



AULA DE EDUCACIÓN MÉDICA

La evaluación del desempeño o de las competencias en la práctica clínica. 2.^a parte: tipos de formularios, diseño, errores en su uso, principios y planificación de la evaluación



Jesús Morán-Barrios

Unidad de Docencia Médica (Formación Sanitaria Especializada), Hospital Universitario Cruces, Barakaldo, Vizcaya, España

Disponible en Internet el 16 de noviembre de 2016

PALABRAS CLAVE

Evaluación;
Competencias;
Principios;
Planificación;
Listas de cotejo;
Modelos de escala;
Rúbricas

RESUMEN Las listas de verificación y métodos de escala son los instrumentos que se emplean en la mayoría de los métodos de evaluación. Los métodos de escala verbales o descriptivos reducen la subjetividad, siendo las rúbricas los más recomendables y el mejor instrumento de información sobre el grado y progreso del desempeño. Los errores y la subjetividad en el uso de formularios se evitan con un adecuado diseño, detallada descripción de las competencias y escalas y con la formación y colaboración de varios evaluadores.

© 2016 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Assessment;
Competencies;
Principles;
Blueprint;
Checklist;
Rating scales;
Rubrics

Assessment of clinical performance. Part 2. Questionnaire, design, errors. Principles and assessment planning (blueprint)

ABSTRACT Checklists and rating scales are the instruments used in the majority assessment methods. Descriptive rating scales reduce subjectivity, being rubrics the most suitable and the best tool to know the progress of learning. Errors and subjectivity in the use forms can be avoided with proper design, detailed description of the competencies and scales and training and collaboration of several evaluators.

© 2016 Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

En la primera parte de esta aula docente se fundamentó la evaluación de las competencias y se analizaron los

diferentes métodos de evaluación del desempeño. La mayoría de estos métodos utilizan 2 tipos de instrumentos: listas de verificación o de cotejo (*checklist*) y escalas de valoración (*rating scales*), también llamados métodos de escala. En esta segunda parte exponemos las claves de su diseño, los errores en su uso y cómo evitarlos, cómo minimizar la subjetividad y cómo planificar la evaluación.

Correo electrónico: jmoranbarrios25@gmail.com

<http://dx.doi.org/10.1016/j.edumed.2016.09.003>

1575-1813/© 2016 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Cirugía Plástica. Competencia: mejora de la calidad y seguridad de los cuidados.			
Descripción: El residente trabaja activamente por la calidad y seguridad de cuidados, es consciente de sus limitaciones, consulta adecuadamente, actúa bajo los más estrictos principios éticos, aplica adecuadamente los recursos asistenciales y protocolos del centro sanitario y promueve la cultura de la seguridad entre su equipo.			
Residente:	fecha		
Evaluador:	Residente	Facultativo	Personal sanitario
Feedback: sí/no			
Problema:			
Breve descripción de la complicación:			
Cuidados del paciente	SÍ	NO	NA
1. Obtiene y documenta apropiadamente el Consentimiento Informado			
2. Trata adecuadamente el problema inicial del paciente			
3. Maneja apropiadamente la complicación			
Conocimientos médicos			
4. Identifica la causa de la morbilidad/mortalidad			
5. Aplica las opciones terapéuticas del problema original del paciente			
6. Aplica las opciones terapéuticas a las complicaciones resultantes			
Práctica basada en sistemas			
7. Demuestra un entendimiento correcto del problema y aplica los recursos disponibles para prestar unos cuidados óptimos al paciente			
8. Aplica las estrategias terapéuticas basadas en la mejor evidencia siendo consciente de los costes.			
9. Informa al Comité de Calidad/Seguridad para la mejora de las prácticas médicas			
Práctica basada en el aprendizaje y la mejora			
10. Revisa la literatura y la evidencia científica sobre esta complicación.			
11. Sugiere modificar prácticas profesionales para evitar que vuelva a ocurrir			
Profesionalismo			
12. Revela la información adecuada al equipo y paciente			
13. Documenta adecuadamente los aspectos ético-legales			
Habilidades de comunicación y relación interpersonal			
14. Aporta propuestas de mejora de manera precisa y atractiva			
15. Presenta el problema en sesión clínica			
Comentarios: tareas para mejorar la seguridad y calidad de cuidados			
Valoración global:	<input type="checkbox"/> No es fiable, precisa supervisión permanente <input type="checkbox"/> Parcialmente fiable, precisa supervisión periódica <input type="checkbox"/> Totalmente fiable, puede trabajar sin supervisión		

Figura 1 Lista de verificación.

Modificado de: The plastic surgery milestone project: Assessment tools [consultado 1 Sep 2016]. Disponible en: <http://www.acgme.org/Portals/0/PDFs/Milestones/PlasticSurgeryAssessments.pdf>

Además, proponemos un guion básico para la evaluación en la formación especializada.

Listas de verificación o de cotejo (*checklist*)

¿Qué son?

Consisten en un listado de criterios o aspectos dirigidos a identificar la presencia o ausencia de una serie de conductas del profesional, para establecer si se ha alcanzado o no el nivel de desempeño requerido. Un ejemplo de ello se ilustra en la [figura 1](#).

¿Para qué se usan?

Para comprobar la presencia o ausencia de una serie de características predefinidas del desempeño. Se pueden construir a la medida de nuestras necesidades. Proporcionan una información precisa del desempeño de actividades o tareas concretas. Son muy útiles para evaluar destrezas psicomotrices (técnicas o procedimientos más o menos complejos), competencias relacionadas con la comunicación (con el paciente, la familia o el equipo de trabajo, calidad de una sesión presentada) o *audits* de la calidad de

la historia clínica, de gestión de recursos, etc. Sirve tanto para la evaluación formativa, sumativa (si se asignan valores a cada criterio) como para la autoevaluación; en este caso, transformando el tiempo del verbo de acción de tercera a primera persona: «informa al paciente.../informo al paciente...».

Diseño básico y contenido ([tabla 1](#))

1. *Propósito y competencia a evaluar*. Se define la tarea o actividad (biopsia renal, ecografía abdominal, seguridad del paciente, etc.) y se describe la competencia a evaluar en esa tarea (por ejemplo «indica, prepara y realiza la biopsia renal adecuadamente acorde al protocolo establecido, garantizando la seguridad del paciente e informando con claridad al paciente y a sus familiares de la necesidad de la misma»). A continuación se identifica un listado de posibles indicadores, criterios, aspectos o aseveraciones a valorar en esa competencia (por ejemplo preparación del material, identificación de la anatomía, consentimiento informado, etc.). Posteriormente se redactan los criterios y finalmente se construye un formulario: criterios en filas y escala en columnas.
2. *Criterios*. Se numera cada criterio y se describen con claridad para evitar al máximo la subjetividad. Emplear

frases cortas que se inician por un verbo de acción como: informa, maneja, coordina, trata, delega, orienta, proporciona, lidera, avisa, etc. Nunca utilizar el verbo «saber o conocer», pues no es una acción observable, ya que podemos comprobar si tiene conocimientos en la acción con el verbo «identifica», «analiza» o «aplica». Cada indicador debe incluir un solo aspecto a evaluar y han de ordenarse en una secuencia lógica (por ejemplo no mezclar descripciones técnicas con aspectos éticos). La lista no debe ser excesivamente larga, cuanto más desagregados los criterios menor validez y fiabilidad; con más de 20 criterios aparecen continuas repeticiones (no discrimina) y además agota al evaluador generando resistencias.

3. **Categorías.** Los criterios se pueden agrupar, por afinidad, en categorías. Un ejemplo es el expuesto en la [figura 1](#): «cuidados del paciente», «conocimientos», etc. En este ejemplo las categorías se corresponden al marco general de competencias del ACGME *Outcome Project* (EE. UU.).
4. **Escala.** Generalmente de 2 niveles: «cumplido/no cumplido», «correcto/incorrecto», «presente/no presente», «sí/no». Se puede incluir un tercer nivel, «parcialmente cumplido», y añadir siempre el nivel «no aplicable». La escala es solo cualitativa, pero se le pueden añadir valores a cada nivel (lista de cotejo con valores), en función de la importancia de esos criterios en el resultado final de la tarea; esto permite establecer puntos de corte en el nivel de desempeño del evaluado: no confiable, novato, aprendiz, competente, experto.
5. **Identificación de evaluador y evaluado, comentarios y valoración global final.** Añadir un espacio para cada uno de estos aspectos y la fecha. La valoración final puede estar basada en criterios de:
 - Supervisión: en casi todos los casos, en pocos casos, para casos complejos, solo ayuda ocasional del personal facultativo, es capaz de actuar independientemente.
 - Autonomía: solo observa, lo hace con tutorización directa, lo hace solo con mínimo apoyo, lo hace sin supervisión y maneja las complicaciones.
 - Destreza: novato, aprendiz, competente, experto.
6. **Borrador preliminar** elaborado tras revisar: guías de práctica clínica, literatura y datos de la experiencia clínica del trabajo a evaluar.
7. **Borrador final** tras la revisión por expertos (por ejemplo por el método Delphi).
8. **Prueba piloto** y test psicométricos (si se consideran necesarios) y obtención el *formulario definitivo* tras la nueva revisión por expertos.
9. **Instrucciones.** Han de ser breves para facilitar su comprensión y correcta cumplimentación.

Ventajas

- Altas propiedades psicométricas con evaluadores entrenados.
- La validez se asegura con el consenso de los expertos, la redacción de los criterios y su secuencia.
- Método económico, fácil de administrar y de estandarizar.
- Permite discriminar el nivel de competencia.

Tabla 1 Recomendaciones para la elaboración de formularios

Propósito. Proceso/actividad/tarea a evaluar y descripción general de las competencias en dicha actividad
Criterios. Definidos en términos de comportamientos observables, con lenguaje claro, específico y empleando verbos de acción de inicio: informa, maneja, coordina, trata, delega, orienta, proporciona, lidera, avisa, gestiona, etc.
Agrupar los criterios por categorías o dominios competenciales
Escalas de valoración. Listas de cotejo: 2 a 3 niveles. Métodos de escala: mejor escalas pares (sin un centro neutro), verbales con o sin números y preferiblemente rúbricas
Identificación, fecha, contexto de trabajo y su dificultad y espacio para comentarios
Valoración global final en función de: grado de autonomía, confianza, destreza, etc.
Borrador preliminar tras revisar guías de práctica clínica, literatura y datos de la experiencia clínica
Borrador final mediante revisión por expertos (por ejemplo por el método Delphi)
Prueba piloto y test psicométricos (si se consideran necesarios)
Formulario definitivo tras nueva revisión por expertos
Redactar unas breves instrucciones de uso

Desventajas

- Interpretación equivocada de algunos criterios si no están consensuados.
- Los niveles de escala (2 o 3) no permiten matizaciones.
- No proporcionan una información resumida sobre el nivel del desempeño, aunque esto se resuelve con una valoración global al final del documento.

Métodos de escala o *rating scales*

¿Qué son?

Al igual que las listas de cotejo, consisten en una relación de criterios sobre las competencias a evaluar, pero con una importante diferencia, ya que permiten graduar la presencia o ausencia de una determinada conducta, no con un sí/no, sino con la variabilidad del comportamiento. Las listas de cotejo se limitan a verificar, mientras que los modelos de escala indican el grado alcanzado en cada uno de los criterios o competencias evaluadas ([fig. 2](#)).

¿Para qué se usan?

La mayoría de los métodos descritos en la 1.^a parte de esta «aula docente» (Mini-CEX, DOPs, ACAT, 360°, evaluación de rotaciones del residente, etc.) utilizan modelos de escala. Estos formularios facilitan la evaluación de actividades o tareas más complejas en diferentes contextos laborales, donde se expresan simultáneamente competencias relacionadas con ética, comunicación, trabajo en equipo, gestión,

Ejemplo para "Información y comunicación con el paciente" ante un procedimiento						
Criterios/Competencias	Tipos de escalas					
<p>1. <i>Información y comunicación con el paciente:</i> Informa al paciente en un lugar privado, de forma clara, correcta y comprensible sobre el procedimiento a realizar, acorde a su nivel cultural y respetando sus creencias.</p> <p>2. <i>Manejo del Consentimiento Informado (C.I.):</i> Entrega el C.I. responde a las preguntas, asegura que lo entiende y solicita su conformidad.</p>	Escala gráfica/visual					
	<p>Nunca ←————→ Siempre</p>					
	Escala numérica					
	1	2	3	4	5	6
	Escala verbal					
	<p>Desempeño pobre Desempeño débil Desempeño bueno Desempeño destacado</p>					
Escala verbal con descriptor adicional y combinada con números						
	<p>1. <i>Muy por debajo de lo esperado</i> No muestra ningún interés por la buena comunicación.</p>	<p>2. <i>Por debajo de lo esperado</i> Muestra interés por la buena comunicación, pero tiene determinadas dificultades.</p>	<p>3. <i>Lo esperado</i> Demuestra lo esperado para su nivel de formación, tal como se indica en el descriptor de la competencia.</p>	<p>4. <i>Por encima de lo esperado</i> Demuestra un desempeño por encima de su nivel y trasmite gran confianza al paciente.</p>		
Competencia	Rúbrica analítica					
Información y comunicación con el paciente	<i>Muy por debajo de lo esperado</i>	<i>Por debajo de lo esperado</i>	<i>Lo esperado</i>	<i>Por encima de lo esperado</i>		
	Utiliza un lenguaje técnico, sin respeto a las creencias del paciente, descuidando la privacidad.	Respeto la privacidad, pero muestra dificultades en la claridad de la comunicación.	Informa al paciente en un lugar privado, de forma clara, correcta y sin tecnicismos sobre el procedimiento a realizar, acorde a su nivel cultural y respetando sus creencias.	Además de lo anterior, empatiza con facilidad y trasmite mucha confianza al paciente. Muestra disposición para la gestión de problemas adicionales que plantee el paciente.		
Manejo del Consentimiento Informado	Emplea el C.I. como un trámite administrativo, sin asegurarse de que el paciente ha entendido.	Explica confusamente el sentido del C.I. y no se asegura que lo ha entendido.	Entrega el C. I., responde a las preguntas, asegura que lo entiende y solicita su conformidad.	Además de lo anterior muestra una clara disposición para la gestión de problemas adicionales que plantee el paciente y ante dudas posteriores que tenga el paciente tras leer el C.I. detalladamente.		

Figura 2 Tipos de escalas para la valoración del desempeño.

cuidados del paciente, conocimiento, etc. Estos instrumentos se han desarrollado en un intento de lograr mayor objetividad y mejorar la fiabilidad de la evaluación.

Diseño y contenido

El proceso es el mismo que en las listas de cotejo (tabla 1). El contenido básico del formulario incluye en filas las competencias a evaluar y en columnas la escala de valoración. Los diseños pueden variar en función de la claridad visual (ver los numerosos ejemplos referenciados en la parte 1.ª de esta aula docente).

1. *Propósito y competencia general a evaluar.* Ejemplo: «dar malas noticias». Definición: «el residente demuestra

que informa adecuadamente a las familias sobre malas noticias, con claridad, respeto y empatía, teniendo en cuenta el nivel cultural y las creencias, tanto religiosas como culturales, dando respuesta adecuada a las necesidades que la familia le plantee». El propósito puede ser más amplio que una tarea concreta, como una rotación por una unidad o un proceso asistencial (cardiología, atención domiciliaria, hemodiálisis, etc.).

2. *Criterios.* Como en las listas de cotejo, deben definirse en términos de comportamientos observables con frases descriptivas amplias de lenguaje claro, específico, fácilmente entendible por el residente y el evaluador y utilizando verbos de acción. La interpretación de la competencia nunca debe quedar abierta a la opinión del observador con descriptores tan simples como: «valore la

información y comunicación con el paciente» sin mayor explicación. Se invita al lector a ver frases en los criterios de la [figura 2](#) y en el siguiente ejemplo para la tarea «dar malas noticias». Criterios:

- Lugar: «informa a los familiares en un lugar privado, cómodo, luminoso, sin interrupciones ni barreras físicas».
 - Información verbal: «se presenta y transmite la información con concisión, sin tecnicismos, con un lenguaje sencillo adaptado al nivel cultural de los familiares, asegurando que lo entienden y respetando los silencios».
 - El número de criterios ha de reflejar el conjunto de la actividad que se evalúa, evitando listas excesivamente largas y complejas, ya que generan resistencias.
3. *Escala*. Puede ser gráfica, numérica, verbal, mixta y tipo rúbrica (ver el siguiente apartado). Se trata de identificar el grado alcanzado en el desempeño, minimizar la subjetividad y elevar la consistencia de la evaluación. Existen discrepancias en cuanto al número óptimo de niveles, pero hay acuerdo entre un mínimo de 3 y no mayor de 10, si bien con no más de 7 es suficiente. Es preferible diseñar primero pocos niveles y aumentarlos si se comprueba que no discriminan. Son preferibles escalas pares (4 o 6 niveles), con un nivel positivo y otro negativo, sin un centro neutro (ni bien-ni mal), para evitar la tendencia central del evaluador no comprometido. Generalmente se mantiene la simetría en estas escalas, 2 o 3 niveles deficientes y 2 o 3 satisfactorios, obligando al evaluador a decantarse. Pero si el objetivo es detectar áreas de mejora la escala debería tener más niveles por debajo del «satisfactorio» y a la inversa si se pretende detectar principalmente excelencias, pero sin términos neutros. En evaluación formativa interesa detectar áreas de mejora. Ejemplo:
 - Claramente menor de lo esperado (tiene muchas áreas de mejora).
 - Por debajo de lo esperado (varias áreas de mejora).
 - En el límite de lo esperado (pocas áreas de mejora).
 - Lo esperado, satisfactorio o adecuado.
 - Por encima de lo esperado (destaca claramente).
 Las escalas deben ser cualitativas, procurando olvidar los números, ya que lo importante es cualificar conductas y no cuantificar. Los números son útiles si simplemente sirven para identificar de qué nivel de la escala hablamos de cara a la gestión de resultados y no para hacer cálculos matemáticos.
 4. *Categorías*. Como en las listas de verificación, los criterios pueden estar agrupados en categorías. Los *Global Rating Scales*, descritos en la 1.^a parte, están categorizados por grandes dominios competenciales. Por ejemplo: <http://www.hospitalcruces.com/universitario/DocPresentacion.asp?lng=es> (pestaña evaluaciones; ver «formulario evaluación rotaciones»).
 5. El resto de recomendaciones de diseño es como en las listas de cotejo.

Tipos de escalas ([fig. 2](#))

1. *Escalas gráficas* (tipo analógica visual). Se emplea una línea continua con 2 descriptores en los extremos: por ejemplo «rara vez-siempre». Implican una gran subjetividad, pero son muy útiles en formularios para evaluar percepciones sobre humanismo o profesionalismo por pacientes. Se redactan con frases cortas: «confío en mi médico», «me trata como a un número».
2. *Escalas numéricas*. Intentan establecer el nivel de competencia observado. El significado de los extremos se puede definir con una expresión verbal: «totalmente de acuerdo-totalmente en desacuerdo». Hay que evitar siempre el cero, porque es improbable que un residente merezca tal valoración y genera conflictos.
3. *Escalas verbales o descriptivas*. Son las escalas que mejor minimizan la subjetividad y las más recomendables. Los números son sustituidos por expresiones verbales que reflejan el menor o mayor grado alcanzado en el desempeño. Se redactan en términos neutros sin términos peyorativos como: «deficiente, muy mal o inaceptable». Son útiles para el evaluado, ya que concreta el estándar deseado de desempeño con la palabra «satisfactorio» o «lo esperado» para su nivel. Tipos de expresiones en las escalas verbales:
 - En función del resultado: menor de lo esperado, límite de lo esperado, lo esperado, por encima de lo esperado.
 - Según el nivel de destreza: novato, aprendiz, competente, experto.
 - En función del comportamiento: muy preocupante, preocupante, nada preocupante.
 - Por la frecuencia: nunca, a veces, frecuentemente, siempre.
 - Por la cantidad: nada, poco, bastante, mucho.
 - En función la percepción: de acuerdo, en desacuerdo.
4. *Rúbricas*
 - *¿Qué son?* Son el método más preciso y el ideal para minimizar al máximo la subjetividad de los evaluadores haciendo la evaluación más simple, transparente y justa. Se trata de un tipo expandido de escala verbal en la que cada nivel de desempeño tiene un descriptor cualitativo específico ([fig. 2](#)). Proporcionan descripciones claras del comportamiento asociado a cada nivel, por lo que delimita claramente el nivel de desarrollo o ejecución de una tarea.
 - *Utilidad*. Facilita la evaluación de actividades o tareas complejas, una mayor coherencia en la medición del progreso, al definir con precisión cada nivel de desempeño, y el diálogo entre evaluado y evaluador (evaluación formativa) sobre los criterios concretos reflejados en la escala. También se pueden emplear como evaluación sumativa.
 - *Tipos de rúbricas* ([figs. 2 y 3](#)):
 - a) *Analíticas*. Son las más empleadas. Diseccionan las competencias por niveles de desempeño e identifican las fortalezas y deficiencias con más claridad. Al evaluarse cada aspecto de forma individual, aportan más información para la autoevaluación y para el *feedback* que las rúbricas holísticas. En EE. UU. (ACGME) los programas de las especialidades se han diseñado por *milestones* o «hitos de educativos», en función de rúbricas analíticas. En el mismo documento se une la redacción de competencias y la evaluación en unos marcadores gráficos. Los programas están redactados

Evaluación de: Presentaciones orales				
Rúbrica analítica				
	Destacado	Bien	En el límite	Mejorable
1. Contacto visual	Mantiene siempre el contacto visual	Mantiene generalmente el contacto visual	Pocas veces mantiene el contacto visual	No mantiene el contacto visual
2. Conclusiones y aplicación	Resume y da conclusiones precisas y claramente aplicables	Resume y da conclusiones, aunque no del todo aplicables	Presenta algunos errores en las conclusiones, poco aplicables	Presenta conclusiones inapropiadas, con numerosos errores y sin aplicación
Etc.	Etc.	Etc.	Etc.	Etc.
Rúbrica holística				
	Destacado	Bien	En el límite	Mejorable
	1-Mantiene siempre el contacto visual 2-Resume y da conclusiones precisas y claramente aplicables. Etc.	1-Mantiene generalmente el contacto visual 2-Resume y da conclusiones, aunque no del todo aplicables. Etc.	1-Pocas veces mantiene el contacto visual 2-Presenta algunos errores en las conclusiones, poco aplicables. Etc.	1-No mantiene el contacto visual 2-Presenta conclusiones inapropiadas, con numerosos errores y sin aplicación. Etc.

Figura 3 Tipos de rúbricas.

Modificado de: Morán-Barrios et al. (2015).

manteniendo el marco general de 6 dominios competenciales del *Outcome Project* del ACGME (EE. UU.).

- b) *Globales, holísticas o comprensivas*. Evalúan la ejecución de la competencia como un todo. No provee información específica del desempeño en áreas particulares. Están enfocadas al proceso en su conjunto, sin dividirlo en categorías separadas. Se utilizan en tareas en las que los diferentes criterios no son fácilmente separables o implican una pérdida de información.

Diseño de las escalas y errores a evitar

Los descriptores de la escala deben estar muy claramente detallados. La diferencia entre los niveles tiene que ser evidente, siendo mejor definir pocas categorías, bien diferenciadas, para mejorar la consistencia y discriminación. En nuestra opinión, las escalas de 10 niveles nos devuelven a tiempos de una pedagogía obsoleta (suspense, aprobado, notable, sobresaliente), que en la evaluación de profesionales no orientan hacia una buena valoración de las cualidades del desempeño, ni promueven la formación basada en competencias. Otro error frecuente es realizar medias aritméticas y ponderaciones de las diferentes competencias. Lo comentamos más adelante.

Ventajas y desventajas de los métodos de escala

Ventajas

1. *Facilidad de uso*. Son muy fáciles de usar y reducen claramente la subjetividad al determinar con precisión los criterios de la escala.
2. *Bajo coste*. El coste es muy bajo, salvo el del tiempo dedicado a su diseño o adaptación. Comparado con los costes de una ECOE, tanto en diseño y aplicación, las ventajas son claramente superiores.
3. *Estandarización*. Permite a los residentes conocer lo que se espera de ellos y los criterios por los que serán evaluados. Facilita la autoevaluación y promueve la responsabilidad y el aprendizaje.
4. *Medida del progreso*. Permite documentar cualitativamente el progreso del residente, detectar deficiencias y proporcionar criterios para la mejora, lo que es una clara ventaja sobre las listas de cotejo. Facilitan la retroalimentación.
5. *Flexibilidad*. El mismo formulario se puede utilizar, actualizar y adaptar con facilidad a diferentes contextos de trabajo o procesos asistenciales; por ejemplo hospitalización, consulta externa, urgencias o paciente crónico.
6. *Facilitan la evaluación de áreas más subjetivas*, como las relacionadas con la comunicación o el profesionalismo.

Desventajas

1. *Subjetividad* si solo evalúa una persona.
2. *Baja fiabilidad* si existe un único punto de vista y si el tiempo de contacto entre evaluador y evaluado es escaso.
3. *Sesgos de elaboración*. Frases poco concretas que reducen su calidad y fiabilidad, quedando abiertos a la interpretación del evaluador.
4. *Errores de los evaluadores*. Generalmente los observadores son inconscientes de lo que están haciendo, con lo cual son incapaces de corregirlos. El mero conocimiento de estos errores no los elimina, lo que obliga a formar a los evaluadores para reducir los mismos. Si una evaluación del desempeño está sesgada, disminuye considerablemente la probabilidad de estimular el aprendizaje del residente/estudiante y aumenta la posibilidad de tomar decisiones erróneas respecto al evaluado.

Errores en el uso de los formularios y cómo evitarlos

Ligados al formulario

- *Mal diseño*. Tarea y competencia general a evaluar no bien descrita. Criterios (competencias) con frases poco claras, sin verbos de acción y no categorizadas.
- *Error de lógica*. Hay características que parecen relacionadas entre sí y no lo son (por ejemplo empatía, comunicación verbal, ética), por lo que hay que separarlas en el formulario, ya que se tiende a valorarlas con la misma calificación.
- *Escala poco discriminativa*. Escalas solo numéricas, o impares y no verbales o descriptivas. Son mejor las pares pues obligan a decantarse. Idóneamente emplear rúbricas.

Ligados al evaluado

- *Primera impresión (impacto)*. Propensión a hacer un juicio inicial de una persona, favorable o desfavorable y después ignorar o distorsionar la información posterior con el fin de confirmar la primera impresión. Se plantea este tipo de distorsión cuando el aprendiz se esfuerza y trabaja mucho al inicio y después se relaja y se le continúa evaluando positivamente.
- *Efecto halo o deslumbramiento*. Predisposición a dejarse impresionar por el primer impacto de la baja o alta evaluación de un individuo en una característica o aspecto parcial de su desempeño, de modo que se extiende este juicio a la generalidad de su conducta. Por ejemplo, un residente puede ser muy brillante en comunicar conocimientos, pero con serias dificultades en la comunicación con pacientes y el equipo. Este defecto, común entre los evaluadores, se acentúa cuando deben evaluar a sus amigos.
- *Proximidad (comportamiento reciente)*. Si algún acontecimiento reciente, ya sea positivo o negativo, afecta a la evaluación, olvidando el desempeño general a lo largo del

periodo evaluado. El evaluado que sabe esto se esmera en el periodo inmediato a la evaluación. Una dificultad que se presenta con la mayoría de los sistemas de evaluación es determinar el período que abarcará. Precisar esta variable es importante debido a que recordamos mejor lo reciente. Por ello, es importante registrar las evaluaciones y las entrevistas periódicas con el residente.

- *Desconocimiento* de los criterios por los que será evaluado. Estos deben ser públicos y consensuados.

Ligados al evaluador

- *Prejuicios (sexo, cultura)*. El evaluador sostiene *a priori* una opinión personal, basada en estereotipos: edad, sexo, religión, nacionalidad, apariencia física, etc. Implica ser honesto y renunciar cuando las interferencias son inevitables.
- *Benevolencia o lenidad*. Postura de la «buena persona» que otorga puntuaciones altas y no discrimina los comportamientos deficitarios.
- *Rigor o severidad*. Postura del «duro» que considera que todos son ineptos. La evidencia indica que se produce una falta de respeto hacia los calificadores «blandos» y desmotivación con los calificadores injustamente «duros».
- *Tendencia central*. Evaluar la mayor parte de los aspectos de la persona en el punto neutro de la escala de valoración. De esta manera se evita un conflicto de conciencia, o se está poniendo muy poco esfuerzo y atención en el proceso evaluativo.
- *Contraste*. Tendencia a evaluar a una persona por comparación con otra o con un grupo. Hay que comparar a A con A y no A con B. Esta distorsión se observa fácilmente cuando se evalúa primero a personas poco o altamente cualificadas. Las valoraciones dadas a las siguientes personas se verán influidas positiva o negativamente por contraste con las primeras.
- *Semejanza*. Se emiten juicios más favorables de aquellas personas más parecidas al evaluador («es mi vivo retrato»). Las personas que se parecen más a nosotros las valoramos mejor. Esto implica que pueden tomarse decisiones erróneas de asignación de funciones y delegación de responsabilidades.

Cómo mejorar el uso de los formularios y evitar errores en la evaluación

1. *Formar a los evaluadores*. Es el punto más importante, ya que mejora la fiabilidad. Los evaluadores tienen que entender bien el formulario, comprender los errores que se cometen y alcanzar un consenso y comprensión de los aspectos a evaluar y sus escalas. La formación se puede realizar mediante reuniones estructuradas, donde se debate sobre los resultados obtenidos, se comentan los errores y se plantean soluciones prácticas para minimizarlos.
2. *Retroalimentación con los evaluadores*. Se trata de «presionar» a los evaluadores, especialmente a los blandos, a los duros, o los poco comprometidos, para que sean autocríticos y reflexionen sobre sus posibles errores.

3. *Utilizar varios evaluadores.* Mejora la objetividad y la fiabilidad y se minimiza la influencia negativa o positiva entre evaluado y evaluador.
4. *Asegurar una relación temporal suficiente entre evaluador y evaluado.* Los periodos de formación práctica de 4 semanas o menos son insuficientes para poder emitir un juicio sobre el residente, salvo que sea conocido de periodos anteriores.
5. *Evitar ponderaciones y cálculos matemáticos.* No es adecuado promediar los resultados de diferentes competencias evaluadas, ni dar más valor a unas que a otras, ya que todas las competencias son igual de importantes en el desempeño y todas deben ser superadas. Un residente valorado deficitariamente en profesionalismo, con el resto de competencias valoradas adecuadamente tendría una media de apto. ¿Es justo y correcto? O, por ejemplo, ponderar para la certificación final un 50% el resultado de las pruebas de conocimiento y el 20% el de aspectos de comunicación, profesionalismo o gestión equitativa de recursos. ¿Qué implicaciones tiene sobre el tipo de profesional que queremos? Está claro que se le está dando más importancia al conocimiento que a otros aspectos vitales en un profesional. Este error se ve con frecuencia. No se trata de clasificar profesionales o poner una nota, se trata de formarles bajo los principios de la formación basada en competencias, con la mirada puesta en el profesionalismo y la seguridad del paciente.

Principios, planificación de la evaluación (*blueprint*) y propuesta para la evaluación en formación especializada

Principios a seguir en todo proceso de evaluación^a

1. *Propósito (¿para qué?).* Adecuado al amplio abanico de fines y necesidades de: evaluados, educadores (evaluación formativa/sumativa/diagnóstica), empleadores, instituciones reguladoras y la ciudadanía.
2. *Contenido (competencias, ¿qué evaluar?).* Basado en los componentes (dominios) de un marco global de competencias de referencia de la especialidad, institución o país. Ejemplos: competencias de Medicina Familiar y Comunitaria en España, modelo competencias Hospital Universitario Cruces, *Outcome Project* (ACGME, EEUU), *Good Medical Practice* (Reino Unido) o *CanMEDS* (Canadá) (ver la tabla 1 de la primera parte de esta "aula"). Dominios y métodos deben estar asociados.
3. *Métodos (¿cómo evaluar?).* Se escogerán por criterios de validez y fiabilidad, si es posible, pero sobre todo por la viabilidad (sencillez de uso), coste-efectividad, por facilitar el *feedback*, por su impacto educativo y en el desempeño y por ser aceptados por todas las partes. Lo importante es que la información que aporten sea útil y relevante. «No hay buenos o malos métodos, esto es

- relativo, ya que la evaluación implica diseñar un plan, establecer el proyecto educativo y analizar los recursos para su implementación» (ver bibliografía de la primera parte: Van Der Vleuten, 2005).
4. *Feedback.* La evaluación debe proporcionar una retroalimentación relevante para el evaluado y para el proceso formativo.
5. *Evaluadores (¿quiénes?).* Seleccionados acorde a su competencia como educadores y evaluadores.
6. *Personas legas.* Pacientes, pacientes simulados y otras personas pueden evaluar y asesorar en el desarrollo de la evaluación.
7. *¿Cuándo y dónde evaluar?* Tanto la relación temporal entre evaluado y evaluadores, como el contexto formativo influyen en el proceso evaluador.
8. *Accesibilidad y transparencia.* Los estándares exigidos de desempeño deben ser públicos y los resultados estar accesibles a todos los agentes implicados.
9. *Recursos.* Se dispondrá de recursos organizativos y humanos suficientes para el proceso evaluador.
10. *Consenso de expertos.* Valoración final por consenso de expertos, sobre la base de las diferentes evaluaciones e informes disponibles.

Planificación del proceso de evaluación

Siguiendo los principios descritos, las pruebas de evaluación deben estar cuidadosamente seleccionadas en función del programa y su desarrollo, ya que el objetivo es formar. Relacionar método evaluador (¿cómo?) con competencias (¿qué?) es esencial, de lo contrario tendremos desconectada la información que aporta la evaluación sobre la formación que se pretende adquirir. A modo de ejemplo y reflexión, en el caso de las competencias de medicina familiar y comunitaria en España, deberemos relacionar el método evaluador con los dominios competenciales de esta especialidad: atención a la comunidad, a las familias, al individuo, comunicación, gestión, ética/profesionalismo, etc. En la [tabla 2](#) mostramos una combinación de diversos métodos de evaluación y su idoneidad para los diferentes dominios competenciales del modelo del *ACGME Outcome Project* (EE. UU.).

Propuesta para la formación especializada

Sobre estas bases realizaremos nuestro proceso de diseño y planificación, o proyecto de evaluación (*blueprint* o *blueprintig*). Se trata de visualizar en un gráfico o diagrama las interacciones de los elementos más importantes del proceso evaluador. En la [figura 4](#) hacemos una propuesta práctica de diseño de un plan de evaluación para que sirva de guía en la formación especializada. Sugerimos, al igual que en otros entornos internacionales, diseñar formularios de escala adaptados a cada área de rotación (cardiología, urgencias, consulta de preoperatorio, etc.) o área de trabajo (unidad de hemodiálisis, quirófano, etc.), o actividades específicas (cesárea, abordaje del final de la vida, garantizar la seguridad del paciente, dispensación de medicamentos). La evaluación a través de actividades seleccionadas es una propuesta en desarrollo. También se las denomina tareas/actividades confiadas o *entrustable*

^a Basado en: Guidance. Developing and maintaining an assessment system - a PMETB guide to good practice. Postgraduate Medical Education Training Board (U.K.), January 2007.

	Metodos de la evaluación del desempeño				
	Durante el trabajo		Después del trabajo		
	Observación		Reflexión personal	Supervisión	Actividad Asistencial
	Un observador	Varios observadores	-Autoevaluación (mismos formularios que para observación/supervisión) -Incidentes críticos -Memoria/Libro del Residente -Videgrabaciones	-Valoraciones globales: de rotación, periodo, área de trabajo, tareas (por uno o varios evaluadores) -Entrevistas tutor/residente -Informes	-Registros -Datos cuantitativos
1. Seleccionar los métodos (ver 1ª parte)					
2. Instrumento. Lista de cotejo, modelos de escala, rúbrica, guía/formulario de entrevista, formularios <i>ad hoc</i> , guion reflexivo, guía de memoria/libro residente, etc.					
3. Qué competencias valora. Relacionar con claridad método y competencias agrupadas por grandes dominios.					
4. Para qué (finalidad). Formativa, diagnóstica, sumativa, certificación, etc.					
5. Aceptabilidad/consenso. ¿Cómo se va a conseguir entre evaluados y el equipo de trabajo?					
6. Quiénes evalúan. Concretar con precisión quiénes serán los evaluadores; profesionales sanitarios/no sanitarios, pacientes, familiares, personas legas, pacientes simulados.					
7. Dónde, situación o contexto laboral. Seleccionar la situación laboral: actos clínicos, actividades (EPAs), áreas de trabajo, rotaciones, procesos.					
8. Cuándo. Tener en cuenta que la evaluación formativa debe permitir rectificar. Equilibrar lo ideal con la realidad (las rotaciones cortas tiene ciertas limitaciones para evaluación formativa).					
9. Viabilidad . ¿Cómo garantizar la viabilidad? ¿Cuántos recursos se necesitan y se disponen?					
	Informe anual/final elaborado por el tutor (sobre las informaciones aportadas por las evaluaciones arriba planificadas). Se propone una estructura basada en: 1) Profesionalismo. 2) Comunicación. 3) Cuidados del paciente y habilidades clínicas. 4) Conocimientos médicos. 5) Práctica basada en el contexto del Sistema de Salud. 6) Práctica basada en el aprendizaje y la mejora. 7) Manejo de la información. Decisión final: consenso de expertos (comité evaluador)				

Figura 4 Guía para la planificación del proceso de evaluación del residente.

professional activities (tabla 3). En el diseño de esos formularios recomendamos mantener un esquema basado en grandes dominios competenciales, con el fin de no descuidar ningún aspecto clave. Implicar siempre a varios evaluadores (profesionales sanitarios o no).

Recuerda

- Las listas de verificación y métodos de escala son los instrumentos que se emplean en la mayoría de los métodos de evaluación.
- Las listas de verificación identifican o comprueban la presencia o ausencia de determinadas competencias.
- Los métodos de escala identifican el grado de desempeño y su progresión, siendo preferibles los de escalas verbales, evitando las escalas impares o con un centro neutro (ni bien-ni mal).
- Las rúbricas son las escalas más objetivas y las que más informan sobre el grado y progreso del desempeño.
- Podemos diseñar tantos formularios como necesidades tengamos en evaluación.
- Los errores y la subjetividad en el uso de formularios se evitan con un adecuado diseño, detallada descripción de las competencias y escalas y con la formación y colaboración de varios evaluadores.
- Debemos pensar menos en los métodos individuales de evaluación y más en la formación, primando el diseño

Tabla 2 Algunos métodos de evaluación

Dominio competencial	P.E.M.	E.C.O.E	Registros clínicos	Portafolios	Mini-CEX y similares	DOPs/PBA	Discusión de casos	360°	Evaluación por pacientes	Formulario de evaluación global
Profesionalidad	-	+	-	±	+	+	+	+	+	±
Comunicación paciente/familia, habilidades interpersonales y trabajo en equipo	-	+	-	±	+	+	±	+	+	±
Cuidados del paciente y habilidades técnicas	-	+	+	.	+	+	+	±	±	±
Conocimientos	+	±	-	.	-	-	+	±	.	±/-
Práctica basada en el contexto sanitario	±	±	+	±	+	+	±	+	±	±
Calidad, gestión, seguridad, organización sanitaria										
Práctica basada en el aprendizaje y la mejora. Manejo de la información	±	±	+	+	.	.	±	+	.	±

+: Claramente idóneo; ±: potencialmente idóneo; -: no idóneo.

Modificado de: Toolbox of Assessment Methods© Accreditation Council for Graduate Medical Education (ACGME) and American Board of Medical Specialties (ABMS). Versión 1.1. September 2000 [consultado 1 Sep 2016]. Disponible en: <http://www.partners.org/Assets/Documents/Graduate-Medical-Education/ToolTable.pdf>

Tabla 3 Características de una actividad confiada (EPA)

Una actividad de gran importancia de la práctica diaria
 Precisa adquirir conocimientos, habilidades y actitudes
 Da lugar a resultados reconocidos en la especialidad
 Son propias de especialistas cualificados
 Ejecutable dentro de unos límites temporales
 Tienen que llegar a realizarse de forma independiente por el residente
 Son observables y medibles, tanto en su proceso como en sus resultados, para poder obtener conclusiones (lo hace bien/lo hace mal)
 Deben expresar más de una competencia

Fuente: Ten Cate et al. (2015).

de un programa integral educativo, estableciendo una clara relación entre competencias a adquirir, las actividades asignadas y el sistema de evaluación combinando métodos sencillos.

Bibliografía recomendada

Bandiera G, Frank JR, Sherbino J, editores. The CanMEDS assessment tools handbook. The CanMEDS assessment tools handbook: An introductory guide to assessment methods for the Can-MEDS competencies. Royal College of Physicians and Surgeons of Canada; 2006.

Gatica-Lara F, Uribarren-Berrueta T. ¿Cómo elaborar una rúbrica? *Inv Ed Med.* 2013;2:61-5 [consultado 1 de Sep 2016]. Disponible en: <http://riem.facmed.unam.mx/sites/all/archivos/V2Num01/10 PEM GATICA.PDF>

Gil Flores J. La evaluación de competencias laborales. *Educación XX1.* 2007;10:83-106 [consultado 1 Sep 2016] Disponible en: <http://revistas.uned.es/index.php/educacionXX1/article/view/298/254>

Gray JD. Global rating scales in residency education. *Acad Med.* 1996;71 1 Suppl:555-63.

Hales B, Terblanche M, Fowler R, Sibbald W. Development of medical checklists for improved quality of patient care. *Int J Qual Health Care.* 2008;20:22-30.

Lunenburg FC. Performance appraisal: Methods and rating errors. *Int J Scholarly Academic Intellectual Diversity.* 2012;14:1-9 [consultado 1 Sep 2016]. Disponible en: <http://www.nationalforum.com/Electronic%20Journal%20Volumes/Lunenburg,%20Fred%20C.%20Performance%20Appraisal-Methods%20And%20Rating%20Errors%20IJSIAID%20V14%20N1%202012.pdf>

Martíññez Ramírez NL, Terrón López MJ, Gallego Izquierdo T, Álvarez Comino MJ, Rubio Alonso M, Hernando Jerez MA, editores. Competencias en las prácticas clínicas en ciencias de la salud. Guía de estrategias y recursos para su desarrollo y evaluación. Editorial Médica Panamericana. Madrid; 2016.

Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ.* 2010;341:c50-64.

Morán-Barrios J. La evaluación del proceso de formación especializada, una propuesta fundamentada en el proyecto de «Formación especializada basada en competencias del Hospital de Cruces 2008». *Educación Médica.* 2012; 15 (supl. 1): S1-S54. [consultado 1 Sep 2016]. Disponible en: http://www.educmed.net/pdf/revista/15S01/2512058.AAVV_mdp_EDU.pdf

Morán-Barrios J, Olmos Rodríguez MA. Formularios de evaluación: tipos, diseño y aplicación. En: Millán J, Palés J, Morán-Barrios J, editores. Principios de educación médica: desde el grado hasta el desarrollo profesional. Editorial Médica Panamericana. Madrid; 2015. p. 583-95.

Olmos Rodríguez MA, Morán Barrios JM, Martínez Clares E. Evaluación del residente. De la teoría a la práctica: evaluar competencias en entornos sanitarios. En: Millán J, Palés J, Morán-Barrios J, editores. Principios de educación médica: desde el grado hasta el desarrollo profesional. Editorial Médica Panamericana. Madrid; 2015. p. 625–39.

Philibert I. Evolution and application of the competencies and educational milestones in physician graduate education in the United States. Evolución y aplicación de las competencias e

hitos educativos en la formación médica del grado (especialidad) en los Estados Unidos. *Educ Med.* 2015;16:17–24 [consultado 1 Sep 2016] Disponible en: <http://www.elsevier.es/es-revista-educacion-medica-71-sumario-vol-16-num-1-S1575181315X00027>

Ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE Guide No. 99. *Med Teach.* 2015;37:983–1002.