

Consenso de expertos sobre evaluación de estudiantes y residentes a partir de la Ottawa Conference 2020

Expert consensus on student and resident assessment from the 2020 Ottawa Conference

Jordi PALÉS-ARGULLÓS, Amando MARTÍN-ZURRO

En 2011, la revista *Medical Teacher* publicaba un documento que exponía un consenso de expertos sobre la evaluación del rendimiento de estudiantes y residentes de Medicina generado en el seno de la Ottawa Conference [1], foro dedicado fundamentalmente a discutir sobre la evaluación en educación médica. Nueve años más tarde, se ha publicado también en *Medical Teacher* un nuevo artículo, titulado 'Performance assessment: consensus statement and recommendations from the 2020 Ottawa Conference' [2], en el que se revisa el documento de 2011, analizando los aspectos que no se han resuelto de forma totalmente satisfactoria en este período de tiempo y poniendo el foco en dos temas concretos, las evaluaciones clínicas objetivas estructuradas (ECOEs) y la evaluación del desempeño en el lugar de trabajo. A partir del documento inicial de 2011, y a la vista de los desarrollos surgidos en evaluación en estos últimos 10 años, los expertos exponen una serie de recomendaciones para establecer buenas prácticas en el uso de estos dos instrumentos.

Dado que todas nuestras facultades de medicina están utilizando las ECOEs de forma generalizada para evaluar a sus estudiantes y que se ha hablado en ocasiones de una futura utilización en el examen de acceso a la formación especializada, quisiéramos, en este editorial, centrar nuestra atención en dicho instrumento, comentando brevemente las recomendaciones que este grupo de expertos aporta para facilitar su buen uso en la práctica. El artículo aborda tres aspectos principales: la necesidad de que se definan con claridad los objetivos y propósitos de su utilización, de que haya evidencias significativas para apoyar y justificar las decisiones tomadas a partir de los resultados obtenidos en la ECOE y, finalmente, de que existan suficientes argumentos o justificaciones para defender las decisiones tomadas en función de los resultados obtenidos.

En el primer aspecto se señala que el propósito de la ECOE debería explicitarse a todas las partes inte-

In 2011, the journal Medical Teacher published a paper outlining an expert consensus on the performance assessment of medical students and residents generated at the Ottawa Conference [1], a forum dedicated primarily to discussing assessment in medical education. Nine years later, a new article has been published again in Medical Teacher entitled 'Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference' [2], which revisits the 2011 document, analysing the aspects that have not been resolved in a totally satisfactory manner in the time elapsed since its publication and focusing on two specific topics, objective structured clinical examinations (OSCE) and workplace-based assessment. Building on the initial 2011 document, and in the light of developments made in assessment over the last 10 years, the experts set out a number of recommendations for establishing good practice in the use of these two tools.

Given that all our medical schools are using OSCEs in a generalised way to assess their students and that there has sometimes been talk of their possible use in the future in the entrance exam to specialised training, in this editorial we would like to focus our attention on this instrument, briefly commenting on the recommendations that this group of experts propose with the aim of facilitating its proper use in practice. The article addresses three main aspects: the need for a clear definition of the intended use and purpose, for there to be significant evidence to support and justify the decisions taken on the basis of the results obtained in the OSCE and, finally, for there to be sufficient arguments or justifications to defend the decisions taken on the basis of the results obtained.

The first aspect states that the purpose of the OSCE should be made explicit to all the stakeholders, such as teachers, candidates, examiners, em-

Fundación Educación Médica

E-mail:
jpales@ub.edu

© 2021 FEM



Artículo open access bajo la licencia CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

ISSN: 2014-9832
ISSN (ed. digital): 2014-9840

resadas, como, por ejemplo, profesores, candidatos, examinadores, empleadores, reguladores, organismos y el público.

En el segundo aspecto, el más extensamente desarrollado, los autores hacen hincapié en diferentes puntos: en el contenido de la ECOE (asegurar que dicho contenido evalúe adecuadamente lo que se pretende evaluar), en la planificación de una ECOE en función de los objetivos de aprendizaje para que realmente nos permita evaluar aquello para lo que está pensada, es decir, habilidades clínicas, de comunicación, de procedimientos técnicos y de razonamiento clínico; en este sentido, nos recuerdan que no se deben incluir en una ECOE exámenes de conocimientos en diferentes formatos, ya que ello amenazaría la validez de la prueba. También se refieren a los requerimientos de la estructura interna que debe cumplir una ECOE, como tener el número adecuado de estaciones (no menos de 12) y su duración (no menos de 150 minutos); disponer de un número adecuado de observadores y de un amplio y diverso mapa de situaciones (recordemos que la competencia no es genérica, sino específica); tener un sistema de puntuación adecuado (listas de comprobación frente a escalas globales de valoración); establecer adecuadamente los niveles de superación del examen utilizando los mejores métodos para ello; y realizar un análisis psicométrico adecuado para asegurar la confiabilidad de la prueba. Además, reflexionan sobre si es necesario adoptar medidas de seguridad para evitar que los examinandos conozcan con antelación las tareas que se les van a solicitar y sobre la necesidad de comparar los resultados de las ECOE con los de otras pruebas mediante una triangulación de los datos obtenidos.

Los expertos abogan también por incluir las ECOE en un enfoque programático de la evaluación, al considerar que las decisiones deberían tomarse no después de cada episodio de evaluación, sino por dominio o competencia en lugar de por instrumento de evaluación. Es decir, defienden la necesidad de combinar la ECOE con el uso de diferentes instrumentos de evaluación que valoren la competencia desde diferentes perspectivas. Finalmente, en este apartado, insisten en la necesidad inexcusable de que las decisiones que se tomen según los resultados del examen sean sólidas, justas y defendibles, lo que exige que los criterios de superación de éste se establezcan utilizando los métodos mejor evidenciados y facilitando obligatoriamente un *feedback* al estudiante con el fin de incrementar su motivación y aprendizaje.

En el tercer apartado, los expertos ven indispensable recopilar todas las fuentes de evidencia de va-

ployers, regulators, agencies and the public.

In the second aspect, the most extensively developed, the authors stress different points: the content of the OSCE (ensuring that the content adequately assesses what it is intended to assess), the planning of an OSCE according to the learning objectives so that it really allows us to assess what it is intended for, that is, clinical skills, communication, technical procedures and clinical reasoning. In this regard, they remind us that we should not include knowledge tests in different formats in an OSCE, as this would jeopardise the validity of the test. They also refer to the requirements in terms of the internal structure that an OSCE must fulfil, such as having an adequate number of stations (no fewer than 12) and their length (no less than 150 minutes); having an adequate number of observers and a wide and diverse map of situations (remember that the competence is specific, rather than generic); having an adequate scoring system (checklists versus global assessment scales); establishing the pass levels of the exam in an appropriate way using the best methods to do so; and carrying out an adequate psychometric analysis to ensure the reliability of the test. They also reflect on the need for security measures to prevent test takers from knowing in advance the tasks they will be asked to perform and on the need to compare the results of OSCEs with those of other tests by triangulating the data obtained.

The experts also advocate the inclusion of OSCEs in a programmatic approach to assessment, considering that decisions should be made not after each assessment episode, but by domain or competence rather than by assessment instrument. In other words, they advocate the need to combine the OSCE with the use of different assessment instruments that rate competence from different perspectives. Finally, in this section, they insist on the inexcusable need for the decisions taken on the basis of the results of the exam to be solid, fair and defensible, which requires that the criteria for passing the exam be established using the best-evidenced methods and providing students with mandatory feedback in order to increase motivation and learning.

In the third section, the experts consider it essential to collect all sources of validity evidence to justify any decision taken on the basis of the test results.

By way of summary, the article concludes with a number of recommendations: 1) design OSCEs in the context of a more global assessment system; 2) adhere to validity criteria; 3) define the purpose of

lidez para justificar cualquier decisión tomada basándose en los resultados de la prueba.

A modo de resumen, el artículo finaliza con unas recomendaciones que enumeramos a continuación: 1) diseñar las ECOE en el contexto de un sistema de evaluación más global; 2) adherirse a los criterios de validez; 3) definir el propósito de la ECOE y hacerlo explícito a las partes interesadas; 4) asegurar la correspondencia entre lo que evaluamos y los resultados del aprendizaje; 5) planificar un muestreo adecuado con suficientes estaciones y tiempo de prueba; 6) utilizar la configuración estándar basada en criterios pertinentes; 7) generar estudios psicométricos para asegurar la confiabilidad global y de estación; 8) valorar la diversidad de los examinadores y centrar su formación en la conducta, los comportamientos y los prejuicios; 9) manejar la seguridad de las pruebas con el diseño de tareas y de circuitos para agrupar estaciones; y 10) triangular los datos del desempeño de la ECOE con otras evaluaciones o resultados para informar de la toma final de decisiones.

El artículo analiza también la evaluación del desempeño en el lugar de trabajo y las diferentes connotaciones que adquiere según su objetivo sea la valoración formativa o la evaluación sumativa, esencialmente en lo que se refiere a *feedback* y a la interacción con los tutores del aprendizaje. Los métodos para la evaluación del desempeño en el lugar de trabajo incluyen diferentes instrumentos centrados en el análisis de encuentros directos o indirectos con uno o más pacientes y deben usarse de forma prudente, intentando minimizar los problemas de interpretación que, sobre todo en el campo psicométrico, pueden aparecer al utilizarlos desde la perspectiva de la evaluación sumativa.

Analizar con detenimiento estas recomendaciones tiene, en nuestra opinión, una gran importancia en España, sobre todo en un momento como el actual, en el que se están planteando cambios relevantes en la formación de estudiantes y residentes. Asumir que las ECOE pueden ser un instrumento nuclear para evaluar el período de grado y para contribuir a definir criterios para la transición con el posgrado necesita la implicación decidida de todos los actores implicados en la corrección de los problemas que pueden tener en su concepción y desarrollo actuales. Las instituciones y organismos responsables de los procesos de formación graduada y posgraduada, con especial referencia a nuestras facultades de medicina, han de analizar críticamente y, en su caso, aplicar las recomendaciones apuntadas para avanzar en el diseño del mejor sistema de evaluación posible de nuestros estudiantes y residentes.

the OSCE and make it explicit to stakeholders; 4) ensure there is a correspondence between what we assess and learning outcomes; 5) plan adequate sampling with sufficient stations and testing time; 6) use standard settings based on relevant criteria; 7) generate psychometric studies to ensure reliability at the global and station levels; 8) value examiner diversity and focus their training on conduct, behaviours and bias; 9) handle test security through task design and circuit design to group stations; and 10) triangulate data from OSCE performance with other assessments or outcomes to inform final decision-making.

The article also discusses workplace-based assessment and the different connotations it can have depending on whether its purpose is formative or summative assessment, essentially in terms of feedback and interaction with learning tutors. Methods for workplace-based assessment include different instruments focused on the analysis of direct or indirect encounters with one or more patients. These tools should be used carefully, striving to minimise the interpretation problems that, especially in the psychometric field, may arise when using them from the perspective of summative assessment.

A careful analysis of these recommendations is, in our opinion, of great importance in Spain, especially at a time like the present, when significant changes are being considered in the training of students and residents. Assuming that OSCEs can be a core instrument to assess the undergraduate period and to contribute to defining criteria for the transition to postgraduate studies requires the decisive involvement of all the stakeholders involved in correcting the problems they may have in their current conception and development. The institutions and bodies responsible for undergraduate and postgraduate training processes, with special reference to our medical schools, must critically analyse and, where appropriate, apply the above recommendations in order to advance in the design of the best possible assessment system for our students and residents.

Bibliografía / References

1. Boursicot K, Etheridge L, Setna Z, Sturrock A, Ker J, Smees S, et al. Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Med Teach* 2011; 33: 370-83.
2. Boursicot K, Kemp S, Wilkinson T, Findyartini A, Canning C, Cilliers F, et al. Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference. *Med Teach* 2021; 43: 58-67.