



La evaluación en educación médica. Principios básicos

M. Nolla-Domenjó

Objetivo. Proporcionar una visión global de los aspectos más relevantes a tener en cuenta en un sistema de evaluación. **Desarrollo.** Se proponen dos modelos fundamentales como marco de referencia de los procesos de evaluación. Se describen las características principales de cualquier sistema de evaluación. Finalmente, se apuntan los retos a los que se enfrenta en la actualidad el amplio campo de la evaluación en educación médica.

Palabras clave. Competencia profesional. Educación médica. Evaluación.

Evaluation in medical education. Basic principles

Aim. To provide a global vision of the most relevant aspects to take into account of an evaluation system. **Development.** The article also proposes two different fundamental models as frameworks for the evaluation process. The main characteristics applicable to any evaluation system are described. Finally the paper points the challenges that evaluation in medical education is facing for the near future.

Key words. Evaluation. Medical education. Professional competence.

Introducción

La evaluación en educación médica es un campo realmente amplio. Desde la evaluación de los aprendizajes a los sistemas de calidad educativa y evaluación de políticas educativas, existen multitud de aspectos que se deben tener en cuenta en los proyectos formativos. Este artículo tratará

especialmente la evaluación de los aprendizajes y, en definitiva, versará sobre uno de los temas de más relevancia en la actualidad: la evaluación de la competencia profesional.

En el módulo 2 de este programa de formación, Palés [1] define el concepto de competencia y describe su complejidad. Introduce el concepto de educación basada en resultados de aprendizaje. Este enfoque lleva implícita la evaluación. Los resultados de aprendizaje son competencias comprobadas, es decir, evaluadas. Este abordaje refuerza la idea de que la evaluación es una parte integrante de un programa o proceso de formación. Debe pensarse en la evaluación desde el inicio, desde la planificación del proceso formativo.

¿Qué? y ¿para qué?

La planificación del sistema de evaluación a utilizar incluye muchas preguntas a plantearse: ¿qué?, ¿para qué?, ¿cómo?, ¿cuándo?, ¿quién?, etc. Todas ellas deben clarificarse en algún momento u otro; sin embargo, las dos más fundamentales que deben iniciar todo el proceso son las dos primeras: ¿qué? y ¿para qué? Las respuestas a estas dos preguntas determina el abanico de posibilidades de las restantes.

¿Qué?

Ya se ha comentado la amplitud de los aspectos susceptibles de ser evaluados. Para no confundirse a la hora de responder a las otras preguntas

es fundamental haber escogido y definido los aspectos que se quieren evaluar.

La utilización de modelos teóricos puede ayudar en la definición del objeto de evaluación. Existen diversos modelos, pero en aras a la brevedad y utilidad del presente artículo, se han escogido dos de los más utilizados por los educadores en general y, en concreto, por los educadores en ciencias de la salud. Como se verá más adelante, ambos modelos son compatibles en el planteamiento de los sistemas de evaluación.

Kirkpatrick [2,3] ha aportado al mundo de la educación un esquema breve y claro en el que basar los procesos de evaluación. Su modelo se centra en cuatro niveles a tener en cuenta, especialmente para los programas de posgrado y desarrollo profesional continuo (Tabla I):

- *Nivel 1: reacción.* Lo que se evalúa es la reacción de los participantes en el programa de formación, en otras palabras, se busca información sobre el grado de satisfacción de los estudiantes.
- *Nivel 2: aprendizaje.* Este nivel se centra de pleno en la evaluación de las competencias adquiridas.
- *Nivel 3: conducta.* Se evalúa la transferencia de los aprendizajes al lugar de trabajo. La dificultad de este nivel estriba en los múltiples factores que influyen en la calidad del desempeño laboral, además de la competencia profesional, y que hacen más complejo relacionar unívocamente, de forma causal, los cambios o no de comportamiento con el grado de eficacia del programa de formación.
- *Nivel 4: resultados.* Lo que se pretende evaluar en este nivel es el impacto en la población o sociedad. Si el nivel anterior tenía sus dificultades en poder atribuir cambios en la conducta profesional a los programas de formación, averiguar su impacto en la sociedad es aún un reto mayor.

En el ámbito de la educación médica, y en concreto en la evaluación de la competencia profesional, el esquema más difundido y utilizado es el ideado por Miller [4] en 1990. Al final de su carrera, Miller invirtió sus últimos años en el reto aún vigente de la evaluación de la competencia profesional. Conceptualizó cuatro niveles de competencia profesional que plasmó en forma de pirámide, y que ha pasado a conocerse más

Tabla I. Niveles del modelo de Kirkpatrick.

- Nivel 1: evaluación de la reacción (satisfacción)
- Nivel 2: evaluación del aprendizaje (competencia)
- Nivel 3: evaluación de la conducta (transferencia)
- Nivel 4: evaluación de los resultados (impacto)

comúnmente como ‘pirámide de Miller’. Relacionando este modelo con el expuesto con anterioridad, se puede decir que el modelo de Miller se centra en los niveles 2 y 3 de Kirkpatrick.

Los niveles de la pirámide de Miller, de la base al vértice, son los siguientes (Fig. 1):

- Miller adjudica la base de la pirámide a los conocimientos en abstracto y la denomina ‘saber’ (*knows-knowledge*).
- El segundo nivel, ‘saber cómo’ (*knows how-competence*), también hace referencia a la parte cognitiva de la competencia, pero este saber es ahora contextualizado e incluye habilidades del tipo toma de decisiones y razonamiento clínico.
- El tercer nivel, ‘demostrar cómo’ (*shows how-performance*) da un salto cualitativo muy importante en la evaluación de la competencia clínica, ya que incluye el comportamiento (habilidades). No obstante, el contexto de aplicación de las competencias no es real. Es un nivel de competencia que se muestra en entornos simulados aunque contextualizados.
- Miller reserva el vértice de la pirámide para el ‘hacer’ (*does-action*). Se trata de la competencia demostrada en situaciones o contextos profesionales reales. Una vez más, subrayar que este nivel constituye el reto actual al que se enfrentan los educadores y, en especial, los especialistas en evaluación de la competencia profesional.

¿Para qué?

Una vez se ha definido el objeto de evaluación se debe clarificar el propósito. Existen diferentes objetivos de la evaluación para los distintos objetos seleccionados [5,6]. Dado que el artículo se centra en la evaluación del aprendizaje profesional se citan algunas posibilidades en la tabla II.



Programa de Formación Médica Acreditada en Educación Médica



Artículo: La evaluación en educación médica. Principios básicos

Autores: M. Nolla-Domenjó

Revista: Educ Med 2009; 12: 223-9

1. Las dos preguntas fundamentales sobre el sistema de evaluación que se deben plantear desde el inicio son:

- a) ¿qué? y ¿cómo?
- b) ¿cómo? y ¿cuándo?
- c) ¿qué? y ¿para qué?
- d) ¿para qué? y ¿para quién?
- e) ¿qué? y ¿cuándo?

2. ¿Cuál de los siguientes aspectos a evaluar no forma parte de los niveles del modelo de Donald Kirkpatrick?:

- a) impacto
- b) reacción
- c) aprendizaje
- d) calidad
- e) resultados

3. Los niveles de la pirámide de Miller son todos los siguientes excepto uno. Indique cuál es el falso:

- a) querer
- b) saber
- c) hacer
- d) demostrar cómo
- e) saber cómo

4. ¿Cuál de las siguientes características no es fundamental en un sistema de evaluación?:

- a) validez
- b) fiabilidad
- c) confianza
- d) transparencia
- e) factibilidad

5. En la actualidad, ¿cuál es la principal dificultad de la evaluación de la competencia clínica?:

- a) la aceptación por parte de los profesionales
- b) el coste del sistema de evaluación
- c) la carencia de medidas correctoras para los que no cumplen los mínimos requeridos
- d) la imposibilidad de medirla de forma directa
- e) la contextualización necesaria de los instrumentos

6. ¿Cuál de las siguientes evaluaciones supone el mayor reto en la actualidad en educación médica?:

- a) la evaluación de los conocimientos
- b) la evaluación de la práctica profesional

- c) la evaluación de las habilidades quirúrgicas básicas
- d) la evaluación de la satisfacción de los estudiantes
- e) la evaluación de la calidad docente

7. ¿Cuál de los niveles de la pirámide de Miller se evalúa mediante las evaluaciones clínicas objetivas estructuradas (ECOE)?:

- a) primero
- b) segundo
- c) tercero
- d) cuarto
- e) todos los niveles

8. ¿Cuál de los siguientes instrumentos evalúa más adecuadamente el cuarto nivel de la pirámide de Miller?:

- a) ECOE
- b) pruebas escritas basadas en contextos clínicos
- c) pacientes simulados
- d) observación directa
- e) pruebas escritas descontextualizadas

9. ¿Cuál de las siguientes definiciones se refiere al concepto de validez de un instrumento de evaluación?:

- a) grado en que un instrumento mide de forma consistente lo que se pretende medir
- b) conocimiento, por todas las partes implicadas, de las reglas del juego
- c) posibilidad de utilizarlo
- d) grado en que un instrumento mide lo que se quiere medir
- e) grado en que los candidatos aceptan la evaluación porque la consideran adecuada

10. Señalar la proposición incorrecta:

- a) una evaluación formativa tiene como objetivo fundamental proporcionar a los estudiantes *feedback* sobre su aprendizaje
- b) una evaluación sumativa es la base de la toma de decisiones sobre el nivel de competencia profesional del candidato
- c) una evaluación sumativa no puede ser en ningún caso formativa, de manera simultánea
- d) la decisión de emplear una forma de evaluación sumativa guiará la utilización de los resultados de la evaluación
- e) la evaluación guía el aprendizaje

Nombre y apellidos DNI.....
 Dirección Código postal.....
 Localidad Provincia/país.....
 Teléfono E-mail.....

Encontrará este cuestionario en la web de la revista: www.educmed.net

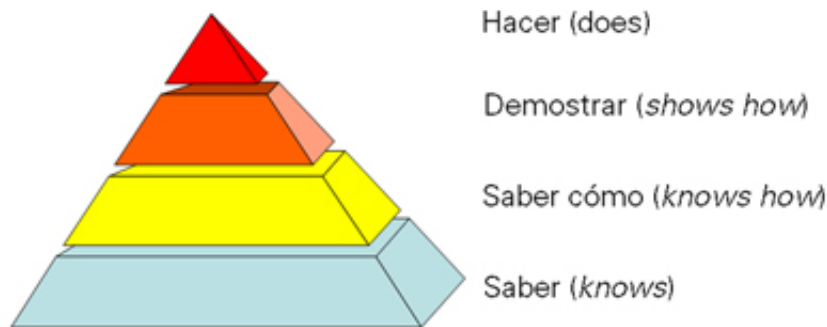


Figura 1. Pirámide de Miller (adaptado de [4]).

Sin embargo, a un nivel más general, además de clarificar el objetivo de la evaluación debe decidirse qué tipo de enfoque se va a dar y qué consecuencias van a tener los resultados de la evaluación.

La evaluación puede ser formativa o sumativa [7]. Se dice que una evaluación tiene finalidades formativas cuando su objetivo fundamental consiste en proporcionar a los estudiantes *feedback* sobre su aprendizaje y facilitar la reflexión sobre su progreso. Una evaluación tiene finalidades sumativas cuando los resultados son la base para la toma de decisiones sobre el nivel de competencia profesional (aprobado/suspenso, obtención de un título, obtención de una acreditación, etc.). No obstante, un sistema de evaluación sumativo puede contener un componente formativo cuando también se proporciona *feedback* a los estudiantes sobre sus fortalezas y sus debilidades.

La decisión de emplear una forma de evaluación sumativa, formativa o una combinación de ambas, guiará la selección de instrumentos, la manera en que se realizará la evaluación, la cantidad de recursos humanos necesarios, la interpretación de los resultados y la utilización de los resultados de la evaluación [8].

Instrumentos de evaluación

Una vez se han respondido las preguntas ¿qué? y ¿para qué?, se puede pasar al ¿cómo?, a la selección de los instrumentos.

Tabla II. Propósitos de la evaluación del aprendizaje.

- Asegurar que se cubren las necesidades de aprendizaje
- Identificar puntos débiles en la docencia
- Aprobar programas docentes
- Proporcionar *feedback* a los docentes y a la organización
- Seleccionar los mejores candidatos
- Etc.

La cantidad de instrumentos desarrollados y la finalidad de visión global del presente módulo hacen del todo imposible su descripción detallada, así como poder citar todas las herramientas disponibles. Un esquema global basado en la pirámide de Miller fue presentado por van der Vleuten [9] en el marco de la 9th International Ottawa Conference on Medical Education (Fig. 2):

- La base de la pirámide ('saber') se relaciona básicamente con pruebas escritas sobre conocimientos abstractos, es decir, descontextualizados.
- En el segundo nivel ('saber cómo') se realizan también pruebas escritas basadas en contextos clínicos, en casos concretos, y se pueden valorar competencias de razonamiento clínico y toma de decisiones.
- El tercer nivel de la pirámide ('demostrar') se relaciona con las pruebas basadas en simulaciones, de ahí que van der Vleuten las englobe



Figura 2. Relación entre nivel de competencia según la pirámide de Miller y los instrumentos susceptibles de ser utilizados (adaptado de [9]).

en la expresión ‘evaluación de la práctica *in vitro*’. A este nivel se encuentran los maniqués, los pacientes simulados y las evaluaciones objetivas clínicas estructuradas (ECOEs). Las ECOEs suponen la combinación de diversos instrumentos para la evaluación de la competencia clínica.

- El vértice de la pirámide (‘hacer’) incluye los instrumentos que permiten evaluar la práctica real, la ‘evaluación de la práctica *in vivo*’. La observación (directa o videograbada), el portafolio, la evaluación de registros clínicos y la utilización de pacientes simulados ocultos son algunos de los más utilizados.

A pesar de la aparición constante de nuevos instrumentos no hay que olvidar que no existe ninguno que, por sí sólo, pueda dar información sobre todos los componentes de la competencia profesional. La evaluación de la competencia global de un profesional siempre debe combinar distintos instrumentos.

Características fundamentales de un sistema de evaluación

Cuando se consideran los instrumentos de evaluación aparecen nuevas preguntas a formular-

se. Las características fundamentales a estudiar y definir son: validez, fiabilidad, transparencia, aceptación, factibilidad e impacto educativo [6].

Validez

La validez es el grado en que un instrumento mide lo que se quiere medir. Una cinta métrica es válida para medir una longitud, pero no para medir el peso. Aunque parezca una obviedad resulta extremadamente importante asegurar la coherencia entre lo que se quiere evaluar y el procedimiento seleccionado para ello [10].

Hay diferentes tipos de validez, pero las tres categorías principales son: validez de contenido, validez de criterio y validez de constructo [11]. La validez de contenido da información sobre en qué medida el instrumento mide el área de conocimiento que se pretende evaluar. La validez de criterio se refiere a la comparación de la medida con un valor conocido (criterio de referencia). La validez de constructo informa de en qué medida el instrumento mide un constructo hipotético (p. ej., la inteligencia, la empatía, la competencia profesional) que no es fácilmente observable.

Fiabilidad

La fiabilidad o precisión es el grado en que un instrumento mide de forma consistente lo que se

pretende medir. Tiene que ver con la repetición de la medida. En el ejemplo anterior de la cinta métrica, es fiable en la medida en que cada vez que se mide un mismo objeto, el resultado es el mismo o muy similar.

La fiabilidad generalmente se mide como un coeficiente de correlación, siendo 1,0 una correlación perfecta, y 0,0, una no correlación total. Los valores superiores a 0,7 se consideran indicativos de que el instrumento es fiable.

La fiabilidad entre evaluadores se refiere al grado de acuerdo entre evaluadores diferentes e independientes del mismo candidato.

Transparencia

La transparencia en un sistema de evaluación tiene relación con el conocimiento por parte de todas las partes implicadas de las 'reglas del juego' (tipo de prueba, criterios de evaluación, etc.).

Aceptación

La aceptación es el grado en que los candidatos aceptan la evaluación porque la consideran adecuada. A mayor validez y fiabilidad contrastadas y documentadas, mayor aceptación por parte de los candidatos. La transparencia del proceso de evaluación también contribuye a la aceptación de la prueba.

Factibilidad

En la selección de un instrumento es de capital importancia comprobar que será posible (factible) utilizarlo. Esto implica calcular el coste de la evaluación en términos de recursos (humanos y materiales) y de tiempo.

El coste debe incluir la puesta en marcha y los cursos necesarios para el desarrollo e implementación. El coste de un sistema de evaluación también debería considerarse en relación al beneficio para la docencia y el aprendizaje.

Impacto educativo

En el diseño de los sistemas de evaluación a menudo no se considera el impacto que puede tener sobre el estudiante. Sin embargo, hace tiempo que se ha observado que los procesos de evaluación dirigen el aprendizaje [12,13]. El tipo de evaluación determina cómo estudian los candidatos.

En 1977, Miller, en su participación en el II Simposio Médico Internacional en Roma, ya aseveró que 'cambiar el sistema de evaluación sin modificar el programa de enseñanza ha tenido una mayor repercusión sobre la naturaleza y la calidad del aprendizaje, que si se hubiera modificado el programa sin tocar los exámenes'.

Van der Vleuten [14] describió al menos cuatro vías por las que el sistema de evaluación dirige el aprendizaje: su contenido, su estructura o formato, lo que se pregunta (p. ej., la información que se proporciona), y la frecuencia, horario y número de evaluaciones repetidas.

Retos actuales de la evaluación en educación médica

Tomando como marco de referencia la pirámide de Miller, el mayor reto actual se sitúa en su parte superior. La evaluación de la competencia profesional en entornos simulados y, más aún, la evaluación de la práctica profesional real suponen un gran desafío porque son conceptos complejos que no se pueden medir de manera directa.

Shumway y Harden [11] consideran que hay que tender a evaluar lo que los aprendices hacen en la práctica y cómo aplican sus conocimientos en ciencias básicas y clínicas para el cuidado de los pacientes. Mientras que muchas facultades de medicina se han decantado por la evaluación de la competencia en situaciones simuladas, sólo unas pocas han puesto un énfasis mayor en la evaluación de la actuación en la práctica clínica a través de abordajes de observación directa o mediante el uso de medidas indirectas como los portafolios.

Según Brailovsky [15], la capacidad reflexiva y el razonamiento en la resolución de problemas clínicos representan el núcleo fundamental de la competencia clínica, y el desarrollo de instrumentos de evaluación que hagan posible la valoración de estos dos elementos cruciales, poco accesibles por los instrumentos de medición empleados tradicionalmente, se ha convertido en un complejo desafío.

Epstein [7] plantea otros retos además de los ya citados. Recientemente han aparecido nuevos dominios de evaluación, en especial debido a la necesidad de evaluar la competencia y la práctica profesional: trabajo en equipo, profesionalismo, comunicación, entre otros... Otro reto actual es

el desarrollo de sistemas de evaluación adecuados a las características especiales de los profesionales con años de experiencia.

Norcross et al [16] concluyen que el futuro de un sistema comprensivo de evaluación de los médicos asistenciales depende del desarrollo y estandarización de instrumentos, técnicas y procedimientos para la evaluación de la competencia y la actuación profesional, incluyendo medidas de la práctica, de programas de evaluación de colaboración en red, del control de costes, y del desarrollo continuo de medidas correctoras que se correspondan con los hallazgos en la evaluación.

Bibliografía

1. Palés J. Planificar un curriculum o un programa formativo. *Educ Med* 2006; 9: 59-65.
2. Kirkpatrick DL. Evaluación de acciones formativas. Barcelona: Gestión; 2000.
3. Kirkpatrick DL. Avaluació d'accions formatives: els quatre nivells. Barcelona: Epise Gestión; 2000.
4. Miller G. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65: S63-7.
5. Wall D. Evaluation: improving practice, influencing policy. Edinburgh: Association for the Study of Medical Education; 2007.
6. Morrison J. Evaluation. *Br Med J* 2003; 326: 385-7.
7. Epstein RM. Assessment in medical education. *N Engl J Med* 2007; 356: 387-96.
8. Dent JA, Harden RM. A practical guide for medical teachers. Edinburgh: Elsevier; 2005.
9. Van der Vleuten CPM. A paradigm shift in education: how to proceed with assessment? 9th International Ottawa Conference on Medical Education. Cape Town, 28 February-3 March 2000. URL: <http://www.educ.unimaas.nl/ottawa>.
10. Villa A, Poblete M. Aprendizaje basado en competencias. Bilbao: Universidad de Deusto; 2008.
11. Shumway JM, Harden RM. AMEE Guide no. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach* 2003; 25: 569-84.
12. Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983; 17: 165-71.
13. Wood D. Formative assessment. Edinburgh: Association for the Study of Medical Education; 2007.
14. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996; 1: 41-67.
15. Brailovsky CA. Educación médica, evaluación de las competencias. In OPS/OMS, eds. Aportes para un cambio curricular en Argentina. Buenos Aires: University Press; 2001. p. 103-20.
16. Norcross WA, Henzel TR, Freeman K, Milner-Mares J, Hawkins RE. Toward meeting the challenge of physician competence assessment: the University of California, San Diego Physician Assessment and Clinical Education (PACE) Program. *Acad Med* 2009; 84: 1008-14.