

¿Ha pasado el tiempo de los exámenes de elección múltiple?

Jordi PALÉS-ARGULLÓS

La evaluación de los estudiantes y de los profesionales de las ciencias de la salud ha sido sin duda uno de los ámbitos de la educación médica en el que más se ha desarrollado investigación educativa. Así, en base a las evidencias que esta investigación aporta, han ido apareciendo en las diferentes épocas múltiples instrumentos de evaluación, cada uno con sus indicaciones, sus puntos fuertes y sus debilidades.

Uno de los instrumentos que siempre ha tenido y sigue teniendo un gran predicamento es el test de elección múltiple, en el cual el examinando, ante una pregunta o una situación, debe elegir la opción correcta o la más adecuada entre varias opciones.

Los exámenes de elección múltiple fueron introducidos a principios del siglo xx por Benjamin Wood (1894-1986), educador, investigador y profesor de la Universidad de Columbia en Estados Unidos, pionero en tecnologías de aprendizaje y métodos de evaluación automatizados que se aplicaron por primera vez en los tests de inteligencia realizados por el ejército estadounidense en la Primera Guerra Mundial. Así mismo, contribuyó al diseño del primer lector automatizado de los resultados de dichas pruebas.

En nuestro país, como estudiante de medicina, recuerdo haber realizado por primera vez un examen de este tipo en la facultad allá por los años setenta, y luego haberlos empleado en mi docencia desde esa época hasta la actualidad. Hoy día, podemos afirmar que es un instrumento ampliamente utilizado en los centros docentes.

Este instrumento nace de la necesidad de evaluar de forma simultánea a un número importante de examinandos con el máximo grado posible de objetividad y fiabilidad. Reúne las ventajas de su factibilidad, ya que es fácil de administrar y corregir mediante lectores automáticos, y de su fiabilidad, siempre y cuando esté bien diseñado. Precisamente es aquí donde radica uno de sus puntos débiles, pues-

Are multiple-choice exams a thing of the past?

The assessment of students and professionals in the health sciences has undoubtedly been one of the areas of medical education in which the most educational research has been conducted. Thus, on the basis of the evidence provided by that research, a number of assessment instruments have appeared at different times, each with its own indications, strengths and weaknesses.

One of the instruments that has always enjoyed and still enjoys great prestige is the multiple choice test, in which the examinee, faced with a question or a situation, must choose the right or the most appropriate answer among several options.

Multiple-choice tests were introduced in the early 20th century by Benjamin Wood (1894-1986), an educator, researcher and professor at Columbia University in the United States, who was a pioneer in automated learning technologies and assessment methods that were first applied in the intelligence tests conducted by the US military in World War I. He also contributed to the design of the first automated reader of the results of these tests.

In our country, as a medical student, I remember taking that kind of exam for the first time in the medical school back in the 1970s, and then using them in my teaching ever since. Today, we can say that it is an instrument that is widely used in educational centres.

This method of assessment arises from the need to evaluate a large number of candidates at the same time and with the highest possible degree of objectivity and reliability. It combines the advantages of its feasibility, since it is easy to administer and mark by means of automatic readers, and its reliability, provided that it is well designed. This is precisely where one of its weak points lies, since the rules are not always followed to produce these exercises in the most appropriate way.

Like any other assessment instrument, it has been criticised and has its weaknesses. In the field of health

Fundación Educación Médica.

E-mail:
jpales@ub.edu

© 2019 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

to que no siempre se siguen las normas para elaborar estos ejercicios de la forma más adecuada.

Como cualquier otro instrumento de evaluación, no está exento de críticas ni debilidades. En el ámbito de las ciencias de la salud, una de las críticas que se le hace es no evaluar de manera suficientemente representativa el amplio abanico de las actividades intelectuales que se asume que deben demostrar los profesionales de la salud, y por otra parte, tener un papel limitado a la medida del conocimiento factual y a la capacidad de aplicación, aunque sí es cierto que puede medir un espectro amplio del conocimiento.

A pesar de estos puntos débiles, se ha llegado a convertir con el tiempo en uno de los instrumentos de evaluación más utilizados y en un componente principal de numerosos procesos de certificación, y ha estado destinado a proporcionar evidencia de que se ha alcanzado un nivel adecuado de competencia.

Sin embargo, en los últimos años, entre los expertos en educación médica han empezado a surgir voces que apuntarían a que este tipo de instrumento debería perder algo de su importancia en beneficio de otros que podrían ser más adecuados.

En este sentido, en su último blog, Ronald Harden [1], bajo el sugerente encabezamiento '*MCQs are dead*', aboga por un cambio sustancial con respecto a los instrumentos que utilizamos en las evaluaciones. Hace suya la afirmación de John Cookson, profesor de la Hull York Medical School, que argumentaba al final de un curso organizado por la Association for Medical Education in Europe (AMEE) que los MCQ (*multiple choice questions*) habían tenido su época y que ahora se debería pasar a utilizar en su lugar otros instrumentos, especialmente el conocido como VSAQ (*very short answer questions*) o 'preguntas de respuesta muy corta'.

Harden afirma que la evidencia actual es clara: las preguntas de respuesta muy corta, en las que el alumno contesta con una, dos o un número muy limitado de palabras en lugar de seleccionar una respuesta de una lista de opciones, como ocurre en un test de elección múltiple, ofrecen grandes ventajas. Considera que son una herramienta de evaluación más válida y que refleja el proceso de decisión clínica sin las indicaciones inherentes al test de elección múltiple porque al profesional, en su práctica diaria, nunca se le ofrece una lista de posibles opciones para elegir entre diagnósticos o tratamientos. Además, se elimina la posibilidad de responder a un elemento simplemente por casualidad. Por ello, Harden considera que las preguntas de respuesta muy corta son más desafiantes y proporcionan una mejor discriminación que el test de elección múltiple.

sciences, one of the criticisms made is that it does not evaluate in a sufficiently representative way the wide range of intellectual activities that health professionals are supposed to demonstrate. Other critics say that it has a limited role in the measurement of factual knowledge and capacity for application, although it is true that it can measure a broad spectrum of knowledge.

Despite these weaknesses, over the years it has become one of the most widely used assessment tools and a major component of numerous certification processes, and has been designed to provide evidence that an adequate level of competence has been achieved. However, in recent years, among experts in medical education voices have begun to emerge that this type of instrument should lose some of its importance to the benefit of others that might be more appropriate.

In this sense, in his last blog, Ronald Harden [1], under the suggestive heading 'MCQs are dead', advocates a substantial change with respect to the instruments we use in assessment. He endorses the statement by John Cookson, professor at Hull York Medical School, who argued at the end of a course organised by the Association for Medical Education in Europe (AMEE) that MCQ (multiple choice questions) have had their heyday and now other instruments should be used instead, especially the one known as VSAQ (very short answer questions).

Harden says the current evidence is clear: very short answer questions, in which the student answers with one, two or a very limited number of words rather than selecting an answer from a list of options as in a multiple choice test, offer great advantages. He considers that they are a more valid assessment tool and that they reflect the clinical decision process without the inherent indications of the multiple choice test because the professional, in his or her daily practice, is never offered a list of possible options from which to choose a diagnosis or treatment. In addition, the possibility of answering correctly simply by chance is eliminated. For this reason, Harden believes that very short answer questions are more challenging and provide better discrimination than the multiple choice test.

Recently, some original articles comparing the two assessment instruments concluded that very short answer questions show high reliability and a higher degree of discrimination and validity than multiple choice tests. In addition, the items are perceived by examinees as more authentic [2,3].

However, the implementation of very short answer questions is not without its drawbacks, especially when it comes to correcting them, which could

Recientemente, algunos artículos originales que comparan ambos instrumentos de evaluación concluyen que las preguntas de respuesta muy corta muestran una alta fiabilidad y un mayor grado de discriminación y validez que los exámenes de elección múltiple. Además, los ítems son percibidos por los examinandos como más auténticos [2,3].

Sin embargo, la implementación de las preguntas de respuesta muy corta no está exenta de inconvenientes, especialmente en su corrección, que podrían dificultar su aplicación ante un número muy considerable de estudiantes.

En cualquier caso, ante estas evidencias, quizás sería conveniente que todos aquellos que estamos implicados en procesos de evaluación de estudiantes y profesionales de las ciencias de la salud tuviéramos en cuenta esta posibilidad y analizáramos la conveniencia y la posibilidad de utilizar este instrumento en nuestras actividades docentes.

make them difficult to apply with a very large number of students.

In any case, in view of this evidence, it would perhaps be advisable for all of us who are involved in processes of assessing students and professionals in the health sciences to take this possibility into account and analyse the convenience and possibility of using this instrument in our teaching activities.

Bibliografía / References

1. Harden RM. Harden's blog: News from Roanoke, MCQs are dead, Learning analytics and ethical issues, Goodhart's law and citation numbers, and Inspiring students (or trying to). URL: <https://www.mededworld.org/hardens-blog/reflection-items/July-2019/HARDEN-S-BLOG-News-from-Roanoke--MCQs-are-dead--Le.aspx>. [05.07.2019].
2. Sam AH, Hameed S, Harris J, Meeran K. Validity of very short answer versus single best answer questions for undergraduate assessment. *BMC Med Educ* 2016; 16: 266.
3. Sam AH, Field SM, Collares CF, Van der Vleuten CPM, Wass VJ, Melville C, et al. Very-short-answer questions: reliability, discrimination and acceptability. *Med Educ* 2018; 52: 447-55.