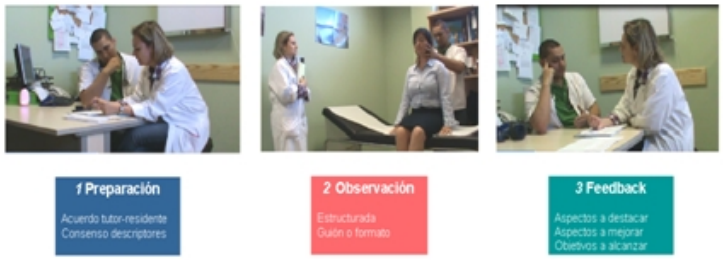
This page was exported from - DocTutor
Export date: Sat Nov 7 20:45:16 2026 / +0000 GMT

¿Queremos (podemos) realmente evaluar la residencia? Retos para un planteamiento riguroso del tema



Doctutor

Resumen: La puesta en práctica de una evaluación sumativa rigurosa de residentes (certificación final del título de especialista) requerirá del uso de una serie de instrumentos y exigirá ciertos requisitos a considerar para hacerlos operativos y eficaces y que tienen que ver sobre todo con su validez y su fiabilidad. Mientras esto no se haga así las instituciones educativas (Ministerio de Sanidad) no podrán certificar que la formación de nuestros residentes es acorde a los estándares de calidad que se plantean, cualesquiera que estos sean.

Si el Sistema Nacional de Formación de Especialistas pretendiera finalmente realizar una evaluación de los residentes con repercusiones reales sobre su formación y sobre todo sobre su certificación final, sería necesario incorporar una serie de instrumentos evaluativos y unos condicionantes de uso de estos para que la aplicación de sus resultados (especialmente en el ámbito de la evaluación sumativa) sea posible y se reduzca al máximo la arbitrariedad y la diversidad. Dad la importancia de este debate Doctutor dará la bienvenida a artículos relacionados con este tema.

Instrumentos evaluativos

Al estar hablando de especialistas MIR, su evaluación final no puede ser realizada utilizando exclusivamente una prueba de conocimientos/razonamiento clínico escrita mediante el uso de diferentes tipos de preguntas pero sobre todo de preguntas de respuesta múltiple y mucho menos confiar exclusivamente (o en un porcentaje muy elevado) en las evaluaciones que hacen los clínicos responsables al final de sus rotaciones. La evaluación del ?mostrar como? o ?saber hacer? en contextos de práctica real o simulada es imprescindible. En la mayoría de los países que realizan este tipo de evaluaciones (especialmente en el Reino Unido) los instrumentos más usados son todos los siguientes o una mezcla de algunos:

- Mini-CEX (ejercicio de evaluación clínica)
- DOPS (observación directa de habilidades procedimentales
- Feedback 360º -también conocido como ?mini-PAT? o ?TAB?(herramienta de evaluación por colegas) (evaluación de conductas por miembros del equipo)
- Discusión de casos (DC)
- Informe del tutor clínico
- Presentación de evidencias de aprendizaje, incluyendo portafolio reflexivo y plan de desarrollo personal.

Formación de Evaluadores

Un aspecto esencial en este proceso evaluativo con enfoque certificativo es la necesidad de disponer de un cuadro amplio y bien formado de evaluadores. En los lugares donde esto se lleva a cabo consideran que entrenar a los evaluadores sobre la forma de utilizar los instrumentos y hacer esto mediante una formación presencial maximiza la validez y la fiabilidad de las evaluaciones, por lo que esto es considerado como un elemento de ?extremada importancia?. El entrenamiento de este tipo en el Reino Unido consiste en:

- Formación en Observación (como observar y planillas a usar para esto)
- Formación en la dimensión práctica (si evaluar conocimientos, habilidades, juicio clínico o profesionalismo
- Formación en el marco de referencia (sobre los acuerdos relacionados con el estándar para el nivel de apto al observar al residente)

Los problemas que se han detectado en el curso de estas evaluaciones son: que muchos evaluadores realmente no observan, que son muy condescendientes o por el contrario muy duros, que no discriminan la actuación de un residente en diferentes dominios, o que ellos mismos no poseen las habilidades que precisamente están observando.

Tomar una decisión: sintetizar la evidencia de una evaluación de la práctica real

Para tomar una decisión evaluativa final sobre un residente se aconseja comprender bien el programa de evaluación que se lleve a efecto como algo en conjunto. Es decir se exige que se de un ?perfil de la evaluación? del residente para determinar si lo que este ha hecho es o no satisfactorio. El hacer un perfil de este tipo requiere hacer triangulaciones sobre la evidencia de sus ?evaluaciones de la práctica? disponibles ( ), a lo que se sigue de la emisión de un juicio experto basado en esto y en el que el supervisor educativo (responsable docente o tutor) tiene un papel importante.

La evidencia en la mayoría de los programas evaluativos contiene tanto información cuantitativa (puntuaciones) como cualitativa (texto libre), esto aumenta el valor de los juicios genéricos al dar más riqueza de información. La evaluación de un portafolio es en muchos casos algo central en este proceso, pero no es lo único

¿Qué vamos a hacer si no estamos seguros sobre si aprobar o no a este residente (o dicho de otro modo: si su ?actuación? ha sido satisfactoria o no)?

Exponemos a continuación un ejemplo práctico de la complejidad y rigurosidad que exige este proceso. Para ello traemos una situación frecuente en cualquier evaluación.

Imaginemos que hemos decidido incorporar en la evaluación las herramientas antes mencionadas. Una duda frecuente con muchos residentes, seguramente la más frecuente será la siguiente:

El número mínimo de evaluaciones requeridas en un programa evaluativo se debe de basar en la confianza que podamos tener en el resultado final (ver la tabla 1). Para la mayoría de los residentes su puntuación final agregada puede colocarse con diferente confianza en uno u otro lado del espectro (satisfactorio/insatisfactorio). Sin embargo un residente ?límite? puede necesitar más evaluaciones. Estas evaluaciones adicionales pueden usarse para describir la naturaleza de los problemas de una forma más clara.

Tabla 1. IC 95% de la puntuación media de un residente según el número de pruebas que contribuyen a esa evaluación

casos	**Mini-CEX**	**DOPS**	**360º**	**DC**
4				
	0,55	0,59	0,57	0,55
6				
	0,45	0,48		

0,47
0,45

8
0,39
0,42
0,40
0,39

12
0,32
0,34
0,33
0,32

Existen diferentes formas en las que las pruebas extras que se hagan aumentarán la fortaleza de las decisiones que finalmente se tomen. Conforme aumenta el número de evaluaciones (pruebas), los Intervalos de Confianza (IC) se estrechan. Por ejemplo, para el mini-CEX el IC del 95% para 4 casos y con una escala de 6 puntos donde el punto de corte sea 4, es de 0,55 pero para 12 es de 0,32. En otras palabras, un residente con una media global de 4,33 sería ?apto? si lo adquiere con 12 casos (o pruebas) (IC: 4,02-4,64) pero no tendríamos una confianza del 95% de que ese residente tuviese el ?apto? solo con 4 casos (IC: 3,77-4,89).
Muestras adicionales permitirían tener una perspectiva cualitativa mayor de las ?actuaciones? de los residentes, posibilitando así una mayor comprensión del tipo de problema que tiene el residente y proporcionando oportunidades adicionales de identificar aspectos que deberían mejorarse.
Donde surjan este tipo de problemas en el contexto de este tipo de evaluación, puede entonces ser apropiado utilizar elementos evaluativos alternativos que describan cual es el problema con más detalles. Por ejemplo: si con un residente tenemos problemas sobre su capacidad para comunicar con los pacientes, entonces lo que habría que hacer sería usar el vídeo para evaluar y analizar sus habilidades de comunicación.