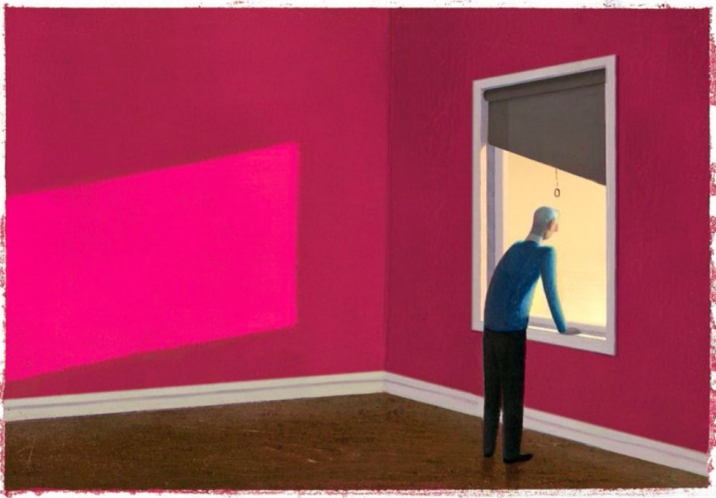


Diez Conceptos Claves sobre evaluación en educación médica basada en la competencia para clínicos docentes no familiarizados.

Resumen: En los últimos años parece que se están imponiendo nuevos enfoques dentro de la educación médica basada en la competencia. El objetivo de este artículo es ofrecer a los tutores y jefes de estudios que no creen estar familiarizados con los nuevos retos que supone la educación médica y evaluación basada en competencias, un vocabulario actualizado sobre los términos más frecuentemente utilizados así como su significado comúnmente admitido.

Ten Key Concepts on evaluation in competency based medical education for faculties who are unfamiliar.

Abstract: In recent years it seems that new approaches are being imposed within competency based medical education. The aim of this article is to offer tutors and educators leaders who do not think they are familiar with the new challenges of competency based medical education assessment, an updated vocabulary about the most frequently used terms, and their commonly accepted meaning.



La educación médica basada en la competencia (EMBC) debería empezar a cambiar la forma en que enseñamos y evaluamos a los residentes, sin embargo, muchos educadores consideran que el sistema actual hace imposible o muy difícil introducir algunos cambios en esta dirección y además se consideran mal preparados para actuar en este nuevo entorno. En los últimos años parece que se están imponiendo nuevos enfoques dentro de este campo de la EMBC, especialmente en algunos países de los que tradicionalmente han sido más punteros en el terreno de la educación médica, nos referimos a las actividades profesionales confiables (EPA) y a los hitos (hitos). La abrumadora tarea que supondría el desarrollarlos en nuestro contexto solo es superada por el desasosiego que nos produce a nosotros simplemente la necesidad de realizar algún cambio efectivo en esta dirección en nuestro país y la dificultad para incorporar alguna estrategia en la que esto podría empezar a cambiar, en un sistema dominado exclusivamente por una actuación puntual (el examen MIR) y que salvo tímidas y limitadas modificaciones (libro del residente, cierto desarrollo de las jefaturas de estudio), se ha mostrado casi inamovible en los últimos treinta años (el ejemplo de esta petrificación quizás sea el status de los clínicos tutores y el no desarrollo de sistemas de evaluación de la competencia del residente y remediación efectivos) .

El objetivo de este artículo es ofrecer a los tutores y jefes de estudios que no creen estar familiarizados con los nuevos retos que supone la educación médica y evaluación basada en competencias, un vocabulario actualizado sobre los términos más frecuentemente utilizados así como su significado comúnmente admitido. Las 10 "perlas de evaluación" prácticas que ofrecemos bajo esta perspectiva reflejan unos fundamentos basados en la evidencia que debería servir como una guía fácil para que estos tutores clínicos y responsables educativos las utilicen y las introduzcan en nuestro contexto. Proporcionamos también un glosario de términos en el cuadro para los que se aproximen por vez primera a este campo.

Glosario de términos

? **Comité de Competencia Clínica (CCC):** en el contexto de educación médica basada en competencias, el CCC es un comité que incluye miembros de la facultad que utilizarán una combinación de datos de evaluación recopilados de múltiples fuentes para evaluar el progreso de los estudiantes y hacer grandes apuestas decisiones.

? **Competencia (Competence):** una gama de habilidades que permite al aprendiz o al médico realizar todas las tareas de la práctica de manera efectiva y consistente. Se considera una construcción compleja.

? **Competencia (Competency):** una habilidad observable de un estudiante o médico. Ejemplo: realizar una exploración física completa y precisa.

? **Educación médica basada en competencias:** un enfoque para educar a los médicos que está orientado en las habilidades resultantes y organizadas en torno a las competencias. Desafía el un entrenamiento evolutivo en el tiempo y pretende más flexibilidad y estar centrado en el alumno (16).

? **Constructo:** una colección intangible de conceptos abstractos que se infieren del comportamiento (5). Por ejemplo, "competencia clínica" y "profesionalismo" son constructos que pueden ser de interés para evaluar, pero se deducen del comportamiento del alumno o residente en el lugar de trabajo.

? **Actividad Profesional Confiable (EPA):** Unidades de práctica profesional, definidas como tareas de la vida real esenciales para una especialidad en particular y que pueden confiarse a un aprendiz una vez que este ha alcanzado el nivel de competencia (3).

? **Ejemplo de un EPA:** "Gestionar la atención de pacientes con enfermedades en múltiples entornos clínicos " (17).

? **Evaluaciones de niveles de riesgo bajos y altos (Low&High Stakes):** las evaluaciones de bajo riesgo tienen consecuencias limitadas para el alumno en términos de promoción, selección o certificación, mientras que las evaluaciones de alto riesgo pueden tener consecuencias de gran alcance, como la no certificación.

? **Milestones (Hitos):** la capacidad esperada de un aprendiz en una etapa de su experiencia, a medida que pasa de novato a experto.

? **Ejemplo de exploración física de neurología:** el nivel 1 realiza una exploración neurológica completa hasta el nivel 4, realiza de manera eficiente una exploración neurológica relevante que incorpora con precisión todas las maniobras apropiadas adicionales (18).

? **Enfoque programático de la evaluación:** el uso de varios métodos de evaluación dispuestos longitudinalmente y contruidos deliberadamente para optimizar el aprendizaje y la evaluación. El programa incluiría varios puntos de datos de evaluación de bajo riesgo que se van agregando para decisiones de alto riesgo tipo aprobado / suspenso (11).

? **Rater Cognition (Proceso Cognitivo del Evaluador):** los procesos mentales que ocurren durante la calificación, ya sea a nivel consciente o inconsciente.

1. Todas las evaluaciones son muestras

No es posible evaluar todo lo que se espera que los residentes demuestren, por lo que debemos tomar muestras deliberadamente del conocimiento y las habilidades representativas utilizando un plan cuidadosamente construido (2). Un "blueprint" o plan evaluativo define lo que se está evaluando; también proporciona evidencia de que la estrategia de evaluación es válida al garantizar un muestreo suficiente y apropiado. Por ejemplo, una EPA para medicina interna es "admitir y gestionar un paciente hospitalizado con un nuevo problema agudo en medicina" (3,4). Se requieren varias competencias para que un residente complete esta EPA: conocimiento de ciencia básica, características clínicas y estrategias de gestión, así como habilidades de comunicación y la capacidad de realizar una exploración física enfocada apropiadamente. Para cada competencia, se requiere una estrategia de muestreo. Parece evidente que no se puede evaluar a un residente en todos los tipos de pacientes que requieran una exploración focalizada, pero sí se puede tomar una muestra deliberadamente, por ejemplo, en función del sistema principal involucrado, como el caso de un paciente con antecedentes respiratorios, cardiovasculares, o problemas musculoesqueléticos.

2. Cuanto más se juega el residente, más muestras se necesitan

Todas las mediciones tienen errores, y cuanto mayor es el riesgo (aprobar o suspender; pasar o no pasar), más evaluaciones (o muestras) se requieren (5). Por ejemplo, si el propósito de la evaluación es proporcionar feedback a los residentes sobre el manejo de pacientes ambulatorios, una única evaluación sería apropiada. Por otro lado, para que un Comité de Evaluación (CE) o de Competencia Clínica (CCC) tome decisiones sobre la promoción del alumno o residente (es decir, una evaluación para decisiones tipo aprueba (pasa de año)/ suspende (repite año) de mucha mayor importancia), entonces usar solo una evaluación no sería defendible. La pequeña muestra sobre el comportamiento observado es insuficiente para hacer una interpretación significativa del

nivel de desempeño del residente para una decisión tan importante. Definir el propósito (decisiones de alta importancia / baja importancia) de la evaluación puede ayudar a determinar cuántas muestras se necesitan.

3. Evalúa lo que es importante, no solo lo que es fácil

Con frecuencia, los educadores predeterminan evaluar lo que es fácil, en lugar de evaluar lo que es importante. Sabemos que usar un examen escrito para evaluar todos los aspectos de la competencia clínica no es adecuado. Los roles intrínsecos (expertos no médicos) en contextos como el CanMEDS 2015 (6) y las competencias del Consejo de Acreditación para la Educación Médica de Posgrado (ACGME) (7) son difíciles de evaluar (por ejemplo, profesionalismo), pero son elementos importantes de la competencia del médico. Se han desarrollado herramientas novedosas que ayudan a enfrentarse a estos desafíos, como feedback de múltiples fuentes para las habilidades de comunicación en el equipo comentarios narrativos para una atención centrada en el paciente, comunicación y profesionalismo (8) y algunos modelos de evaluación en el quirófano (9). Claramente, se ha progresado en el desarrollo de herramientas para evaluar las diferentes áreas de competencia, no solo las competencias más fáciles de evaluar, como el conocimiento médico. En nuestro país un modelo de evaluación de competencias que va en esta dirección, desarrollado en los últimos años en el contexto de la evaluación a residentes en ámbito hospitalario, es el modelo de "evaluación de competencias del Hospital de Cruces" (10)

4. Toda evaluación implica juicio

La objetividad perfecta y la estandarización no son posibles ni deseables. Teniendo en cuenta que para que se genere una calificación de una prueba, se debe observar un rendimiento y luego convertirlo en "nota" o "calificación", que luego deberá interpretarse: cada paso implica un juicio. A nivel individual, el juicio ocurre cuando el tutor (o evaluador) evalúa el desempeño de un residente en un contexto particular, y determina qué feedback proporcionar. La "Cognición del evaluador" (ver glosario) recibe una mayor atención como fortaleza, donde múltiples evaluadores ofrecen diferentes perspectivas y tienen el potencial de proporcionar datos más ricos y variados sobre el desempeño de un residente (11). Por otro lado, la excesiva variabilidad en las calificaciones de diferentes calificadores influye negativamente sobre la confiabilidad y la validez (12). Evaluadores que son formadores pueden mejorar la calidad de la evaluación, pero los resultados son mixtos (13). El juicio también se produce colectivamente cuando los miembros de un CE revisan todo el portafolio de un residente. El juicio en ese contexto determinará qué elementos del portafolio se deben proporcionar al comité para su revisión, y cómo ponderar los elementos individuales. Por ejemplo, el comité puede juzgar que un elemento del área de profesionalismo es tan nefasto que anularía evaluaciones excelentes de conocimiento retrasando así la promoción al siguiente año o etapa. Determinar el umbral o estándar para la promoción requiere claramente juicios colectivos y el consenso de los expertos (11,14).

5. Los métodos cuantitativos y cualitativos se complementan entre sí.

Los datos cuantitativos se han considerado tradicionalmente más deseables para su supuesta objetividad, aunque una limitación de los números es que no proporcionan datos específicos sobre cómo mejorar. Se ha demostrado que las narrativas capturan elementos de desempeño que una acumulación de números puede enmascarar (15); Además, en situaciones no estandarizadas, como la mayoría de las situaciones de evaluación en el lugar de trabajo, la narrativa proporciona datos mucho mejores para el feedback y el aprendizaje.

6. Ninguna herramienta de evaluación individual puede captar todos los aspectos de la competencia clínica

La competencia clínica es una construcción compleja que necesita un conjunto diverso de herramientas y estrategias de evaluación. Para ilustrarlo, la competencia implica conocimiento, que se puede evaluar mejor con exámenes escritos; las habilidades clínicas requieren observación directa, como un examen clínico objetivo estructurado (ECO); administrar pacientes en una unidad de hospitalización implica el uso de herramientas de evaluación basadas en el lugar de trabajo, como el feedback de 360 "grados"; y, por último, la evaluación de las habilidades de razonamiento diagnóstico puede requerir múltiples herramientas para la evaluación en el entorno clínico mediante el análisis retrospectivo de casos, y para eventos poco frecuentes utilizando simulaciones de casos basados "en ordenador" (19).

7. El feedback es un elemento esencial de evaluación

La evaluación debe informar a los estudiantes sobre cómo están progresando para convertirse en expertos, y el feedback formativo es una parte esencial de esto. Van der Vleuten y sus colegas (11,20), al describir la evaluación programática, indicaron que la mejor práctica de evaluación no solo es tratar de hacer las cosas lo suficientemente bien como para pasar un nivel de instrucción, sino también brindar una oportunidad de feedback formativo que contribuya a un mejor rendimiento.

8. La evaluación impulsa el aprendizaje

Los alumnos "estudiarán para la prueba" siempre que sea posible, enfocando sus estrategias de estudio sobre los conceptos que saben que saldrán en el examen. El aprendizaje basado en pruebas implica aprender mientras los alumnos se preparan para la prueba,

la completan y luego reciben comentarios. Los individuos evaluados sobre la materia retienen más en su memoria que aquellos que simplemente la estudiaron (21) De manera similar, la evaluación impulsa el aprendizaje en el lugar de trabajo: cuando los estudiantes ven que se evaluarán todos los aspectos de ser un médico, la importancia de dominar esas competencias expertas no médicas se vuelve mucho más claro. Por lo tanto, las estrategias de evaluación deben diseñarse teniendo esto en cuenta.

9. La validez es la característica más importante de los datos de evaluación

En pocas palabras, la validez es el criterio general del grado en que la teoría y la evidencia respaldan la interpretación de las calificaciones de una evaluación para un propósito específico (5,22). Si un residente obtiene una calificación perfecta en un examen de opción múltiple, ¿podemos concluir que el residente está listo para atender la primera llamada de todas las consultas que llegan a la medicina interna? Aunque una interpretación de la calificación de preguntas de opción múltiple de este alumno es que Sí, muchos argumentan que demostrar conocimiento no es suficiente. También necesitamos conocer las habilidades del residente en la toma de antecedentes, la exploración física, la gestión y las habilidades de procedimiento. Por lo tanto, la interpretación de la calificación de la prueba de opción múltiple como prueba de la disposición para atender la primera llamada no es válida: la evidencia (prueba de conocimiento) no respalda el propósito (competencia general). Esto resalta un concepto importante: que no existe una prueba válida o no válida. La validez siempre se refiere a lo apropiado de las inferencias o juicios basados ??en calificaciones de exámenes para un propósito específico. Está más allá del alcance de este artículo elaborar más este asunto, pero la validez es un concepto unitario y requiere múltiples fuentes de evidencia para apoyar o refutar la interpretación significativa de una calificación (5).

10. La evaluación perfecta es una ilusión

Hay muchos criterios que son relevantes para cualquier evaluación: validez, reproducibilidad, equivalencia, factibilidad, efecto educativo, efecto catalítico y aceptabilidad (23,24). En última instancia, la evaluación siempre implica cierto grado de compromiso. Considere una evaluación de bajo riesgo donde el propósito es proporcionar a los residentes datos de progreso y feedback. En este caso, la viabilidad (facilidad de administración), la aceptabilidad (para los residentes y la facultad), el efecto educativo (facilita el feedback) y el efecto catalítico (proporciona resultados que mejoran la educación) se considerarían todos importantes. Si esta fue una evaluación de alto riesgo con consecuencias significativas, entonces la reproducibilidad (prueba estadísticamente confiable) y la equivalencia (cada residente se somete al mismo tipo de prueba) sería primordial para defender los resultados.

Hemos delineado 10 perlas de evaluación que pueden ayudar a guiar a los tutores clínicos y diseñadores y responsables de programas docentes o jefes de estudio y unidades docentes encargados de modificar o transformar o hacer avanzar sus programas de evaluación en la línea de los nuevos requisitos de EMBC. Los conceptos importantes en EMBC tendrán implicaciones directas para la evaluación: Si se apuesta por incluir los EPA esto exigirá evaluaciones más rigurosas en el lugar de trabajo; incorporar los hitos requerirán una observación directa continua y feedback, así como un programa continuado de evaluación; y los CE tendrán la tarea de desafiados integrar todos estos datos para tomar decisiones evaluativas sobre los residentes. Esperamos que estos principios básicos puedan servir como guías para facilitar la transición tan necesaria como inaplazable que requiere nuestra formación de residentes en particular y educación médica en general.

Referencias

1. Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676-682.
2. Hamdy H. Blueprinting for the assessment of healthcare professionals. *Clin Teach*. 2006;3(3):175-179.
3. Hauer KE, Kohlwes J, Cornett P, et al. Identifying entrustable professional activities in internal medicine training. *J Grad Med Educ*. 2013;5(1):54-59.
4. ten Cate O, Chen HC, Hoff RG, et al. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach*. 2015;37(11):983-1002.
5. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837.
6. Frank JR, Snell L, eds. The Draft CanMEDS 2015 Physician Competency Framework. Royal College of Physicians and Surgeons of Canada. http://www.royalcollege.ca/portal/page/portal/rc/common/documents/canmeds/framework/framework_series_1_e.pdf. Accessed October 4, 2017.
7. Educational Commission for Foreign Medical Graduates. ACGME core competencies. <http://www>.

ecfmg.org/echo/acgme-core-competencies.html. Accessed October 4, 2017

8. Donnon T, al Ansari A, Al Alawi S, et al. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med.* 2014;89(3):511?516.
9. Gofton WT, Dudek NL, Wood TJ, et al. The Ottawa Surgical Competency Operating Room Evaluation (OSCORE): a tool to assess surgical competence. *Acad Med.* 2012;87(10):1401?1407.
10. Morán Barrios J. La evaluación del proceso de formación especializada, una propuesta fundamentada en el proyecto 'Formación especializada basada en competencias del Hospital de Cruces 2008'. *Educ Med* 2012; 15 (Supl 1): S1-S54.
11. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205?214.
12. Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: rethinking the etiology of rater errors. *Acad Med.* 2011;86(suppl 10):1?7.
13. Lee V, Brain K, Martin J. Factors influencing mini-CEX rater judgments and their practical implications: a systematic literature review. *Acad Med.* 2017;92(6):880?887.
14. Hauer KE, ten Cate O, Boscardin CK, et al. Ensuring resident competence: a narrative review of the literature on group decision making to inform the work of clinical competency committees. *J Grad Med Educ.* 2016;8(2):156?164.
15. Cook DA, Kuper A, Hatala R, et al. When assessment data are words: validity evidence for qualitative educational assessments. *Acad Med.* 2017;91(10):1359?1369.
16. Frank JR, Mungroo R, Ahmad Y, et al. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach.* 2010;32(8):631?637.
17. Caverzagie KJ, Cooney TG, Hemmer PA, et al. The development of entrustable professional activities for internal medicine residency training: a report from the Education Redesign Committee of the Alliance for Academic Internal Medicine. *Acad Med.* 2015;90(4):479?484.
18. Leep Hunderfund AN, Reed DA, Starr SR, et al. Ways to write a milestone: approaches to operationalizing the development of competence in graduate medical education. *Acad Med.* 2017;92(9):1328?1334.
19. Ilgen JS, Humbert A, Kuhn G, et al. Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. *Acad Emerg Med.* 2012;19(12):1454?1461.
20. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. 12 tips for programmatic assessment [published online ahead of print November 20, 2014]. *Med Teach.* 2014;1?6. doi:10.3109/0142159X.2014.973388.
21. Larsen DP, Butler AC, Roediger HL III. Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ.* 2013;47(7):674?682.
22. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 2006;119(2):166.e7?166.e16.
23. Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):206?214.
24. van der Vleuten CP. The assessment of professional competence: developments, research, and practical implications. *Adv Health Sci Educ.* 1996;1(1):41?67.