

Impacto educativo de una evaluación basada en el desempeño en residentes de medicina del Programa de Segunda Especialización de la Universidad Privada Antenor Orrego, Trujillo, Perú

Juan A. Díaz-Plasencia, Julio E. Gamarra-Sánchez, Hugo Valencia-Mariñas

Objetivo. Determinar el impacto de la evaluación basada en el desempeño de los residentes utilizando el modelo de Kirkpatrick.

Sujetos y métodos. Estudio descriptivo tipo encuesta a 93 residentes (72 de primer año y 21 de segundo año) y 34 tutores del Programa de Segunda Especialización de la Universidad Privada Antenor Orrego. Se consideró una actitud positiva cuando los encuestados calificaron un ítem con una media igual o superior a 4 en una escala Likert de cinco puntos para los niveles de Kirkpatrick 1 a 3, y el nivel 4 se midió evaluando las tasas de mortalidad neta e infecciones intrahospitalarias en cinco hospitales académicos de Trujillo, Perú, entre los años 2012-2014.

Resultados. La discusión de casos clínicos e incidentes críticos y la retroalimentación del portafolio virtual tuvieron un impacto positivo en los residentes en los niveles 1 a 3 de Kirkpatrick. Los tutores consideraron que las pruebas de progreso, portafolio, ejercicio de evaluación miniclínico (mini-CEX) y observación directa de actividades procedimentales (DOPS) tuvieron un efecto positivo en el desempeño de los residentes en los niveles 1 a 3 de Kirkpatrick. La evaluación tuvo impacto positivo en el nivel 4.

Conclusiones. Se debe enfatizar la discusión de casos clínicos e incidentes críticos y retroalimentación del portafolio virtual y, en general, la evaluación basada en el desempeño de los residentes tendría impacto en el nivel 4 de Kirkpatrick al mejorar la mortalidad neta e infecciones intrahospitalarias en hospitales académicos.

Palabras clave. Competencias. Desempeño. DOPS. Educación médica. Evaluación. Formación. Mini-CEX. Portafolio.

Universidad Privada Antenor Orrego.
Trujillo, Perú.

Correspondencia:

Dr. Juan Alberto Díaz Plasencia.
Las Casuarinas, 208-210. Urb. Santa
Edelmira. Distrito de Víctor Larco.
Trujillo, Perú.

E-mail:

alberdiaz@hotmail.com

Recibido:

01.10.15.

Aceptado:

20.01.16.

Conflicto de intereses:

No declarado.

Competing interests:

None declared.

© 2016 FEM

Educational impact of a performance based evaluation in residents of the Program of Second Specialization in medicine at the Antenor Orrego Private University, Trujillo, Peru

Aim. To investigate the impact of formative and reflective assessment on performance of residents according the Kirkpatrick's model.

Subjects and methods. Cross-sectional study in which a survey was sent to 93 residents (first year, 72; second year, 21) and 34 mentors of the Program of Second Specialization at the Antenor Orrego Private University. Each item was rated on a Likert scale of five points, and was considered relevant when respondents rated an item with an equal or higher average score of 4 (which indicated a positive attitude) and level 4 was evaluated considering net mortality and intrahospitalary rates in five academic hospitals in Trujillo, Peru, among 2012 and 2014.

Results. The components of the corresponding portfolio to reflective and critical incidents and discussion of clinical cases obtained educational impact on levels 1, 2 and 3 in residents. Clinical mentors felt that progress tests, portfolios, Mini-Clinical Evaluation Exercise (Mini-CEX) and Direct Observation of Procedural Skill (DOPS) had a positive effect on the performance of residents in levels 1, 2 and 3 of Kirkpatrick. The assessment system had impact on the level 4 of Kirkpatrick.

Conclusions. It should encourage case-based discussion, reflective critical incidents and feedback components of virtual portfolio in order to improve the performance of residents and performance-based evaluation would have a positive impact in the level 4 of Kirkpatrick in reducing the rates of net mortality and nosocomial infections in academic hospitals.

Key words. Competencies. DOPS. Evaluation. Medical education. Mini-CEX. Performance. Portfolios. Training.

Introducción

La educación médica de posgrado ha evolucionado notablemente en años recientes con programas de formación que se llevan a cabo en los niveles más altos de responsabilidad por sus resultados educacionales [1]. Durante la última década, Estados Unidos [2], Reino Unido [3] y Canadá [4] han cambiado el formato de sus programas de formación de los residentes médicos. El cambio en las necesidades y expectativas de los pacientes y la sociedad en la era de la información de Internet [5], la evolución en la tecnología médica [6], la aplicación de la legislación restrictiva sobre las horas de trabajo de los residentes, la feminización de la profesión médica [7] y el renovado equilibrio de ambición profesional sopesado contra los estilos de vida y responsabilidades privadas [8,9] eran todos los acontecimientos que exigían una reforma de la formación para obtener especialistas competentes. No cabe duda de que el mejor sistema de aprendizaje es el basado en la práctica, a la cabecera del enfermo, en especial cuando esta práctica clínica se basa en la evidencia y se apoya en las tecnologías de la información y la comunicación. Este método, con el apoyo de un mentor experto, es el que más y mejor forma y motiva a los residentes.

La utilidad de cualquier evaluación se conceptualiza como un producto de la validez, fiabilidad, transparencia, viabilidad, aceptación e impacto educativo [10,11]. Las recientes reformas en la educación médica de posgrado han dado lugar a un mayor énfasis en la evaluación basada en el lugar de trabajo (EBT), un sistema mediante el cual se evalúan las habilidades clínicas y otros atributos de los médicos en el contexto de su entorno de trabajo. Se utilizan varios métodos para proporcionar esta información, incluyendo el portafolio de aprendizaje, el ejercicio de evaluación clínica reducido (mini-CEX), la discusión basada en casos, la observación directa de habilidades procedimentales (DOPS), la evaluación basada en procedimientos, la evaluación objetiva y estructurada de habilidades técnicas y la retroalimentación de múltiples fuentes. La retroalimentación y calificación son dadas por el tutor y esta información se compila y se retroalimenta a los supervisores educativos. Adicionalmente, a pesar de que el portafolio, mini-CEX y DOPS son instrumentos muy utilizados en ciencias de la salud en los países anglosajones, en Perú todavía no se ha introducido en otras unidades de segunda especialización.

Una implementación exitosa de un programa modernizado de entrenamiento de residencia depende en gran medida del apoyo de sus usuarios. Por lo tan-

to, es esencial que poco después de la aplicación, se obtenga una retroalimentación para asegurar una evaluación adecuada. Desde esta perspectiva, se realizó un estudio multicéntrico entre los residentes y los tutores asistenciales en el programa de residentado médico de la Universidad Privada Antenor Orrego (UPAO) para examinar la efectividad del programa de evaluación basada en el desempeño de acuerdo a los niveles de Kirkpatrick, utilizado en los médicos residentes durante los años 2012-2014 y, de esa manera, reestructurar el programa.

Sujetos y métodos

Diseño del estudio

Este estudio descriptivo transversal tipo encuesta contó con la aprobación del Comité de Residentado Médico de la UPAO. En julio de 2014 se distribuyeron encuestas anónimas a los tutores ($n = 34$) y a todos los residentes ($n = 93$), 72 de primer año (R1) y 21 de segundo año (R2), pertenecientes al Programa de Segunda Especialización en medicina de la UPAO, que realizaban su formación especializada en cinco hospitales (Hospital Belén, Hospital Regional Docente, Hospital Víctor Lazarte Echegaray, Instituto Regional de Enfermedades Neoplásicas y Hospital de Chocope), distribuidos en 30 programas. Ambos grupos fueron encuestados para evaluar la satisfacción y la comprensión del sistema de evaluación autorreflexivo sumativo y formativo, basado en el desempeño en el lugar del trabajo, implementado por el Programa de Residentado Médico desde el año 2011. El sistema de evaluación comprende: un portafolio virtual implementado en la plataforma Moodle, con una evaluación mensual de un caso clínico real prototipo y de un incidente crítico reflexivo, ambos analizados de acuerdo a la medicina basada en la evidencia; documentación reflexiva de su práctica clínica diaria en los diferentes contextos hospitalarios, y evaluación de 360° con autoevaluación, evaluación por pares y heteroevaluación. Además, los residentes rinden dos pruebas teóricas de progreso, seis mini-CEX y cuatro DOPS por año, tomados por los tutores de la plantilla docente, quienes brindan retroalimentación directa y oportuna a los residentes, tras lo cual se establece un plan de mejora a corto y medio plazo.

Encuesta y distribución

La encuesta consistió en 256 preguntas para los residentes y 160 preguntas para los tutores, desarro-

lladas y valoradas en una escala tipo Likert de cinco puntos (de 'muy en desacuerdo' a 'totalmente de acuerdo').

El instrumento de la encuesta fue creado por el autor para el seguimiento y evaluación del programa de EBT institucional y fue revisado, validado y aprobado por un panel local de educadores médicos y coordinadores de las cinco sedes hospitalarias. Una carta de presentación se adjuntó a la encuesta, explicando el propósito del estudio y asegurando el anonimato y la confidencialidad. La participación fue voluntaria.

Dos preguntas adicionales documentaron el año de formación del residente y la cantidad de tiempo dedicada al mes para completar el portafolio. Para aumentar la tasa de respuesta, una segunda solicitud para completar las encuestas se hizo a través de correo electrónico, a finales de julio de 2014.

En total, 125 de los 126 participantes (99%) respondieron a la encuesta: 93 (100%) residentes y 32 (99%) médicos asistentes (tutores). El análisis de valores perdidos reveló que 135 encuestados completaron todos los ítems de la encuesta: cuatro residentes encuestados no declararon su edad. De los 93 residentes que respondieron, la edad media era de 32,1 años (rango: 24-39 años), 58 eran hombres y 35 eran mujeres. Sesenta y un residentes ingresaron al sistema en la modalidad libre y 32 por destaque. Los 32 tutores que respondieron tenían una edad media de 49,8 años (rango: 36-67 años), 25 de los cuales eran hombres, y contaban con una experiencia docente-asistencial media de 13,4 años. Los R1 tuvieron, antes de ingresar al programa, una experiencia media previa como médicos generales de 4,9 años, y los R2, de 8,2 años.

Impacto educativo

Se analizó la efectividad del sistema de evaluación usando el modelo de Kirkpatrick [12], descrito en Freeth et al [13], mediante el cual pueden explorarse los efectos en cuatro niveles (Figura) [11,14,15]:

- *Nivel 1 (reacción)*: satisfacción de los participantes con el programa, organización, contenidos, métodos, profesores y experiencia de aprendizaje.
- *Nivel 2 (aprendizaje)*: adquisición de conocimientos y destrezas o cambio en las percepciones hacia lo aprendido.
- *Nivel 3 (conducta)*: disposición para aplicar los conocimientos y destrezas o practicar lo aprendido en el lugar de trabajo.
- *Nivel 4 (resultados)*: resultados en los pacientes o la salud como consecuencia del programa (p. ej., número de vidas salvadas).

Figura. Cuatro niveles de Kirkpatrick.



Se utilizó una encuesta anónima y semiestructurada para explorar los tres primeros niveles de Kirkpatrick que fue aplicada a todos los tutores y médicos residentes al finalizar el año académico 2014. Se incluyó un cuestionario sobre percepción del sistema de evaluación y de la retroalimentación ofertada-recibida con una escala tipo Likert de cinco puntos como evidencia para el nivel 1 [16]. Los aprendizajes adquiridos de acuerdo a las competencias ACGME se usaron como fuente de información para el nivel 2, y los cambios efectuados en la práctica clínica, como evidencia para el nivel 3. Cada ítem se clasificó en una escala tipo Likert de cinco puntos. El nivel 4 fue explorado a través de los indicadores de impacto de las sedes hospitalarias relacionados con las tasas de mortalidad e infecciones intrahospitalarias desde los años 2012 a 2014, considerando el 2011 como año base en que se aplicó el programa de evaluación.

Análisis estadístico

Se utilizó estadística descriptiva para todos los ítems de la encuesta, incluyendo medias $\pm$ desviación estándar y frecuencias. Para evaluar si los elementos de la encuesta se consideraron pertinentes, se calculó el índice de validez de contenido. Debido a que no existe un acuerdo acerca del valor aceptable del índice de validez de contenido, de acuerdo con las directrices de expertos, se escogió un valor de corte del 80% [17,18] o, de acuerdo con nuestra propuesta, cuando el promedio de todos los encuestados fue igual o superior a 4 (lo que indicó una actitud

Tabla I. Promedios de los niveles de Kirkpatrick de los residentes y tutores según formatos de evaluación.

		Examen teórico	Portafolio	Mini-CEX	DOPS	Retroalimentación	Discusión basada en casos	Incidente crítico	Evaluación 360°
Residentes	Nivel 1	3,78 ± 0,86	2,77 ± 1,04	3,48 ± 1,06	3,79 ± 0,93	4,62 ± 0,57	4,00 ± 0,82	4,07 ± 0,94	3,66 ± 0,89
	Nivel 2	3,68 ± 0,86	3,04 ± 1,05	3,37 ± 1,04	3,80 ± 0,80	4,00 ± 0,75	4,22 ± 0,77	4,22 ± 0,77	3,59 ± 0,96
	Nivel 3	3,77 ± 0,84	3,09 ± 0,99	3,51 ± 0,96	3,82 ± 0,94	4,40 ± 0,68	4,08 ± 0,81	4,08 ± 0,81	3,62 ± 0,88
Tutores	Nivel 1	4,34 ± 0,48	4,10 ± 0,70	4,34 ± 0,65	4,25 ± 0,62	–	–	–	–
	Nivel 2	4,63 ± 0,49	4,44 ± 0,67	4,25 ± 0,62	4,41 ± 0,62	–	–	–	–
	Nivel 3	4,34 ± 0,83	4,19 ± 0,78	4,28 ± 0,58	4,31 ± 0,54	–	–	–	–

DOPS: observación directa de habilidades procedimentales; mini-CEX: ejercicio de evaluación miniclínico.

positiva) en la escala de Likert de cinco puntos. Para la estadística inferencial se utilizó la prueba de chi al cuadrado para variables categóricas y la *t* de Student para variables numéricas, y se consideró estadísticamente significativo un valor $p < 0,05$. Los análisis estadísticos se realizaron con el programa SPSS v. 21.0.

Resultados

Los R1 sólo consideraron el formato DOPS como aceptable en los niveles 1 y 3 de Kirkpatrick –en relación al índice de validez de contenido con una puntuación de 4-5 mayor del 80%–, siendo esta puntuación menor a ese nivel en los R2 en los niveles de Kirkpatrick 1, 2 y 3. En orden descendente, el examen teórico, el mini-CEX y el portafolio obtuvieron menor puntuación (< 4) en los tres niveles tanto en los R1 como en los R2. Sin embargo, los tutores generaron puntuaciones superiores a 4 en todos los formatos de evaluación en los niveles 1 a 3 de Kirkpatrick.

La retroalimentación ofrecida en los formatos de portafolio, mini-CEX y DOPS obtuvo puntuaciones aceptables entre los R1 y R2 en los niveles 1, 2 y 3 de Kirkpatrick. Con respecto a la evaluación global del programa (nivel 3 de Kirkpatrick), los R1 mostraron un mayor cambio en su desempeño relacionado con los objetivos del plan de estudios (R1: $72,8 \pm 14,9$; R2: $61,9 \pm 20,6$; $p = 0,032$) y en su desempeño general (R1: $76,5 \pm 14,9$; R2: $63,9 \pm 20,6$; $p = 0,014$) como producto del sistema de evaluación del programa de residentado médico, aunque sin superar las puntuaciones de 80% consideradas como aceptables.

Los residentes consideraron que la discusión basada en casos clínicos fue positiva para su desempeño en los niveles 1, 2 y 3 de Kirkpatrick. Con relación al incidente crítico y a la retroalimentación proporcionada en el portafolio virtual, los residentes señalaron una puntuación superior a 4 en los niveles 1, 2 y 3. La evaluación de 360° no obtuvo una puntuación aceptable en estos tres niveles entre los residentes (Tabla I).

Con relación a la valoración del nivel 4 de Kirkpatrick mediante indicadores de calidad de atención hospitalaria, hubo una disminución significativa en los años 2012-2014, cuando se aplicó el sistema de evaluación basado en el desempeño de las tasas de mortalidad neta y de infecciones intrahospitalarias –con respecto al año basal 2011– en los cinco hospitales docentes de Trujillo, donde los residentes realizan su formación supervisada (Tabla II).

Discusión

Examen teórico de progreso

Fue desarrollado inicialmente para evitar un comportamiento de aprendizaje direccionado por la evaluación de los planes de estudios de pregrado basado en problemas [19]. Por esta razón, la prueba está en sintonía con el nivel de posgrado y abarca todos los temas que se van a tratar durante el programa de estudios de posgrado. Sin embargo, todos los esfuerzos de estudio pertinentes deben ser recompensados, al preparar al residente para una carrera de aprendizaje permanente [20]. En la UPAO, las pruebas de progreso permiten una buena im-

presión del avance del conocimiento de los residentes. Además, los resultados de las pruebas constituyen una rica fuente de información que puede ayudar a los residentes y profesores en la identificación de los objetivos individuales de aprendizaje. Sin embargo, como el objetivo de la evaluación utilizada formativamente es estimular el aprendizaje, los determinantes psicométricos no son los únicos determinantes de la utilidad. Se puede argumentar también que la aceptación por parte de los interesados y el impacto educativo que se alcanza son aspectos muy importantes [21]. En esta perspectiva, la aceptación puede definirse como el grado en que partes interesadas apoyan la evaluación y la interpretación asociada de las calificaciones [10]. El impacto educativo, por su parte, es el grado en que la retroalimentación resultante de la evaluación se convierte en una parte integral del proceso de aprendizaje e influye en el comportamiento de aprendizaje del residente. A pesar de que la evaluación formativa ha tomado vuelo durante la formación de posgrado en el ámbito internacional [22], la evidencia sobre la aceptación o el impacto educativo de este formato de evaluación es todavía difícil de encontrar [23].

A pesar de que la prueba de progreso se introdujo específicamente por sus propiedades de retroalimentación bajo el presentimiento de que los resultados de las pruebas podrían guiar el aprendizaje de los residentes, el impacto educativo, con excepción de los R1, quienes tuvieron puntuaciones de 4 y 5 en la escala de Likert superiores al 70% en los niveles de Kirkpatrick 1, 2 y 3, no cumplió las expectativas. Se pueden dar varias posibles explicaciones para este hallazgo: calidad de la retroalimentación, efecto de la clasificación y estrictez de los tutores.

Portafolio electrónico

En este trabajo se evidenció que, por parte de los residentes, el portafolio alcanzó puntuaciones medias neutras de 3 en los niveles de Kirkpatrick 1 (reacción), 2 (aprendizaje) y 3 (conducta); sin embargo, la valoración global del portafolio y de sus componentes (caso clínico reflexivo, incidente crítico) y la retroalimentación por parte de los tutores alcanzaron promedios superiores a cuatro, ofreciendo una opción viable para demostrar las competencias ACGME de mejora de conocimientos médicos y de aprendizaje basado en la práctica. La investigación del portafolio informa de efectos positivos para los niveles 1, 2 y 3. Los resultados obtenidos en el presente estudio en los residentes difieren de otros trabajos en que la mayoría de los usuarios del portafolio se sienten alentados y apoyados en su

Tabla II. Evaluación del nivel 4 de Kirkpatrick según indicadores de impacto en cinco hospitales de Trujillo.

	Año 2011 (basal)	Año 2012	Año 2013	Enero-junio 2014	<i>p</i> ^a
Altas	44.723	44.641	45.301	33.843	–
Mortalidad bruta (%)	3,2	2,6	3,5	3,3	0,000000
Mortalidad neta (%)	2,8	2,8	2,7	2,7	0,045
Infecciones intrahospitalarias (%)	3,2	2,6	2,5	2,2	0,000000

^a χ^2 ; $p < 0,05$ estadísticamente significativo.

desarrollo profesional [24,25] y reportan que el portafolio facilita la consecución de los objetivos de aprendizaje [26]. Los cambios producidos en el rendimiento correspondieron a: ser más reflexivo [27], conocimientos médicos actualizados, y un mejor mantenimiento de registros médicos [28]. En un estudio se demostró la ausencia de cualquier efecto en el nivel 3 de Kirkpatrick y no hubo un aumento significativo en las puntuaciones del portafolio en los años siguientes [29].

Tochel et al [30] realizan una revisión sistemática de 56 artículos que describen el uso de un portafolio para el aprendizaje en un ambiente de trabajo o estudio profesional. Concluyen que, a pesar de existir una extensa base de pruebas, ésta contiene pocos estudios de alta calidad con mensajes generalizables acerca de la efectividad de los portafolios. Sin embargo, hay buena evidencia de que si se implementan bien, los portafolios son eficaces y prácticos en diversas formas, incluyendo el aumento de la responsabilidad personal por el aprendizaje y el apoyo al desarrollo profesional. Las versiones electrónicas son mejores para fomentar la reflexión y los usuarios pasan más tiempo en ellos voluntariamente. La retroalimentación regular de un mentor realza este éxito, a pesar de demandas que compiten con el tiempo de los usuarios y el escepticismo ocasional sobre el propósito de un portafolio. Informes de fiabilidad entre evaluadores para la evaluación sumativa de datos del portafolio son variados y hay beneficios que se pueden obtener a partir de la triangulación con otros métodos de evaluación. Estos resultados de éxito y preocupaciones son consistentes con los de la revisión sistemática sobre los portafolios de aprendizaje de Driessen et al [31]. Refinar aún más nuestras rúbricas y el aumento de la capacitación del evaluador podrían allanar el camino para una evaluación sumativa individual precisa.

Tromp et al [32] encuentran que el formato basado en la web ha mejorado la administración del portafolio. Los residentes entienden el propósito del portafolio y su potencial para promover el crecimiento y el aprendizaje autorreflexivo. Con las limitaciones de tiempo impuestas a los residentes, es importante que el portafolio se introduzca en un momento en que se va a maximizar la motivación y dedicación de los residentes al programa [31]. La retroalimentación que se proporciona promueve aún más el pensamiento y la investigación sobre un tema determinado. Debido a que la participación del profesorado es tan importante para el éxito de un programa del portafolio de aprendizaje, se necesitan datos subsiguientes para identificar la percepción y la perspectiva de la plantilla docente en la eficacia de los portafolios de aprendizaje como una herramienta para la práctica reflexiva. En el presente trabajo de investigación, el portafolio de aprendizaje basado en la web proporcionó a los residentes una experiencia positiva en la autorreflexión y el aprendizaje basado en la práctica, y los tutores estuvieron satisfechos con la facilidad de supervisión. Un trabajo adicional es necesario para evaluar la calidad y la correlación de la calidad del portafolio con el rendimiento clínico.

Casos clínicos reflexivos

Existe bibliografía que apoya la aplicación del caso clínico, como Chamberlain et al [33], quienes afirman que la revisión dirigida de 'caso clínico' posibilita en los estudiantes el desarrollo del pensamiento crítico, clínico, comunicación efectiva, que sea competente para producir conocimientos ya que implica comprensión, reconstrucción e interpretación de la información que se le provee, desarrolla la capacidad para seleccionar, organizar y transformar la información, vincular los conocimientos previos y los nuevos, y aumenta la posibilidad de autonomía. Además, ayuda a generar una hipótesis, interpretación de pruebas diagnósticas, diagnóstico diferencial y verificación diagnóstica, proporcionando tanto el lenguaje como los métodos para la resolución de un problema clínico [34].

El presente estudio evaluó la satisfacción de los residentes con la discusión de casos clínicos como estrategia didáctica, en el contexto de un cambio curricular hacia un modelo basado en competencias. Los resultados mostraron que la discusión de casos clínicos fue bien aceptada por los residentes y efectiva en los tres niveles iniciales de Kirkpatrick, lo cual es muy favorable para su implementación en un programa basado en competencias. Estos resul-

tados concuerdan con los de otros autores, en cuanto a la importancia de desarrollar en los alumnos un aprendizaje significativo, especialmente con Chamberlain et al [33], quienes aplican casos clínicos en un curso con dos grupos de alumnos. El primero recibió sus clases de forma tradicional y el segundo participó en la revisión de casos clínicos. Los resultados obtenidos en las calificaciones finales del grupo 1 fueron del 84% y 80%, comparado con 88% y 86,7% del grupo 2 ($p < 0,001$; $p < 0,001$), respectivamente. Estos autores midieron otros parámetros como la adquisición de conocimientos, habilidades y competencias, desenlaces clínicos en los pacientes y otros, los cuales corresponden al segundo nivel de evaluación de impacto de Kirkpatrick.

En el presente estudio, las opiniones negativas fueron menores al 20% (categorías 'nunca' y 'casi nunca'), lo cual habla indiscutiblemente de la gran satisfacción obtenida con los casos clínicos. Por lo anterior consideramos que su uso resulta satisfactorio, recomendando su aplicación en el sistema basado en competencias, y es más compatible con modelos curriculares novedosos. Como limitación en la enseñanza, el 'caso clínico' (la historia clínica) se presenta en un tiempo limitado, de forma total, simplificada, con una perspectiva global y compila todo el proceso diagnóstico; carece del desarrollo cronológico del proceso de enfermedad, lo que podría distar de la realidad del aprendizaje y de la práctica del diagnóstico clínico. Por lo tanto, la reconstrucción cronológica permite en parte agregar un valor educativo a los ejercicios de resolución de 'problemas clínicos', a través de la presentación de fragmentos de información para discutirlo secuencialmente, conforme van surgiendo en el tiempo [35].

Aprendizaje reflexivo e incidente crítico

Durante la residencia, las reflexiones sobre los problemas que surgen en la práctica asistencial son una fuente inagotable de aprendizaje. No hay mejor maestro que la propia experiencia. Al basarse en las cuestiones que plantea el incidente crítico, el portafolio facilita que el residente se responsabilice de sus competencias y adopte una actitud más activa sobre sus necesidades de aprendizaje [36]. Por eso, era de esperar que en el presente estudio la sección de incidentes críticos haya sido valorada con puntuaciones superiores a 4 en los niveles 1, 2 y 3 de Kirkpatrick por parte de los residentes, radicando su valor agregado en la reflexión de estos eventos críticos que les permite evaluar con mayor responsabilidad y gestionar sus necesidades de aprendiza-

je, conduciéndolos hacia el aprendizaje experiencial y autodirigido.

La presentación de informes de incidentes críticos sigue siendo un aspecto importante de la mejora de la seguridad del paciente y se utiliza en todos los organismos del National Health Service para monitorizar eventos adversos [37]. Las habilidades de reflexión de los estudiantes se consideran muy importantes en el desarrollo de conductas profesionales [38]. Mediante el estudio de las reflexiones de los estudiantes sobre el comportamiento de incidentes críticos, se podría explorar el desarrollo de buenos comportamientos profesionales. Su introducción podría proporcionar herramientas para que los estudiantes examinen su propio comportamiento y actitudes, para considerar las razones por las que se produjeron los incidentes, y para permitir plantear soluciones que se deben desarrollar para su prevención.

Evaluación basada en el lugar del trabajo

Las principales características comunes a las EBT incluyen la recopilación de datos de la actividad de aprendizaje que constituye el foco de la evaluación, oportunidad tanto para el aprendiz como para el docente de introducir comentarios en texto libre relacionados con el evento o retroalimentación recibida, y mapear el plan de estudios a través de la EBT. Para las cuatro principales EBT (discusión basada en casos, mini-CEX, evaluación basada en procedimientos y DOPS), los dominios son marcados como 'no evaluado', 'requiere desarrollo' o 'satisfactorio', y un nivel global de 1 (bajo) a 4 (alto). Los niveles globales, sin embargo, son inconsistentes, con algunas puntuaciones de los residentes sobre su capacidad para su nivel actual de formación (discusión basada en casos, mini-CEX y DOPS), mientras que otros ofrecen una clasificación basada en la progresión hacia el certificado de culminación del entrenamiento (discusión basada en casos), resultando en lo que se ha descrito como una 'economía mixta' [39]. Epstein et al [40] hacen hincapié en la necesidad de una evaluación multidimensional basada en la observación de los residentes en situaciones reales, en la retroalimentación de los compañeros y de los pacientes, y en mediciones de los resultados.

Un hallazgo importante de este estudio es que los residentes, a diferencia de los tutores, consideraron los instrumentos de evaluación de observación directa (DOPS) predominantemente inadecuados para la evaluación de competencias. Una explicación para este hallazgo podría ser la reciente puesta en marcha del programa de formación o no estar fami-

liarizados con el uso de las herramientas de evaluación. De todas ellas, sólo DOPS logró un índice de validez de contenido de al menos el 80% en los residentes de segundo año. DOPS es un instrumento utilizado para la evaluación de las habilidades procedimentales de los residentes. Como era de esperar, los tutores y los residentes de mayor experiencia (R2) valoraron mejor DOPS y, por ende, la forma en que se evaluaba tradicionalmente a los residentes. Esto confirma que las interacciones personales con los tutores y las evidencias de habilidades médicas y quirúrgicas todavía son consideradas fundamentales para la formación por los encuestados.

Shalhoub et al [41] revisan y discuten la evidencia educativa empírica existente que rodea la introducción y el uso continuo de las EBT en la formación quirúrgica. Recomendamos que los mini-CEX y las discusiones de casos clínicos deben llevarse a cabo por lo menos seis veces al año cada uno durante el entrenamiento quirúrgico básico, y su uso en la formación especializada depende de la especialidad y el nivel de formación. DOPS quirúrgicos también deben realizarse al menos seis veces al año durante el entrenamiento quirúrgico básico. El documento de implementación de la EBT por el General Medical Council y la Academia de los Reales Colegios Médicos [42] ofrece pocos datos o pruebas que apoyan la validez o fiabilidad de las EBT. Sin embargo, citan el trabajo del psicólogo George Miller en los años noventa [43] y prosiguen afirmando que las EBT exploran todas las áreas de la pirámide de Miller. El primer trabajo que mira directamente a la EBT después de su introducción se centró en la satisfacción del usuario por medio de una encuesta en línea [44]. De los 539 encuestados, el 49% describió la EBT como 'mala' o 'muy mala'. En general, el 6% declaró que el *Intercollegiate Surgical Curriculum Program* (ISCP) había afectado negativamente en su formación y un 41% informó de un efecto negativo global sobre su formación. Sólo el 6% comunicó un efecto positivo en la formación. En su respuesta a este artículo, Chris Munsch, presidente del Comité Conjunto de Formación Quirúrgica, reiteró la intención original de las EBT como 'herramientas de retroalimentación formativa para orientar y dirigir la formación, en lugar de evaluaciones basadas en competencias sumativas' [45]. En la repetición de la encuesta tres años más tarde, de los 201 que respondieron, las calificaciones modales habían cambiado a un 'promedio' [46]. Centrándose en la EBT, el 35% la calificó como 'pobre' o deficiente, y un 22%, como 'buena' o mejor, con calificaciones similares para la retroalimentación multifuente, que se evaluó indivi-

dualmente. Llegaron a la conclusión de que, a pesar de las mejoras generales en el rendimiento de parte del ISCP, la EBT había sido señalada como un foco de descontento. Los autores destacaron que la EBT siguió careciendo de las pruebas necesarias de validez para su uso.

En una revisión sistemática posterior en la que se examinaba específicamente el efecto de la EBT en la educación y el desempeño de los médicos, se identificaron 16 estudios pertinentes de calidad mixta [47]. De ellos, 15 eran estudios descriptivos no comparativos o de observación, mientras que el otro era un ensayo controlado aleatorizado. De los estudios incluidos, y a pesar de resultados mixtos, la mayoría de los médicos sintió que la EBT tenía valor educativo, aunque la evidencia de cambio en la práctica fue contradictoria. Algunas de las pruebas más contundentes en esta revisión sistemática se derivó de datos detallados de grupos focales [48]. Un total de cuatro estudios examinaron el mini-CEX; uno, la DOPS, y tres se ocuparon de múltiples EBT. Todos estos estudios informaron de resultados positivos para el efecto educativo de la EBT, pero no había evidencia objetiva alguna de un mejor rendimiento con estas herramientas. Como tal, se precisan trabajos adicionales para determinar si las EBT pueden alcanzar niveles de Kirkpatrick de cambio más altos que resultne en mejoras en la atención y la evolución del paciente. Beard et al [49], en una respuesta crítica a Miller y Archer [47], encuentran evidencia fuera de medicina y cirugía que apoya la causa formativa de la EBT, señalando también que es probable que se hayan perdido trabajos actuales en curso por la revisión sistemática y que la falta de la evidencia actual para las EBT no constituye una prueba de su ineficacia y, ciertamente, no una justificación para su abandono.

En una revisión sistemática integral de las herramientas para la observación directa y la evaluación de las habilidades clínicas de los médicos residentes, se identificaron 55 instrumentos, de los cuales 11 tenían evidencias de validez basada en la estructura interna y la relación con otras variables [50]. Entre estos 11, el mini-CEX fue la única EBT que formaba parte del ISCP. Además, se demostró que el mini-CEX tuvo la evidencia de validez más fuerte. Un posterior metaanálisis de 11 estudios publicados que examinaron el mini-CEX utilizando el modelo de efectos aleatorios demostró que la validez de criterio y de constructo –siendo la validez de criterio la capacidad de una evaluación para predecir futuros desempeños [51]– del mini-CEX fueron sustentadas por diferencias pequeñas y grandes de tamaño del efecto, basadas en mediciones

entre el rendimiento de los residentes y el desempeño de habilidades clínicas [52]. Esto fue interpretado por los autores como una indicación de que el mini-CEX es un importante instrumento para la observación directa del desempeño clínico de los residentes.

Overeem et al [53] realizaron una revisión sistemática con búsquedas en las bases de datos Medline, PsychINFO, CINAHL, EMBASE y Cochrane para los artículos en idioma inglés, y se complementó con una búsqueda manual de las listas de referencias de los estudios y bibliografías de los artículos de revisión pertinentes. Se incluyeron los estudios que trataban de evaluar el desempeño de los médicos individuales en la práctica habitual. Un total de 64 artículos cumplieron los criterios de inclusión. Observaron seis métodos diferentes de evaluación del desempeño: pacientes simulados, observación de videos, observación directa, evaluación por pares, auditoría de las historias clínicas y portafolio. La evaluación por pares fue el método más factible en términos de coste y tiempo. En 19 estudios se informaron efectos positivos, mientras que en dos no se comunicó efecto alguno. Ocho estudios informaron sobre la eficacia en términos de satisfacción del alumno (nivel 1). Cuatro estudios investigaron el efecto en los resultados del aprendizaje (nivel 2). El logro de la mejora del rendimiento se investigó en 12 estudios (nivel 3). El nivel 3 puede dividirse en las mejoras reportadas en el rendimiento (nivel 3a) y mejoras medidas en el rendimiento (nivel 3b). No se encontraron estudios en relación con la eficacia en términos de los resultados en pacientes y en la salud (nivel 4).

En la presente investigación, las pruebas de progreso y las evaluaciones basadas en el trabajo (mini-CEX, DOPS y evaluación de 360°) no condujeron a una mejora en el desempeño. El portafolio, de manera general, no tuvo impacto educativo en los residentes, pero la retroalimentación mejoró su rendimiento. Los casos clínicos y los incidentes críticos obtuvieron impacto educativo en los niveles 1, 2 y 3. Los tutores consideraron que las pruebas de progreso, portafolio, mini-CEX y DOPS tuvieron un efecto positivo en el desempeño de los residentes en niveles 1, 2 y 3 de Kirkpatrick. La evaluación global basada en el desempeño formativa y sumativa utilizada en la UPAO tendría impacto en el nivel 4 de Kirkpatrick. La mortalidad hospitalaria de las primeras 48 horas de ingreso (bruta) se encuentra más afectada con los servicios médicos ambulatorios, no así la muerte que acontece a las 48 horas o más de hospitalización (neta), en que la responsabilidad de la atención hospitalaria es mayor. La tasa

de mortalidad neta (mortalidad institucional) sirve para evaluar la calidad de atención que se brinda a los pacientes hospitalizados, permitiendo un mayor nivel de precisión que la tasa de mortalidad bruta, ya que se presume que las muertes ocurridas están asociadas con la atención hospitalaria [54]. Durante los tres años en que se aplicó el sistema de evaluación en nuestro programa, que coincide con el tiempo de formación de esta cohorte de residentes (2012-2014) en su camino hacia la competencia, la mortalidad bruta fue variable en los cinco hospitales académicos donde se desarrolló la residencia. Adicionalmente, durante ese período hubo una tendencia hacia la disminución significativa y progresiva de la tasa de infecciones intrahospitalarias y de la mortalidad neta en los cinco hospitales. Si bien es cierto que este resultado podría ser especulativo por la naturaleza multifactorial que afecta los indicadores de calidad hospitalaria, ni la mortalidad neta ni las infecciones se incrementaron en el período transcurrido durante la implementación del programa, tal como fue la hipótesis de trabajo. La residencia supervisada impulsa la actividad académica, promoviendo y fortaleciendo de manera programática las reuniones de morbilidad y mortalidad, rondas académicas, club de revistas, presentaciones de casos, sesiones didácticas de los residentes, en un entorno que favorece el aprendizaje a través de las interacciones con los pacientes, la observación de médicos y compañeros en acción. Ello, aunado a la práctica reflexiva con retroalimentación constante por parte de la plantilla docente-asistencial y la medicina basada en la evidencia, tendría impacto en una mejora de la calidad asistencial. En los tiempos actuales, la sociedad exige mayores resultados, mayor calidad y mayor rendición de cuentas. Un reflejo de esto son los movimientos por la seguridad del paciente que están siendo liderados no sólo por organismos internacionales (Organización Mundial de la Salud), sino nacionales. En cierta forma, la sociedad exige mayores resultados y menores efectos adversos en la prestación de servicios. Hoy día, hay una tendencia a la reclamación legal por estos efectos. Factores como las largas jornadas laborales, la impericia médica, la imprudencia médica, las condiciones para la prestación de servicios, la falta de normas y protocolos en hospitales, la falta de trabajo en equipo y otras más actúan en sinergia para la generación de efectos adversos en los pacientes. En ese orden de ideas, los hospitales, las universidades, las sociedades médicas y los estudiantes presionarán cada día más por mejorar estas condiciones, lo que en última instancia beneficiará a los profesionales en formación y a los pacientes.

Retroalimentación

En el presente estudio, los residentes valoraron muy alto la retroalimentación ofertada por los tutores académicos, sobre todo cuando fue administrada directamente tras el encuentro clínico, planteándose de acuerdo a los resultados objetivos a desarrollar para la consecución de la competencia clínica. Los diseños de investigación futuros deben prestar especial atención a develar la eficacia de las evaluaciones formativas en términos de mejora del rendimiento. El concepto de que la evaluación direcciona el aprendizaje se reconoce cada vez más en la educación médica como la representación de un principio fundamental de la buena práctica en la evaluación [55]. En nuestra opinión, este concepto debe ser extrapolable a la atención clínica. Falta la evidencia empírica que apoye la mejora de la práctica habitual de los médicos sometidos a evaluaciones. Los resultados del aprendizaje de los médicos pueden ser determinados en función de los cambios observados en la práctica, en lugar de la percepción subjetiva de los cambios por los médicos.

Hay una base para la hipótesis de que la retroalimentación se puede utilizar para mejorar el rendimiento de los médicos. Sin embargo, las conclusiones de los estudios empíricos sobre el impacto de la retroalimentación sobre el desempeño clínico han sido ambiguas. La retroalimentación es importantísima [48], conduce a la formulación de objetivos de aprendizaje [56] y el 61-72% de los médicos informan de un cambio en su comportamiento [57,58].

Veloski et al [59] buscaron en la bibliografía desde 1966 hasta 2003 en Medline, HealthSTAR, el *Science Citation Index* y otras ocho bases de datos electrónicas, y sólo 41 estudios cumplieron todos los criterios de selección y evaluaron el efecto independiente de la retroalimentación sobre el desempeño médico. De ellos, 32 (74%) demostraron un impacto positivo. La retroalimentación era más probable que fuera eficaz cuando la proporcionaba una fuente autorizada durante un período prolongado. Otro subconjunto de 132 estudios examinó el efecto de la retroalimentación en combinación con otras intervenciones como los programas educativos, las guías de práctica y los recordatorios. De éstos, 106 (77%) demostraron un impacto positivo. Se revisaron dos subconjuntos adicionales de 29 estudios de retroalimentación que involucraban médicos residentes en formación y 18 estudios que examinaban medidas sustitutivas de rendimiento de los médicos a través de entornos clínicos o grupos de pacientes. La mayor parte de estos dos subconjuntos también informó de que la retroalimentación

tuvo efectos positivos en el rendimiento. Estos autores concluyen que los efectos de la evaluación formal y la retroalimentación sobre el desempeño médico se ven influidas por el origen y la duración de la retroalimentación. Otros factores, como la participación de los médicos activos en el proceso, la cantidad de información presentada, la sincronización y la cantidad de comentarios, y otras intervenciones simultáneas, como educación, directrices, sistemas de recordatorio y de incentivos, también parecen ser importantes.

Las EBT son ahora una parte establecida de la formación del residente y están aquí para quedarse. Actualmente, hay una significativa incertidumbre entre los residentes y los formadores en cuanto a su función y valor. Carecen de una fuerte base de evidencia sustentatoria, pero esto no significa necesariamente que han fracasado o que son de utilidad limitada; de hecho, los residentes apoyan ampliamente los principios que las sustentan. Es imperativo que los datos recogidos, analizados y publicados sean fiables y precisos acerca de su calidad y validez. Estos hallazgos deberían entonces retroalimentarse y conducir a un cambio evolutivo en la estructura y la aplicación de las EBT. El mensaje claro que emerge es que debe reducirse la función sumativa de las EBT y los aspectos formativos deben ocupar un lugar preponderante. Esto contribuirá a conseguir una verdadera aceptación por parte de profesores y residentes. Nuevas vías para el trabajo futuro serían de gran beneficio práctico si centraran su atención en los programas de evaluación que muestran, de manera conjunta, la utilización de los instrumentos de la EBT. Por último, tenemos que descubrir si las estrategias de evaluación formativa, como la EBT, pueden alcanzar niveles de cambio más altos, lo cual permitiría mejorar la prestación de atención y los resultados del paciente.

Bibliografía

- Shumway JM, Harden RM, Amee G. The assessment of learning outcomes for the competent and reflective physician. *Med Teach* 2003; 25: 569-84.
- Brennan MF, Debas HT. Surgical education in the United States: portents for change. *Ann Surg* 2004; 240: 565-72.
- Smith AJ, Aggarwal R, Warren OJ, Paraskeva P. Surgical training and certification in the United Kingdom. *World J Surg* 2009; 33: 174-9.
- Frank JR, Langer B. Collaboration, communication, management, and advocacy: teaching surgeons new skills through the CanMEDS project. *World J Surg* 2003; 27: 972-8.
- Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach* 2007; 29: 642-7.
- Jones R, Higgs R, De Angelis C, Prideaux D. Changing face of medical curricula. *Lancet* 2001; 357: 699-703.
- Phillips SP, Austin EB. The feminization of medicine and population health. *J Am Med Assoc* 2009; 301: 863-4.
- Davenport DL, Henderson WG, Hogan S, Mentzer RM Jr, Zwischenberger JB. Surgery resident working conditions and job satisfaction. *Surgery* 2008; 144: 332-35.
- Wallenburg I, Van Exel J, Stolk E, Scheele F, De Bont A, Meurs P. Between trust and accountability: different perspectives on the modernization of postgraduate medical training in the Netherlands. *Acad Med* 2010; 85: 1082-90.
- Van der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996; 1: 41-67.
- Morrison J. ABC of learning and teaching in medicine: evaluation. *BMJ* 2003; 326: 385-7.
- Kirkpatrick DL. Evaluating training programs: the four levels. San Francisco, CA: Berrett-Koehler Publishers; 1994.
- Freeth D, Hammick M, Reeves S, Barr H. Critical review of evaluations of interprofessional education. London: Higher Education Academy Learning and Teaching Support Network for Health Sciences and Practice; 2002.
- Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach* 2006; 28: 497-526.
- Cook DA. Twelve tips for evaluating educational programs. *Med Teach* 2010; 32: 296-301.
- McLeod PJ, Steinert Y, Snell L. Use of retrospective pre/post assessments in faculty development. *Med Educ* 2008; 42: 543.
- Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health* 2006; 29: 489-49.
- Grant JS, Davis LL. Selection and use of content experts for instrument development. *Res Nurs Health* 1997; 20: 269-74.
- Arnold L, Willoughby TL. The quarterly profile examination. *Acad Med* 1990; 65: 515-6.
- Dijksterhuis MGK, Schuwirth LWT, Braat DDM, Scheele F. An exploratory study into the impact and acceptability of formatively used progress testing in postgraduate obstetrics and gynaecology. *Perspect Med Educ* 2013; 2: 126-41.
- Watling CJ, Lingard L. Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Adv Health Sci Educ* 2012; 17: 183-94.
- Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011; 33: 478-85.
- Wrigley W, Van der Vleuten CPM, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Med Teach* 2012; 34: 683-97.
- Van Tartwijk J, Driessen EW. Portfolios for assessment and learning: AMEE Guide. *Med Teach* 2009; 31: 790-801.
- Boylan O, Bradley T, McKnight A. GP perceptions of appraisal: professional development, performance management, or both? *Br J Gen Pract* 2005; 55: 544.
- Mathers NJ, Challis MC, Howe AC, Field NJ. Portfolios in continuing medical education –effective and efficient? *Med Educ* 1999; 33: 521-30.
- Campbell C, Parboosingh J, Gondocz T, Babitskaya G, Pham B. A study of the factors that influence physicians' commitments to change their practices using learning diaries. *Acad Med* 1999; 74 (Suppl): 34-6.
- Bruce D, Phillips K, Reid R, Snadden D, Harden R. Revalidation for general practitioners: randomised comparison of two revalidation models. *BMJ* 2004; 328: 687-91.
- Melville C, Rees M, Brookfield D, Anderson J. Portfolios for assessment of paediatric specialist registrars. *Med Educ* 2004; 38: 1117-25.
- Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, et al. The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No. 12. *Med Teach* 2009; 31: 299-318.
- Driessen E, Van Tartwijk J, Van der Vleuten C, Wass V. Portfolios in medical education: why do they meet with mixed success? A systematic review. *Med Educ* 2007; 41: 1224-33.
- Tromp F, Vernooij-Dassen M, Grol R, Kramer A, Bottema B. Assessment of CanMEDS roles in postgraduate training: the

- validation of the Compass. *Patient Educ Couns* 2012; 89: 199-204.
33. Chamberlain NR, Stuart MK, Singh VK, Sargentini NJ. Utilization of case presentations in medical microbiology to enhance relevance of basic science for medical students. *Med Educ Online* 2012; 17: 139-43.
 34. Gonçalves PJA, Rozenfeld LM, Murano ML. Opinión de un grupo de estudiantes de medicina sobre la importancia de la discusión de casos clínicos como técnica didáctica. *Aten Fam* 2011; 18: 56-8.
 35. Kassirer J. Teaching clinical reasoning: case-based coached. *Acad Med* 2010; 85: 1118-24.
 36. Swallow V. Learning in practice: but who learns from who? *Nurse Educ Pract* 2006; 6: 1-2.
 37. National Patient Safety Agency. Seven steps to patient safety. An overview guide for NHS staff. London: NHS; 2004.
 38. Friedman BD, Davis MH, Harden RM, Howie PW, Ker J, Pippard MJ. AMEE Guide No. 24: portfolios as a method of student assessment. *Med Teach* 2001; 23: 535-51.
 39. Association of Surgeons in Training. Response to the ISCP review. London: Association of Surgeons in Training; 2013.
 40. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002; 287: 226-35.
 41. Shalhoub J, Vesey AT, Fitzgerald JEF. What evidence is there for the use of workplace-based assessment in surgical training? *J Surg Educ* 2014; 71: 906-15.
 42. General Medical Council. Academy of Medical Royal Colleges. Workplace based assessment: a guide for implementation. London: GMC; 2010.
 43. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65 (Suppl 9): S63-7.
 44. Pereira EA, Dean BJ. British surgeons' experiences of mandatory online workplace-based assessment. *J R Soc Med* 2009; 102: 287-93.
 45. Munsch C. British surgeons' experiences of mandatory online workplace-based assessment –reply to authors. *J R Soc Med* 2009; 102: 357-8.
 46. Pereira EA, Dean BJ. British surgeons' experiences of a mandatory online workplace based assessment portfolio resurveyed three years on. *J Surg Educ* 2013; 70: 59-67.
 47. Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ* 2010; 341: c5064.
 48. Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multi-source feedback: perceptions of credibility and usefulness. *Med Educ* 2005; 39: 497-504.
 49. Beard J, Marriott J, Purdie H, Rowley DI, Jolly B, Norcini J. Workplace-based assessments as opportunities for learning. Response to: impact of workplace-based assessment on doctors' education and performance: a systematic review. *BMJ* 2010; 341.
 50. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *J Am Med Assoc* 2009; 302: 1316-26.
 51. Woolfolk A, Hughes M, Walkup V. *Psychology in education*. Harlow: Pearson; 2013.
 52. Al Ansari A, Ali SK, Donnon T. The construct and criterion validity of the mini-CEX: a meta-analysis of the published research. *Acad Med* 2013; 88: 413-20.
 53. Overeem K, Faber MJ, Arah OA, Elwyn G, Lombarts KM, Wollersheim HC, et al. Doctor performance assessment in daily practice: does it help doctors or not? A systematic review. *Med Educ* 2007; 41: 1039-49.
 54. Indicadores básicos para el análisis del estado de salud de la población. Fichas técnicas, definiciones, interpretación, cálculo y aplicaciones. La Habana: Dirección Nacional de Registros Médicos y Estadísticas de Salud, Ministerio de Salud Pública de Cuba; 2009.
 55. Handfield-Jones RS, Mann KV, Challis ME, Hobma SO, Klass DJ, McManus IC, et al. Linking assessment to learning: a new route to quality assurance in medical practice. *Med Educ* 2002; 36: 949-58.
 56. Van de Camp K, Vernooij-Dassen M, Grol R, Bottema B. Professionalism in general practice: development of an instrument to assess professional behaviour in general practitioner trainees. *Med Educ* 2006; 40: 43-50.
 57. Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003; 326: 546-8.
 58. Sargeant JM, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Hayes VM, et al. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad Med* 2003; 78 (Suppl 10): 42-4.
 59. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach* 2006; 28: 117-28.