pero que se muestra en entornos simulados, mientras que el vértice de la pirámide hace alusión al “hacer”, es decir demostrar las competencias en contextos reales^{7,11} (Figura 1).

Métodos de evaluación

Miller postuló que “ningún método de evaluación único puede proporcionar todos los datos necesarios para juzgar algo tan complejo como la prestación de servicios profesionales por parte de

un médico exitoso”¹¹. Esto lleva a la necesidad de utilizar distintas estrategias para poder medir los diversos niveles de aprendizaje propuestos, dado que un solo tipo de evaluación no puede medir todas las competencias que se requieren.

En este sentido, para una evaluación basada en competencias significativas se deben realizar evaluaciones continuas y oportunas, con revisiones periódicas, usar múltiples evaluadores y evaluaciones que permitan asesorar de manera y con el propósito correcto la EM, sintetizar datos de forma grupal, la capacitación docente de todos los evaluadores y relaciones óptimas entre los participantes de la evaluación para una retroalimentación adecuada, como a su vez estas evaluaciones deben de ser capaces de proporcionar información sobre el desempeño real, la capacidad de adaptación al cambio, encontrar y generar nuevos conocimientos, y mejorar el desempeño general^{2,10}.

Para planificar estas evaluaciones se debe tener en consideración dos objetivos principales: la evaluación del aprendizaje (sumativa) y la evaluación para el aprendizaje (formativa)¹⁰.

Con respecto al primer objetivo (evaluación sumativa), en base a la pirámide diseñada por Miller se puede atribuir a cada nivel descrito anteriormente un método de evaluación. En la

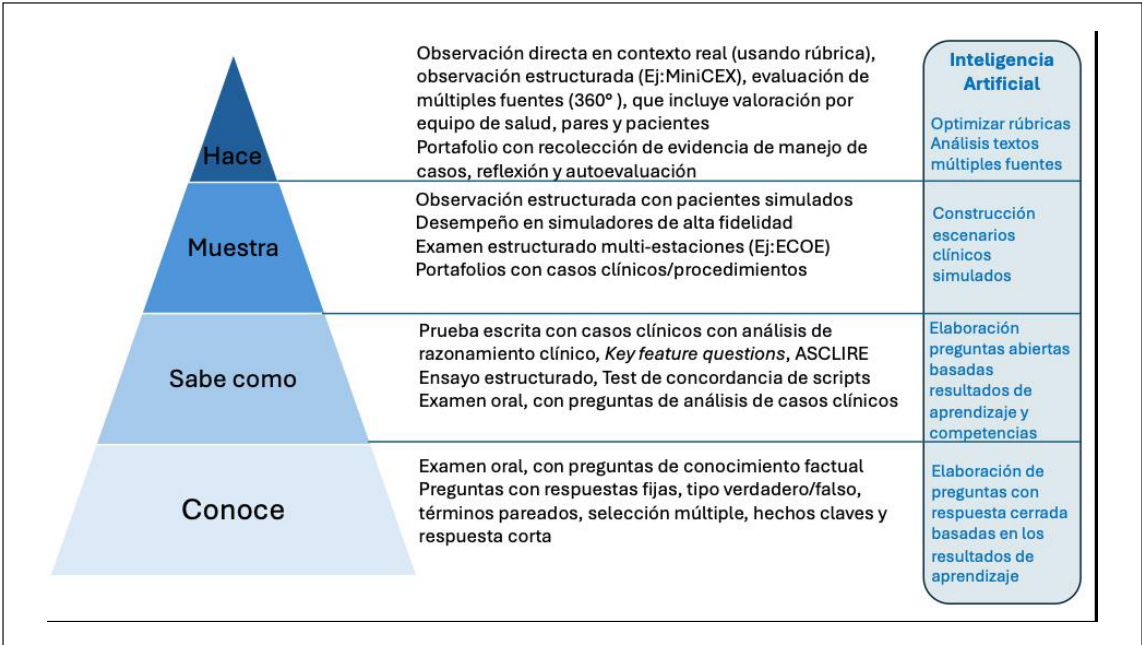


Figura 1. Adaptación de la pirámide de Miller. Se incluye ejemplos de estrategias de evaluación de los diferentes niveles, y ejemplos de aplicación de herramientas de inteligencia artificial para el desarrollo de instrumentos de evaluación. Significado de abreviaturas: MiniCEX=Mini Clinical Evaluation Exercise;360°= evaluación en 360°; ECOE=Evaluación Clínica Objetiva Estructurada; ASCLIRE=Assessing Clinical Reasoning.

base de ésta se evalúa el conocimiento mediante pruebas escritas con preguntas de selección múltiple o preguntas de desarrollo breve. En el segundo nivel, en el que se busca aplicar el conocimiento mediante ensayos, ejercicios para resolver problemas clínicos y preguntas de selección múltiple extendidas. En el tercer nivel, para medir la demostración de las competencias, se usan entornos de simulación, ya que estos generan una aproximación de cómo se desempeñaría la persona que está siendo evaluada en una situación similar a la real. En este nivel está el Examen Clínico Objetivo Estructurado (ECO), también conocido como OSCE (*Objective Structured Clinical Examination*). Se pueden usar grabaciones y pautas de observación estructuradas. Por último, en el vértice de la pirámide se mide la práctica por una observación directa en contextos clínicos reales mediante observación directa o videograbada, portafolio, evaluación de registros clínicos, entre otros. Estas evaluaciones deben de ir acompañadas de retroalimentación formativa para poder atribuir al proceso de aprendizaje y además lograr una autoevaluación informada, en donde el mismo estudiante en base a distintas fuentes y el uso de un portafolio, documenta y reflexiona sobre su propio aprendizaje, lo cual se denomina como “búsqueda de evaluación autodirigida”^{6-7,10-15} (Figura 1).

Para cumplir el segundo propósito (evaluación para el aprendizaje) se busca el uso de múltiples métodos para compensar las deficiencias de los diferentes métodos individuales, como también el uso de múltiples evaluadores para compensar sesgos, efectos de halo (una impresión general positiva o negativa interfiere con la evaluación de aspectos específicos) y efecto lenidad (falta de rigor en la evaluación para evitar un conflicto), dentro de lo cual es de ayuda la capacitación de quienes sean encargados de efectuar las evaluaciones, de forma de garantizar que los pacientes que reciban atención de los estudiantes sea de forma segura, efectiva y centrada en el mismo paciente, como también de contar con evaluadores competentes que cuenten con los conocimientos de las competencias a evaluar y comprendan las tareas de observación y registro del rol del evaluador.

Los instrumentos de evaluación deben poder valorar la capacidad de razonamiento clínico de cada estudiante, el cual es definido como una competencia compleja, central en el desempeño del profesional médico, que implica la síntesis de datos clínicos e investigativos para llegar a un diagnóstico diferencial, y con ello proponer planes de manejos apropiados; siendo así el proceso

de pensamiento y toma de decisiones asociado con la práctica clínica¹⁵.

El razonamiento clínico, se puede evaluar también en los niveles superiores de la pirámide mediante el ECOE y el desempeño en simulaciones para el nivel de demostrar cómo, y observaciones directas realizadas por docentes clínicos, mini-CEX y evaluaciones 360^{2,7,10}.

Estos procesos generalmente se han centrado en evaluaciones de forma individual, pero también se ha sugerido el uso de procesos grupales para el empleo de toma de decisiones de forma de elevar la confiabilidad y encontrar deficiencias en el profesionalismo, lo cual se puede realizar mediante comités de competencias clínicas¹⁰.

A su vez, otra innovación que se ha buscado en el proceso de la evaluación en la EM es la aplicación de la tecnología como herramienta para facilitar el proceso, ya que puede proporcionar retroalimentaciones, generar representaciones gráficas del desempeño tanto individual como de un grupo colectivo, para estimular la reflexión de las habilidades propias, como de las brechas de éste, y las cuales también pueden ser de orientación para directores de programas y así comparar los procesos académicos, además de servir como canal de información de decisiones¹⁰.

A continuación, se enumerarán algunos instrumentos utilizados para evaluar en la EM y su utilidad dividiéndolos según los niveles de la pirámide de Miller:

A. Primer nivel (conoce)

Este se realiza principalmente mediante interrogatorios orales de preguntas teóricas, exámenes escritos y, ocasionalmente, prácticas clínicas observadas por docentes clínicos, enfocando esta evaluación en el proceso de memorización de conocimientos teóricos, con menos énfasis en habilidades clínicas y de razonamiento.

En relación con los exámenes escritos, éstos son un método que principalmente evalúa el contenido teórico, y para ello tiene distintas formas de hacerlo, ya sea preguntas de selección opción múltiple, de respuesta abierta, entre otras. A su vez estas preguntas pueden encontrarse descontextualizadas, o pueden enriquecerse entregando contexto clínico².

- Preguntas de opción múltiple

Son comúnmente utilizadas en instancias evaluativas, debido a que pueden abarcar un gran número de contenidos, la facilidad de calificación y administración en un periodo corto de tiempo, de forma de generar una evaluación estandarizada de bajo costo. Estas a su vez pueden generar

una respuesta basada en el reconocimiento de la alternativa, que en una pregunta abierta, sin alternativas pre-establecidas, no hubiera tenido respuesta, llamado “efecto cueing” que genera una problemática en torno al razonamiento diagnóstico y puede llevar a errores en la práctica por decisiones prematuras antes de llegar al diagnóstico correcto². En la Tabla 2 se describen las ventajas, desventajas y sugerencias para desarrollar este instrumento.

- Preguntas de respuesta corta abierta

Permiten respuestas libres, ya que no cuentan con opciones de selección, fomentando así una respuesta detallada y contextualizada, involucrando procesos cognitivos más complejos y minimizando así el cueing².

- Exámenes orales

Tradicionalmente se han usado tanto a nivel formativo como a nivel sumativo y en este último caso se exige que sean dos o más examinadores.

Dentro de sus limitaciones se describe que tiende a ser subjetiva y que tiene sesgos de diferentes tipos, siendo una de las principales fortalezas la retroalimentación inmediata que puede ser entregada en este tipo de evaluación².

B. Segundo nivel (Sabe cómo)

- Prueba de concordancia de guiones (Scripts concordance test)

En una prueba escrita se presentan una serie de casos clínicos, y mediante una escala tipo Likert, se atribuye un valor a un elemento dado, con base en cuánto hace más o menos probable el diagnóstico dado o un resultado particular, siendo situaciones ambiguas que pueden proporcionar información sobre el juicio clínico del estudiante, y así sus resultados relacionarse con el nivel de formación y desempeño futuro. La calificación se asigna comparando las respuestas de cada estudiante con las respuestas de un grupo de expertos¹⁶. La confiabilidad de la prueba

Tabla 2. Ventajas, desventajas y sugerencias para desarrollar preguntas de opción múltiple

Ventajas	Desventajas	Sugerencias de confección
Permiten evaluar una gran cantidad de contenidos del <i>curriculum</i> en poco tiempo.	Si no se cumplen con los criterios técnicos de calidad, se induce a respuesta correcta o incorrecta.	Pregunta posible de responder, sin necesidad de mirar las alternativas.
Permiten evaluar los niveles de “conocimiento” y la “aplicación del conocimiento” (nivel 1 y 2 de Miller)	Siempre existe la posibilidad de respuesta al azar.	Pregunta debe formularse de manera breve, sencilla, directa y en orden lógico. Evitar los términos confusos, abreviaturas desconocidas, ambigüedades e información irrelevante.
A mayor cantidad de preguntas, aumenta la confiabilidad del instrumento de evaluación.	Se requiere tiempo y entrenamiento para construir preguntas válidas, confiables y de orden superior del pensamiento.	Debe existir concordancia gramatical entre la pregunta y las alternativas: forma verbal, género y número para no inducir la respuesta.
Fáciles de confeccionar, aplicar y corregir, lo que reduce el costo.	Preguntas mal diseñadas favorecen el recuerdo superficial en vez del aprendizaje profundo.	Pregunta debe expresarse en forma positiva. Evitar en negativo: “lo falso”, “lo incorrecto”, “la excepción”.
Existen criterios definidos de calidad técnica para su confección.	No permiten evaluar habilidades de comunicación efectiva, trabajo en equipo, argumentación, redacción o desempeño clínico.	Debe evitar que la pregunta solicite una opinión. Ej. “cuál cree usted”, “según lo que usted piensa”, “según su opinión”, “de acuerdo con sus conocimientos”.
Tienen una corrección objetiva, que puede ser manual o automática, que facilita obtener un puntaje y calificación (nota).		Use 4 alternativas verosímiles, lógicas, homogéneas en longitud y contenido.
Son de alta aceptabilidad por los docentes y estudiantes.		Alternativas con cifras deben ordenarse en forma ascendente o descendente.
		Alternativas sin expresiones ambiguas, absolutas, ninguna de las anteriores, todas las anteriores o formato k (números romanos).

es alta si se usan al menos 4 ítems de preguntas en al menos 25 casos clínicos diferentes¹⁷. Este tipo de preguntas pueden tener un importante rol en la evaluación formativa y en la de resultados intermedios. En la Tabla 3 se exponen ventajas, desventajas y sugerencias para la realización de este instrumento.

**- Preguntas de características clave
(Key feature questions)**

Se trata de preguntas escritas en las que los estudiantes deben identificar elementos esenciales (claves) dentro de una viñeta clínica para el diagnóstico o manejo terapéutico. Este tipo de preguntas ha demostrado tener una buena consistencia interna (entre 0,49 y 0,95), moderada a alta validez de contenido, buena validez de constructo (pudiendo diferenciar adecuadamente a los expertos y novatos), y sus resultados muestran adecuada predicción del desempeño clínico futuro, lo que las convierte en un buen formato de pregunta para exámenes escritos realizados a grupos grandes de estudiantes en forma simultánea¹⁸.

**- Evaluación del Razonamiento Clínico o
ASCLIRE (Assessing clinical reasoning)**

Consiste en una serie de casos clínicos pre-

sentados en un computador, que permiten analizar como el estudiante obtiene datos clínicos relevantes, los interpreta y los sintetiza. Cada caso se inicia con una viñeta clínica escrita y un video, y se va complementando con información solicitada por el estudiante, de manera interactiva. Finalmente, cada estudiante selecciona el diagnóstico desde una lista de posibilidades. Se califica tanto a la precisión diagnóstica, como el tiempo invertido en obtener el resultado. Este método ha demostrado tener adecuada validez, con una tasa mayor de precisión diagnóstica, menor tiempo de decisión y mejor elección de información relevante entre los expertos al compararlos con estudiantes iniciales¹⁹.

C. Tercer nivel (Muestra)

**- Evaluación Clínica Objetiva Estructurada
(ECOE u OSCE)**

Es considerado un buen instrumento de evaluación para evaluar el desempeño clínico en un contexto de simulación, mediante un sistema de evaluación de diversas habilidades clínicas, como la toma de historia clínica, examen físico y diagnósticos diferenciales justificados mediante el uso de estaciones en un circuito con pacientes

**Tabla 3. Ventajas, desventajas y sugerencias para la implementación de la Prueba de concordancia de guiones
(Scripts concordance test)**

Ventajas	Desventajas	Sugerencias para implementación
Permite visualizar el razonamiento del estudiante al resolver un caso clínico con información incompleta.	Requiere un panel de al menos 10 expertos que resuelvan todos los casos.	Los casos deben ser breves y representativos del área clínica que se está evaluando. El nivel de complejidad del caso debe ser adecuado para el nivel de formación del estudiante.
No tiene una sola respuesta correcta. Se construye usando casos reales y la opinión de un panel de expertos, que pueden elegir como correcta diferentes alternativas. El puntaje de cada pregunta se otorga de acuerdo con la cantidad de respuestas del panel de expertos que coincide con el estudiante.	La puntuación es relativa a la frecuencia de respuestas que se obtuvo del panel de expertos. Esto puede ser difícil de aceptar para estudiantes y académicos no familiarizados con el instrumento.	Los casos deben ser breves. Se describe una situación inicial, una hipótesis y los ítems con nueva información. El estudiante debe contestar en qué medida la nueva información afecta a la probabilidad de la hipótesis planteada (escala de Likert con 5 elementos, desde -2 “descartada” hasta +2 “comprobada”).
No busca respuestas aprendidas “de memoria”. Las respuestas permiten evaluar razonamiento y fomentar el pensamiento crítico.	No es útil para evaluar conocimiento factual.	Tiene buena confiabilidad usando al menos 25 casos, con al menos 4 ítems cada uno.
Puede adaptarse a diferentes niveles de formación, desde pregrado hasta formación de especialistas, ajustando la complejidad del caso.	Necesita entrenamiento docente.	Convocar a un panel de 10-20 expertos, lo más diverso posible. Sus respuestas se usan para calcular la calificación de los estudiantes.

simulados de forma sistemática y estandarizada, lo que disminuye la variabilidad dependiente del paciente y del examinador, mediante la capacitación de los involucrados (académicos o actores como pacientes simulados) y permite evaluar el desarrollo de competencias en comunicación y profesionalismo. Es un examen con múltiples estaciones, con tiempo estandarizado para desarrollar cada actividad. En cada estación el estudiante se enfrenta a diferentes situaciones que buscan evaluar diferentes competencias clínicas dentro del área cognitiva, actitudinal y destrezas psicomotoras, intercaladas con estaciones de descanso entre ellas.

El OSCE permite la evaluación en el tercer nivel de la pirámide de Miller y cuenta con validez, confiabilidad y factibilidad, además permite evaluar un espectro amplio de habilidades debido a las diferentes estaciones de simulación, tanto para su evaluación de manera formativa como en un formato sumativo. Tiene la ventaja de poder evaluar exactamente de la misma manera a un conjunto amplio de estudiantes, y en una amplia variedad de escenarios clínicos²⁰⁻²³. En la Tabla 4 se exponen ventajas, desventajas y sugerencias para la realización de un OSCE.

- *Evaluación clínica integral sistemática (ECIS)*

Es un sistema conformado por tres unidades básicas: un examen teórico, estaciones de simulación estáticas y dinámicas tipo ECOE, y un módulo de autoevaluación. El examen teórico

contiene preguntas en las que se describen casos clínicos con preguntas de selección múltiple, el cual se resuelve en una plataforma digital. El segundo módulo evalúa el desempeño en situaciones clínicas específicas mediante 4 estaciones que miden habilidades en interrogatorio, exploración física y actitud en cada una de ellas. Por último, la autoevaluación del estudiante permite establecer objetivos como identificar fortalezas y dificultades, y una propia valoración del proceso formativo, mediante una comparación con un estándar grabado al finalizar el apartado de estaciones clínicas²⁴.

D. Cuarto nivel (Hace)

- *Observaciones directas*

Se realizan en situaciones médicas reales por docentes clínicos. Habitualmente en nuestro medio, la observación es informal y basada en impresiones, más que analizadas de forma sistemática. Para mejorar este tipo de evaluaciones, se puede analizar el encuentro clínico mediante el modelo de Bowen que enumera pasos secuenciales, desde la recolección de datos como la historia y el examen físico, para luego representar precisamente el problema y generar diagnósticos diferenciales, de los cuales se selecciona uno principal y con este se plantea un manejo. Otra estrategia para esto es la herramienta de razonamiento mediante cinco componentes: la recolección de datos dirigidos por hipótesis, representación del problema, diagnóstico diferencial

Tabla 4. Ventajas, desventajas y sugerencias para la realización de Evaluación Clínica Objetiva Estructurada (ECOE)

Ventajas	Desventajas	Sugerencias
Objetivo y estandarizado, todos los estudiantes se someten al mismo tipo de evaluación.	No evalúa un enfoque holístico de manejo de un paciente	Realizar un estudio previo de factibilidad de recursos docentes, tecnológicos, económicos, entre otros
Tiene alta confiabilidad y validez	Las estaciones tienen un tiempo acotado, por lo cual, puede generar estrés en el estudiantado	Diseñar el OSCE lo más cercano a la práctica clínica real
Permite evaluar una amplia gama de competencias en poco tiempo	Requiere una planificación con mucho tiempo invertido de parte del equipo docente	Realizar una inducción a los estudiantes para que conozcan este tipo de evaluación
Simula situaciones reales, por lo cual se puede usar con fines sumativos o formativos	Algunos aprendizajes son difíciles de evaluar con este instrumento	Elegir las competencias más representativas de lo que se requiere evaluar, dado el número limitado de estaciones.
Se puede brindar una retroalimentación completa en diferentes áreas	Requiere un alto costo de implementación	Debe ser explícito el ámbito de la competencia evaluada en cada estación.

priorizado, pruebas de alto valor y metacognición, en esta se emplea una matriz de puntuación simple calificando el grado de dominio de cada componente. Otro sistema que puede ser utilizado por los docentes para la observación directa es el marco de 'IDEA' que evalúa la documentación escrita en busca del razonamiento mediante un resumen interpretativo, diagnóstico diferencial, explicación del razonamiento y diagnósticos alternativos considerados, y por último, el modelo de preceptor de un minuto da un marco simple para una evaluación mediante la presentación de casos cortos, evaluando la capacidad de síntesis junto al razonamiento y, la llegada y defensa de un diagnóstico¹⁵.

Otra manera de mejorar las observaciones directas es incluir pautas de observación de desempeños que permitan registrar una evaluación estructurada y a partir de los resultados, realizar retroalimentación al estudiante, promoviendo la reflexión sobre el desempeño observado, incluyendo conocimientos, razonamiento clínico, profesionalismo, comunicación y registro adecuado de los hallazgos para facilitar la continuidad de los cuidados. Algunos de estos instrumentos se comentan en la siguiente sección.

- *Evaluación en 360 grados*

Es un método de valoración de competencias que se origina del mundo corporativo²¹. En el ámbito de la salud se ha utilizado para evaluar de forma integral el proceso formativo, mediante el recibimiento de retroalimentación desde distintos grupos de interés con una pauta de observación estructurada, dentro de los que se encuentran autoridades académicas, estudiantes, autoevaluación, pares académicos hasta otros profesionales no médicos y los mismos pacientes, para así generar una evaluación más integral del estudiante. Su énfasis se da en la evaluación de las habilidades comunicacionales, de trabajo interprofesional y de profesionalismo, en donde después de que cada participante dé la retroalimentación y se termina de generar esta información, se debe recopilar de manera gráfica de forma de reflejar las calificaciones promedio del estudiante en cada ítem y compararlas con el resto de los estudiantes del mismo nivel, junto a su autoevaluación y los comentarios textuales, pero de forma anónima, y así entregarle esta información al evaluado de forma presencial. Presenta validez de contenido, indicadores de aceptable a buena confiabilidad en las distintas escalas y una buena capacidad de discriminación de los ítems de cada una de ellas, pero a la vez requiere de muchas encuestas para llegar a estos aspectos²⁵⁻²⁷.

- *Mini-CEX (Mini Clinical Evaluation Exercise) u Observación estructurada de la práctica clínica (OEPC)*

El Mini-CEX se ha consolidado como una herramienta eficaz para evaluar el desempeño clínico en contextos reales, al integrar la observación directa con una retroalimentación formativa inmediata. Su aporte reside en la valoración del razonamiento clínico, la comunicación y la toma de decisiones, dimensiones alineadas con los niveles superiores de la pirámide de Miller. La evidencia clásica y reciente respalda su uso sistemático como promotor de una cultura de evaluación continua que aproxima la enseñanza a la práctica profesional auténtica^{28,29}.

Para optimizar su efectividad, es clave asegurar validez contextual y consistencia evaluativa. Esto exige formación específica de evaluadores y múltiples observaciones en escenarios diversos con pacientes de distinta complejidad, estrategias que incrementan confiabilidad y reducen sesgos de observación. El marco de las recomendaciones sobre observación directa –los “do’s y don’ts”– ofrece pautas operativas transferibles a la aplicación del Mini-CEX, tales como aumentar la frecuencia de observaciones, usar criterios explícitos y promover un diálogo de mejora³⁰.

La calidad del *feedback* (retroalimentación) es el principal determinante del impacto educativo: debe ser específico, oportuno y co-construido para favorecer la autorregulación y la planificación de acciones de mejora. La literatura sobre coaching en educación médica identifica habilidades conversacionales que incrementan la receptividad al *feedback* y su uso por parte del estudiante³¹.

En cuanto al soporte digital, la evidencia muestra que formularios y aplicaciones móviles mejoran la trazabilidad, el análisis longitudinal y la oportunidad de la retroalimentación, siempre que no sustituyan la conversación formativa. Implementaciones recientes de Mini-CEX en plataformas *web* y aplicaciones móviles han mostrado validez, aceptabilidad y confiabilidad adecuadas, además de un mayor tiempo efectivo dedicado a la retroalimentación; no obstante, requieren cuidar el componente relacional del proceso³².

Finalmente, el impacto del Mini-CEX depende menos del formato y más del clima educativo, la formación del evaluador y la sistematicidad del proceso. Integrarlo dentro de un plan de evaluación programática –con múltiples puntos de observación, estándares explícitos y reuniones de revisión– potencia su valor formativo y su alineación con el desempeño profesional esperado³³⁻³⁷.

Tabla 5. Ventajas, desventajas y sugerencias para la implementación del Mini-CEX
(*Mini Clinical Evaluation Exercise*)

Aspecto	Ventajas	Desventajas	Sugerencias para su implementación
Validez	Evalúa el desempeño real en contextos clínicos auténticos	Puede verse afectada por la variabilidad de los casos clínicos.	Seleccionar casos representativos de competencias clave y estandarizar criterios de observación.
Confiabilidad	Mejora con múltiples observaciones y evaluadores	Riesgo de sesgos del evaluador y efecto halo.	Capacitar a los evaluadores y emplear rúbricas estandarizadas.
Factibilidad	Instrumento breve (15–30 minutos), de bajo costo y adaptable	Requiere compromiso docente sostenido.	Integrar el Mini-CEX dentro del plan curricular y del desarrollo profesional docente.
Impacto educativo	Favorece la retroalimentación inmediata y la reflexión del estudiante	La calidad de la retroalimentación puede ser heterogénea.	Establecer pautas de retroalimentación estructurada y sesiones reflexivas posteriores.

La Tabla 5 muestra las ventajas, desventajas y recomendaciones para la confección de un Mini-CEX.

Portafolio

Es un método de evaluación del nivel superior de la pirámide de Miller, principalmente formativo, basado en una recopilación de documentos sobre el aprendizaje que permite generar una instancia de autorreflexión sobre áreas específicas que demuestran la obtención de competencias, mediante el seguimiento de estas actividades, las que pueden ser monitorizadas también por el docente. Aporta información sobre las capacidades que se buscan lograr y su proceso de adquisición, aumenta la capacidad reflexiva, el desarrollo personal y profesional, habilidades de comunicación, de gestión de información, pensamiento crítico e interacción social. Este se puede hacer en formato manual como también digital, este último es llamado portafolio electrónico (e-P) el cual necesita el análisis de situaciones y acciones que se requieren realizar para llegar a los conocimientos esperados, mediante la movilización del estudiante con la recopilación de textos (notas clínicas, registros de procedimientos, otros), presentaciones, audios, imágenes, videograbaciones, páginas web, infografías, entre otros materiales que den cuenta del proceso de formación y reflexión sobre la práctica con autoevaluaciones, planes de aprendizaje y ensayos reflexivos.

El portafolio evalúa tanto el proceso como el resultado, generando una motivación a la autorreflexión del desempeño, con logros y dificultades, el desarrollo de habilidades colaborativas y capacidad de resolución de problemas, siendo

un enfoque que proporciona la posibilidad de demostrar su aprendizaje a lo largo del tiempo, en diversos contextos, reduciendo la dependencia de una sola oportunidad de evaluación como indicador de desempeño, acumulando evidencias de cómo actuaría el estudiante en contextos similares reales. Además, este puede ser utilizado como herramienta docente para la planificación del curso, siendo una herramienta útil en varios participantes de la EM^{2,38-40}.

Evaluación basada en el lugar de trabajo (*Workplace-based assessment*)

Este tipo de evaluación está en relación con el ápice de la pirámide de Miller ya que evalúa a los estudiantes en un contexto real de atención de pacientes. El objetivo es dar retroalimentación basado en la observación en un marco clínico real y se han utilizado diferentes instrumentos como mini-CEX (mini, observación directa de habilidades en procedimientos (sigla DOPS en inglés), discusión basada en casos (sigla Cbd en inglés), evaluación estructurada objetiva de habilidades técnicas (sigla OSATS en inglés) y también diferentes evaluaciones en 360° usando Mini-PAT (*mini-peer assessment tool*) y TAB (*team assessment of behavior*)⁴¹.

- *Uso de inteligencia artificial*

La inteligencia artificial (IA) está transformando progresivamente la evaluación en educación médica al automatizar tareas repetitivas, mejorar la retroalimentación y permitir análisis más profundos del aprendizaje. Su aplicación se extiende desde la generación automatizada de ítems y la corrección de respuestas abiertas mediante pro-

cesamiento del lenguaje natural, hasta la creación de escenarios clínicos simulados y entornos de realidad virtual para la evaluación de competencias^{42,43}. Estos desarrollos se enmarcan en un enfoque programático de la evaluación, en el que la IA se concibe como un recurso complementario que fortalece la validez y la oportunidad de las decisiones educativas, pero que no sustituye la observación directa ni el juicio clínico docente.

Entre los beneficios observados, la IA puede aumentar la confiabilidad y la factibilidad de los procesos evaluativos al reducir la variabilidad entre evaluadores y al ofrecer datos longitudinales sobre el desempeño de los estudiantes. En particular, los sistemas de análisis del aprendizaje (*learning analytics*) permiten integrar grandes volúmenes de información procedente de diversas fuentes —como portafolios electrónicos, resultados de exámenes y evaluaciones clínicas—, generando visualizaciones que orientan la toma de decisiones y la planificación de intervenciones educativas⁴⁴. Sin embargo, su implementación debe basarse en criterios pedagógicos y éticos claros, con supervisión docente permanente.

La incorporación de IA en la evaluación médica también plantea desafíos relevantes. La transparencia de los algoritmos, la protección de los datos personales y la detección de sesgos son aspectos críticos que deben abordarse mediante marcos institucionales de gobernanza. La literatura reciente enfatiza que toda herramienta automatizada debe ser evaluada en función de su impacto educativo y de su capacidad para mejorar la equidad y la personalización del aprendizaje, evitando la dependencia excesiva de decisiones algorítmicas⁴⁵. En este sentido, la supervisión humana sigue siendo indispensable para garantizar la interpretación contextualizada de los resultados.

Desde una perspectiva práctica, la integración efectiva de la IA en la evaluación médica requiere tres condiciones fundamentales: 1) desarrollo de pilotos controlados que comparen el rendimiento de los nuevos sistemas con los métodos tradicionales; 2) formación docente en competencias digitales y en el uso pedagógico de los datos generados; y 3) interoperabilidad de las plataformas tecnológicas con los sistemas curriculares y de gestión académica existentes⁴³.

En síntesis, la IA tiene un potencial significativo para optimizar la evaluación médica al fortalecer su validez, confiabilidad, factibilidad e impacto educativo. No obstante, su valor real depende de su integración crítica en un ecosistema educativo guiado por principios de ética, transparencia y acompañamiento humano.

Conclusiones

El proceso de evaluación de competencias en el ámbito de salud es un proceso complejo. La selección de los instrumentos de evaluación adecuados a cada etapa de formación debe ser coherente con los resultados de aprendizaje y las competencias que se busca desarrollar. La recomendación actual es planificar estrategias de evaluación con múltiples puntos de control, con retroalimentación oportuna, que permita acompañar la reflexión del estudiante sobre su proceso de aprendizaje, y que den cuenta del desarrollo de competencias complejas, como el razonamiento clínico. Conocer instrumentos diseñados para la evaluación de los niveles más altos de la pirámide de Miller, permite acercarse al desempeño profesional auténtico con el objetivo de robustecer las estrategias de evaluación, impactando profundamente el proceso de aprendizaje.

Referencias bibliográficas

1. WORLD MEDICAL ASSOCIATION (WMA). Statement on medical education, 2020. <https://www.wma.net/policies-post/wma-statement-on-medical-education/> (accedido en octubre 2025).
2. EPSTEIN RM. Assessment in medical education. *N Engl J Med*. 2007 Jan 25;356(4):387-96.
3. HARDEN R, LAIDLAW J. Essential skills for a medical teacher. Second Edition. Edinburgh: Editorial Elsevier; 2017.
4. SILVA-VELASCO E, LÓPEZ-ABALLE M, MAYEDO-NÚÑEZ Y. La educación médica y su papel en la formación clínica de los estudiantes de medicina. *Luz* 2023;22(3):150-60. Epub 15 de septiembre de 2023. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1814-151X2023000300150&lng=es&tlng=es.
5. HARDEN R, LILLEY P. The eight roles of the medical teacher. Edinburgh: Editorial Elsevier; 2018.
6. NOLLA-DOMENJÓ M. La evaluación en educación médica: Principios básicos. *Educación Médica* 2009;12(4):223-9. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1575-18132009000500004&lng=es&tlng=es
7. ZEPEDA G, HERESI C. Educación médica. En Zepe-da G, Molina V, Mira M y Acuña M (editores). Manual de atención pediátrica del Hospital Roberto del Río. 6ª Edición. Santiago de Chile: Editorial Mediterráneo; 2023. Págs. 501-10.
8. VAN DER VLEUTEN CP. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ Theory Pract*. 1996 Jan;1(1):41-67. doi: 10.1007/BF00596229. PMID: 24178994.

9. BARÓN M. Sistemas de evaluación. *Educación Médica* 2005; 8(Supl. 2):15-7. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1575-18132005000600006&lng=es&tlng=es
10. LOCKYER J, CARRACCIO C, CHAN MK, HART D, SMEE S, TOUCHIE C, et al. Core principles of assessment in competency-based medical education. *Med Teach*. 2017 Jun;39(6):609-16. doi: 10.1080/0142159X.2017.1315082. PMID: 28598746.
11. MILLER GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990 Sep;65(9 Suppl):S63-7. doi: 10.1097/00001888-199009000-00045. PMID: 2400509.
12. DOMÍNGUEZ-TORRES LC, VEGA-PEÑA NV. Pyramids in medical education: Their conceptualization and utility summarized. *Rev Colomb Obstet Ginecol*. 2023 Jun 30;74(2):163-74.
13. WITHERIDGE A, FERNS G, SCOTT-SMITH W. Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. *Int J Med Educ*. 2019 Oct 25;10:191-2.
14. CRUESS RL, CRUESS SR, STEINERT Y. Amending Miller's Pyramid to Include Professional Identity Formation. *Acad Med*. 2016 Feb;91(2):180-5.
15. THAMPY H, WILLERT E, RAMANI S. Assessing Clinical Reasoning: Targeting the Higher Levels of the Pyramid. *J Gen Intern Med*. 2019 Aug;34(8):1631-6. Erratum in: *J Gen Intern Med*. 2019 Dec 17. doi: 10.1007/s11606-019-05593-4.
16. DORY V, GAGNON R, VANPEE D, CHARLIN B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ*. 2012 Jun;46(6):552-63.
17. GAGNON R, CHARLIN B, LAMBERT C, CARRIÈRE B, VAN DER VLEUTEN C. Script concordance testing: more cases or more questions? *Adv Health Sci Educ Theory Pract*. 2009 Aug;14(3):367-75.
18. HRYNCHAK P, TAKAHASHI SG, NAYER M. Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ*. 2014 Sep;48(9):870-83.
19. KUNINA-HABENICHT O, HAUTZ WE, KNIGGE M, SPIES C, AHLERS O. Assessing clinical reasoning (ASCLIRE): Instrument development and validation. *Adv Health Sci Educ Theory Pract*. 2015 Dec;20(5):1205-24.
20. HARDEN R, LILLEY P, PATRICIO M. The definitive guide to the OSCE. Edinburgh: Editorial Elsevier; 2016.
21. FORNELLS J. La evaluación en educación médica *Medwave* 2009;09(1):3702 doi: 10.5867/medwave.2009.01.3702
22. TICSE R. El Examen Clínico Objetivo Estructurado (ECO) en la evaluación de competencias de comunicación y profesionalismo en los programas de especialización en Medicina. *Rev Méd Hered* 2017; 28(3):192. Disponible en: <https://revistas.upch.edu.pe/index.php/RMH/article/view/3188>
23. TRIVIÑO X, VÁSQUEZ A, MENA A, LÓPEZ A, ALDUNATE M, VARAS M, et al. Aplicación del Examen Clínico Objetivo Estructurado (OSCE) en la evaluación final del internado de pediatría en dos escuelas de medicina *Rev Med Chil*. 2002 Jul;130(7):817-24.
24. GUZMÁN-VALDIVIA G, DOMÍNGUEZ A, GARCÍA A, MENESES D, VELASCO MT. Un sistema de evaluación integral en educación médica (ECIS). *Revista del Centro de Investigación de la Universidad La Salle* 2023;15(60):161-78. <https://doi.org/10.26457/recein.v15i60.2940>
25. COINDREAU R. Evaluación 360 grados - guía completa. Integratec (Internet). México (2025, 01, 22). Disponible en <https://www.integratec.com/blog/evaluacion-360-grados-guia-completa.html>
26. BÁEZ-ROJAS C, CÓRDOVA-LEÓN K, FERNÁNDEZ-HUERTA L, VILLAGRA-ASTUDILLO R, ARAVENA-CANESE L. Modelo de retroalimentación mediante evaluación de 360° para la docencia de pregrado en ciencias de la salud. *Educación Médica* 2021; 24(4):173. <https://doi.org/10.33588/fem.244.1133>
27. CORREA D, GONZÁLEZ C. Uso de la evaluación en 360 grados para medir competencias en residentes de programas de postítulo de especialidades médicas. *ARS med [Internet]*. 26 de marzo de 2020 [citado 19 de octubre de 2025];45(1):26-31. Disponible en: <https://www.arsmedica.cl/index.php/MED/article/view/1562>
28. NORCINI J, BURCH V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007 Nov;29(9):855-71.
29. RAMANI S, KÖNINGS KD, MANN KV, PISARSKI EE, VAN DER VLEUTEN CPM. About Politeness, Face, and Feedback: Exploring Resident and Faculty Perceptions of How Institutional Feedback Culture Influences Feedback Practices. *Acad Med*. 2018 Sep;93(9):1348-58.
30. KOGAN JR, HATALA R, HAUER KE, HOLMBOE E. Guidelines: The do's, don'ts and don't knows of direct observation of clinical skills in medical education. *Perspect Med Educ*. 2017 Oct;6(5):286-305.
31. ARMSON H, LOCKYER JM, ZETKULIC M, KÖNINGS KD, SARGEANT J. Identifying coaching skills to improve feedback use in postgraduate medical education. *Med Educ*. 2019 May;53(5):477-493.
32. FUENTES-CIMMA J, FUENTES-LÓPEZ E, ISBEJ ESPÓSITO L, DE LA FUENTE C, RIQUELME PÉREZ A, CLAUDSORFF H, et al. Utility analysis of an adapted Mini-CEX WebApp for clinical practice assessment in physiotherapy undergraduate students. *Frontiers in Education* 2023; 8, 943709. <https://doi.org/10.3389/feduc.2023.943709>
33. BABOOLAL SO, SINGARAM VS. Specialist training: workplace-based assessments impact on teaching, learning and feedback to support competency-based postgraduate programs. *BMC Med Educ*. 2023 Dec 11;23(1):941.

34. LIU C. An introduction to workplace-based assessments. *Gastroenterol Hepatol Bed Bench*. 2012 Winter;5(1):24-8.
35. FORNELLS-VALLÉS JM. El ABC del Mini-CEX. *Educación Médica* 2009;12(2):83-89. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1575-18132009000300004&lng=es&tlng=es
36. OLASCOAGA A, RIQUELME A. Aplicación longitudinal del Mini Clinical Examination (Mini-CEX) en médicos residentes. *Educación médica* 2019; 20 (S1): 25–28. <https://doi.org/10.1016/j.edumed.2017.07.014>
37. DURNING SJ, CATION LJ, MARKERT RJ, PANGARO LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med*. 2002 Sep;77(9):900-4.
38. GUTIÉRREZ M, MEDEROS H, GÓMEZ G, MONTALVO A, HERNÁNDEZ E, MIRALLES E. El uso del portafolio para el pregrado en Ciencias Médicas. *Educ Med Super [Internet]*. 2019 Jun [citado 2025 Oct 19];33(2). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412019000200016&lng=es
39. MORÁN-BARRIOS J. La evaluación del desempeño o de las competencias en la práctica clínica. 1ª Parte: principios y métodos, ventajas y desventajas. *Educación médica* 2016; 17(4): 130–139. <https://doi.org/10.1016/j.edumed.2016.07.001>
40. DÍAZ J. Bases teóricas del portafolio en la educación médica basada en competencias. *Educación Médica Superior* 2016; 30(1) Recuperado en 21 de mayo de 2025, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412016000100016&lng=es&tlng=es
41. BARRETT A, GALVIN R, STEINERT Y, SCHERPBIER A, O'SHAUGHNESSY A, HORGAN M, et al. A BEME (Best Evidence in Medical Education) review of the use of workplace-based assessment in identifying and remediating underperformance among postgraduate medical trainees: BEME Guide No. 43. *Med Teach*. 2016 Dec;38(12):1188-98.
42. MASTERS K. Artificial Intelligence developments in medical education: a conceptual and practical framework. *Med Ed Publish* (2016). 2020 Oct 26; 9:239.
43. MASTERS K, MACNEIL H, BENJAMIN J, CARVER T, NEMETHY K, VALANCI-AROESTY S, et al. Artificial Intelligence in Health Professions Education assessment: AMEE Guide No. 178. *Med Teach*. 2025 Sep;47(9):1410-1424.
44. SETHI A. Artificial intelligence in health professions education. *Journal of Shalamar Medical & Dental College* 2024;5(1):1-3. <https://doi.org/10.53685/jshmdc.v5i1.227>
45. MASTERS K. Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158. *Med Teach*. 2023 Jun;45(6):574-584.

Correspondencia a:
Dr. Guillermo Zepeda Flores
Subdirector Escuela de Medicina
Universidad de Chile
Avda. Independencia 1027
Santiago-Chile
Email: gzepedaflores@uchile.cl