



LECTURA CRÍTICA DE ARTÍCULO

¿Podemos utilizar la simulación para evaluar la formación durante la residencia?

Artículo original: Blum RH, Muret-Wagstaff SL, Boulet JR, Cooper JB, Petrusa ER. Simulation-based assessment to reliably identify key resident performance attributes. *Anesthesiology* 2018; 128:821-31. ([HTML](#))

Albero Roselló I, Belltall Olmos A, Romero García E, Argente Navarro P.

Hospital Universitario y Politécnico La Fe, Valencia.

Resumen

Los médicos residentes (MR) durante el periodo de formación deben adquirir conocimientos y habilidades de forma progresiva para abordar cualquier situación clínica. La simulación clínica es una herramienta que permite entrenar los conocimientos y habilidades adquiridas y aplicarlos a situaciones cuasi reales previamente a entrar en contacto con el paciente.

El objetivo de este estudio consiste en identificar qué competencias dependen del entrenamiento (año de residencia) y son clave en el rendimiento, y proporcionar más datos para mejorar la validación de la escala HARP (Harvard Assessment of Anesthesia Resident Performance), un instrumento de medición que valora el rendimiento de los residentes en diferentes escenarios simulados.

Se realizó un estudio prospectivo observacional donde se seleccionaron cincuenta y un residentes de primer año y dieciséis de tercero de tres programas de formación diferentes. Se diseñaron siete escenarios acordes al nivel formativo donde los residentes asumen el rol de anestesiólogo principal en el desarrollo del caso.

Los resultados de este estudio añaden validez a la evidencia de que la simulación puede servir para evaluar el rendimiento de los residentes, así como encontrar vacíos en las destrezas en áreas seleccionadas, incluyendo algunas que no se encuentran en las evaluaciones clásicas.

Los resultados obtenidos sugieren que la evaluación con el método HARP, si está adecuadamente estructurada, puede proporcionar información detallada de alta calidad sobre las capacidades clínicas de cada residente, con puntuaciones razonablemente precisas y estandarizadas.

Introducción



Los médicos residentes (MR) durante el periodo de formación deben adquirir

conocimientos y habilidades de forma progresiva para abordar cualquier situación clínica. La simulación clínica es una herramienta que permite entrenar los conocimientos y habilidades adquiridas y aplicarlos a situaciones cuasi reales previamente a entrar en contacto con el paciente.

El objetivo de este estudio consiste en identificar qué competencias dependen del entrenamiento (año de residencia) y son clave en el rendimiento, y proporcionar más datos para mejorar la validación de la escala HARP (Harvard Assessment of Anesthesia Resident Performance), un instrumento de

medición que valora el rendimiento de los residentes en diferentes escenarios simulados.

Se realizó un estudio prospectivo observacional donde se seleccionaron cincuenta y un residentes de primer año y dieciséis de tercero de tres programas de formación diferentes. Se diseñaron siete escenarios acordes al nivel formativo donde los residentes asumen el rol de anestesiólogo principal en el desarrollo del caso.

Los resultados de este estudio añaden validez a la evidencia de que la simulación puede servir para evaluar el rendimiento de los residentes, así como encontrar vacíos en las destrezas en áreas seleccionadas, incluyendo algunas que no se encuentran en las evaluaciones clásicas.

Los resultados obtenidos sugieren que la evaluación con el método HARP, si está adecuadamente estructurada, puede proporcionar información detallada de alta calidad sobre las capacidades clínicas de cada residente, con puntuaciones razonablemente precisas y estandarizadas.

Introducción

Se ha demostrado mucha variabilidad en la formación de los médicos residentes (MR) y no siempre se alcanzan los estándares cuando finaliza el periodo de formación. Es importante detectar de forma precoz los problemas formativos durante este periodo a la hora de diseñar planes adaptados a corregir estas brechas.

La Accreditation Council for Graduate Medical Education (ACGME) establece unas competencias básicas que hay que alcanzar de forma progresiva. Para ello es importante disponer de una metodología de evaluación objetiva que incluya también aspectos no técnicos.

La simulación clínica, además de ser útil para entrenar todo tipo de habilidades, ha demostrado su utilidad también para evaluar el desarrollo de competencias globales.

Blum et al, en una publicación previa, realizan la validación de una herramienta basada en simulación: Harvard Assessment of Anesthesia Resident Performance (HARP), un instrumento de medición que valora el rendimiento de los residentes en diferentes escenarios simulados considerando 5 dominios definidos como básicos.

Descripción de Artículo

El objetivo de este artículo consiste en identificar, a través de un programa de simulación, qué competencias dependen del entrenamiento (año de residencia) y son clave en el rendimiento, así como proporcionar más datos para mejorar la validación del HARP. Para la validación se utilizó el método de Shaw adaptado de Kane (Anexo 1).

Se realizó un estudio prospectivo observacional donde se seleccionaron cincuenta y un residentes de primer año y dieciséis de tercero de tres programas de formación diferentes del Harvard Medical School (Boston) acreditados por el ACGME. El estudio se llevó a cabo en tres centros de simulación diferentes, pero con características similares. Utilizando metodología Delphi, con anestesiólogos expertos y profesores universitarios, se eligen cinco dominios clave en la formación de los residentes que deben obtenerse para alcanzar un nivel mínimo de competencia.

1. Sintetiza información para formular un plan anestésico claro.

2. Es capaz de modificar el plan anestésico bajo condiciones cambiantes.
3. Demuestra habilidades efectivas de comunicación con los pacientes y el resto del equipo.
4. Identifica estrategias de mejora.
5. Reconoce sus propios límites personales.

Se diseñan siete escenarios acordes al nivel formativo donde los residentes asumen el rol de anestesiólogo principal en el desarrollo del caso (Anexo 2). Todos los participantes en el estudio como instructores o evaluadores eran anestesiólogos experimentados implicados en programas de formación de residentes que se entrenaron con un programa específico según su rol.

Cada uno de ellos evaluó dos o tres escenarios sin conocer el nivel de formación del residente. No evaluaron a los residentes de sus propios programas de formación. Cada escenario fue puntuado por un solo evaluador, seleccionándose de forma aleatoria un 10% para ser valorados por un segundo, y así calcular la concordancia interevaluador.

Cada residente obtuvo una puntuación, por dominio, por escenario y una puntuación total. La media de los cinco dominios permite calcular la puntuación de cada escenario. La puntuación de cada dominio es la media de ese dominio dentro de cada uno de los escenarios. La puntuación global es la obtenida en la media de los 7 escenarios.

Se considera que los residentes presentan un rendimiento preocupante cuando obtienen una puntuación menor o igual a dos en

alguno de los escenarios o en su totalidad. En la Tabla 1 se describen los comportamientos y la asignación de la puntuación.

Se realizó un análisis estadístico descriptivo para resumir el rendimiento (media \pm DE). Se examinaron las DE de las puntuaciones de los distintos escenarios y dominios para explorar los posibles efectos del entrenamiento. Para respaldar la validez se correlacionaron las puntuaciones de cada dominio. Para evaluar las diferencias en el rendimiento por año de residencia se realizaron seis medidas repetidas ANOVA (ANOVA RM). La medida repetida fue el escenario (7), y la variable dependiente fue la puntuación del escenario o la puntuación de dominio individual. Los tamaños de efecto (ES) se calcularon para cuantificar la magnitud de las diferencias que se encontraron. Se eligió ANOVA como test estadístico por la posibilidad de que las distribuciones no fueran normales.

Para explorar las posibles diferencias en el bajo rendimiento entre los dos grupos, se calcularon los porcentajes de calificaciones 2 o inferiores para cada cohorte (CA-1, CA-3) en cada uno de los siete escenarios por dominio y en general. Se realizó el Chi-cuadrado de Pearson para probar las diferencias entre los grupos.

Los resultados de este artículo muestran que el 96% de los escenarios se completaron, hubo diecisiete pérdidas, cuatro por problemas técnicos y trece por superar el máximo de tiempo permitido.

En cuanto a los datos psicométricos la fiabilidad interevaluador fue del 0.76, la fiabilidad total del test fue de 0.7. Los índices de discriminación de cada escenario fueron altos (≥ 5). El dominio uno estaba altamente correlacionado con el dominio 2 ($r = 0.91$). El dominio

cuatro tuvo las correlaciones más bajas con los otros dominios (promedio $r = 0.59$). El dominio 3 mostró correlaciones modestas con el dominio 4 ($r = 0.57$) y el dominio 5 ($r = 0.64$).

Comparando a los residentes de primer y tercer año, el test RM-ANOVA no mostró diferencias significativas entre años. Estadísticamente, los residentes de tercer año superaron a los de primer año para los escenarios 3, 4, 5, 6 y 7. Ningún escenario tuvo una puntuación total promedio superior a 5. Las puntuaciones mínimas y máximas de cada escenario indican que hubo puntuaciones bajas y altas en ambas cohortes de residentes. Además, algunos residentes de tercer año tuvieron puntuaciones más bajas que algunos residentes de primer año. La tabla 1 muestra las puntuaciones para cada uno de los cinco dominios y la puntuación total según el año de residencia.

	Dominio	Año residencia	Media	DE	Baja	Alta	
1	R-1	4.5	1.3	3.0	6.0		
	R-3	5.0	1.3	4.4	5.7		
2	R-1	4.4	1.5	2.6	6.3		
	R-3	5.1	1.3	4.4	6.2		
3	R-1	4.7	1.4	2.3	6.4		R-1: Residentes de primer año
	R-3	5.0	1.4	4.3	5.9		
4	R-1	4.0	1.3	3.2	6.0		
	R-3	4.7	1.4	3.6	6.3		R-3: Residentes de tercer año
5	R-1	4.6	1.5	2.5	5.8		
	R-3	4.8	1.4	3.1	6.1		
Total score	R-1	4.6	1.6	3.1	6.8		
	R-3	4.9	1.4	4.1	6.0		

Tabla 1: Puntuación total y por dominios para cada año de residencia.

En cuanto al bajo rendimiento, contando cualquier calificación menor o igual a dos como preocupante, los R1 tuvieron mayor proporción de puntuaciones preocupantes que los R3. El escenario 3 presentó el mayor número de calificaciones menores o iguales a 2. El escenario 6 muestra mayor porcentaje de puntuaciones menores o iguales a 2 para los residentes de primer año, mientras que para los de tercer año fue el mejor escenario.

Análisis crítico del Artículo

Este estudio intenta responder a la pregunta de si la simulación es útil para evaluar los conocimientos y habilidades de los residentes durante su periodo

formativo, y si esta evaluación se puede realizar de forma estandarizada.

Los resultados de este estudio añaden validez a la evidencia de que la simulación puede servir para evaluar el rendimiento de los residentes, así como encontrar vacíos en las destrezas en áreas seleccionadas, incluyendo algunas que no se encuentran en las evaluaciones clásicas.

La población a estudio fue 67 residentes, sin conocer los criterios de selección de los mismos y siendo la distribución asimétrica (51 R1, 16 R3). Se seleccionaron de 3 programas formativos locales, con lo que la generalización puede ser difícil. A pesar de ello, cada sujeto fue tratado de la misma forma durante la simulación. Cada uno de ellos proviene de hospitales con programas formativos en simulación ya establecidos, por lo que pueden estar acostumbrados a este método de trabajo y de ahí las escasas diferencias entre los R1 y los R3.

Los programas de formación de residentes son propios de cada región, pudiendo presentar diferencias sustanciales. Aunque los escenarios simulados se desarrollaron para representar situaciones reales de atención del paciente, el vínculo entre el rendimiento en el entorno simulado y las situaciones reales de atención del paciente aún no se ha establecido con firmeza.

En cuanto a la medición de resultados se utilizan variables subjetivas que dependen de la valoración del instructor, aunque el cociente de fiabilidad interevaluador fuera elevado. En un estudio previo de estos mismos autores se demuestra la validez de las variables utilizadas.

Los resultados obtenidos sugieren que la evaluación con el método HARP, si

está adecuadamente estructurada, puede proporcionar información detallada de alta calidad sobre las capacidades clínicas de cada residente, con puntuaciones razonablemente precisas y estandarizadas. A nivel individual estas evaluaciones pueden informar sobre mejoras específicas y a nivel del programa educativo pueden guiar el desarrollo curricular de habilidades conductuales consideradas críticas para la práctica segura de la anestesia.

Con respecto a la generalización, el diseño del estudio puede ser valorado como representativo del ámbito anestésico.

Se necesita más trabajo para abordar dos elementos importantes del marco de validez: (1) extrapolación: ¿son los escenarios seleccionados representativos de la competencia en dominios temáticos más amplios, por ejemplo, en entornos clínicos reales? Y (2) toma de decisiones: ¿existe una guía para que las partes interesadas sepan qué significan las puntuaciones y cómo deben usarse los resultados? Para la mayoría de las evaluaciones basadas en la simulación, se ha trabajado poco para vincular el rendimiento en simulación con el de los entornos clínicos reales.

A medida que el campo de la evaluación basada en la simulación madure, los escenarios basados en simulación con métricas de evaluación apropiadas deberían permitir a los directores de los programas de anestesiología evaluar las habilidades de los residentes con más confianza y, cuando sea necesario, proporcionar intervenciones educativas que podrían mejorar la calidad del rendimiento de sus especialistas.

Los resultados obtenidos en el estudio podrían aplicarse a nuestro medio utilizando la herramienta de la simulación clínica como un elemento

más de evaluación para la formación de los residentes.

Conclusiones

Creemos que la simulación clínica asociada a una buena herramienta de evaluación, como la utilizada en este estudio debería ser aplicada en los programas formativos de anestesiología y reanimación de nuestro entorno, para así mejorar la calidad asistencial en pro de la seguridad de nuestros pacientes.

Bibliografía

Blum RH, Boulet JR, Cooper JB, Muret-Wagstaff SL. Simulation-based assessment to identify critical gaps in safe anesthesia resident performance. *Anesthesiology* 2014; 120:129-41. ([PubMed](#))

ANEXO 1

Los cinco argumentos utilizados para validar la escala de evaluación HARP fueron:

1. Validez de la representación: ¿los escenarios reflejan la intención del estudio (seguridad durante la práctica anestésica)?
2. Puntuación: ¿La calificación es fiable en cuanto a sus medidas?
3. Generalización: ¿Permiten los escenarios mostrar adecuadamente los ítems definidos como importantes?
4. Extrapolación: ¿Reflejan los escenarios de simulación situaciones reales de los pacientes?
5. Toma de decisión: ¿Existe una guía para los participantes sepan que significan las puntuaciones y cómo se utilizarán los resultados?

ANEXO 2

1. Valoración preoperatoria en paciente sometido a laparotomía exploradora urgente.
2. Manejo intraoperatorio del paciente con hemorragia por úlcera.
3. Cuidado anestésico monitorizado en el paciente con disconfort durante cirugía de carcinoma basocelular.
4. Cuidado postanestésico en paciente con broncoaspiración tras cirugía de resección tumoral.

5. Manejo de la anafilaxia en paciente sometido a RTU y biopsia vesical.
6. Cuidado del paciente con retraso del despertar tras RTU.
7. Identificación y manejo de la situación de intubación selectiva secundaria a la tos en paciente sometido a tiroidectomía total.

Correspondencia al autor

Ignacio Albero Roselló
nachoalbero@gmail.com

*Residente Anestesiología y Reanimación.
Hospital Universitario y Politécnico La Fe,
Valencia.*

Eva Romero García
Adjunta Anestesiología y Reanimación.
*Hospital Universitario y Politécnico La Fe,
Valencia.*

Aceptado para el blog en noviembre de 2019

