

Efecto del orden de las preguntas de selección múltiple en los resultados de exámenes de cardiología

Effect of order of multiple-choice questions on the results of cardiology exams

AMANDA GALLI[✉], MIRTA DANIELA LOBIANCO^{MTSAC}, SANDRA SWIESZKOWSKI^{MTSAC},[✉] LUCÍA KAZELIAN^{MTSAC},[✉] GRACIELA GIMENO, SUSANA LAPRESA^{MTSAC}, MARISA PAGES^{*MTSAC}, HUGO GRANCELLI^{MTSAC}

RESUMEN

Introducción: Algunos autores han señalado que el armado de un examen con preguntas aleatorias puede perjudicar el rendimiento de los estudiantes.

A partir de la pandemia por COVID -19 y del aislamiento social obligatorio, las actividades de educación médica de posgrado pasaron a la modalidad virtual y los exámenes se implementaron *online*.

Objetivo: Estudiar si el ordenamiento al azar de las preguntas tiene algún efecto en los resultados de los exámenes.

Material y métodos: Se redactaron 2 exámenes: uno sobre Prevención de la enfermedad cardiovascular y otro sobre Valvulopatías. Para cada uno de los temas se confeccionaron dos cuestionarios: una versión con las preguntas en el orden lógico y según dificultad estimada, y otra con las mismas preguntas distribuidas al azar. Cada examen tenía 50 preguntas de selección múltiple con 3 opciones. Puntaje máximo posible: 50

Los exámenes fueron administrados en plataforma Moodle, modalidad sincrónica, tiempo disponible 75 minutos.

Los resultados se expresaron en puntajes obtenidos (rango y valores de tendencia central) y según el índice de dificultad de las preguntas.

Resultados: Respondieron 284 residentes, alumnos del Curso Biannual de Cardiología. Dos grupos de 1° año (tema Prevención) y dos grupos de 2° año (Valvulopatías). No hubo diferencia entre los resultados de las dos versiones del mismo examen.

Conclusiones: Este estudio careció de la potencia suficiente para fundamentar una u otra forma de ordenar las preguntas de los exámenes de ciencias clínicas. Surgieron nuevas preguntas que deberán ser respondidas en futuros estudios.

Palabras clave: Preguntas de selección múltiple - Preguntas aleatorias - Evaluación de conocimientos

ABSTRACT

Background: Some authors have pointed out that setting up an exam with random questions can impair student performance. Since the COVID-19 pandemic and compulsory social isolation, postgraduate medical education activities became virtual and exams were implemented online.

Objective: The aim of the study was to analyse whether the random order of questions has any effect on test results.

Method: Two exams were written: one on the Prevention of Cardiovascular Disease and another on Heart Valve Diseases. For each topic, two questionnaires were designed: a version with the questions in logical order and according to the estimated difficulty and another with the same questions randomly distributed. Each exam had 50 multiple-choice questions with 3 options, and the maximum possible score was 50

The exams were taken on the Moodle platform, with synchronous modality, and the time available was 75 minutes.

The results were expressed as scores obtained (range and central tendency) and according to the index of difficulty of the questions.

Results: The number of respondents was 284 residents, students of the Biannual Cardiology Course: two 1st year groups (Prevention topic) and two 2nd year groups (Valve diseases). There was no difference between the results of the two versions of the same exam.

Conclusions: This study lacks sufficient power to support either way of question order in clinical sciences exams, giving rise to new queries.

Key words: Multiple choice questions - Random questions - Knowledge evaluation

REV ARGENT CARDIOL 2022;90:141-145. <http://dx.doi.org/10.7775/rac.es.v90.i2.20503>

Recibido: 22/10/2021 - Aceptado: 11/02/2022

Dirección para separatas: Amanda Galli - Juncal 2388. CABA - E-mail: amandaelisagalli@gmail.com

INTRODUCCIÓN

En la literatura sobre Evaluación de los Aprendizajes, (1-4) específicamente en el tema de construcción de exámenes estructurados, se hacen indicaciones para asegurar que el instrumento de evaluación brinde resultados válidos y confiables.

El National Board of Medical Examiners (NBME) ha publicado muchas recomendaciones sobre cómo construir exámenes en el área de ciencias básicas y clínicas. Insiste en que se deben explorar contenidos relevantes más que el recuerdo de datos puntuales y que se debe preguntar sobre conductas a seguir en situaciones clínicas. Recomienda que todas las opciones de respuestas sean homogéneas. (5)

Varios autores han explorado cuántas opciones debe tener una pregunta de selección múltiple y han demostrado que en el área de ciencias de la salud, 3 opciones resultan suficientes. (6,7)

Diversas investigaciones han evaluado el impacto del ordenamiento de las preguntas en el rendimiento de los estudiantes. Algunas publicaciones afirman que los estudiantes tienen mejor rendimiento cuando las preguntas se presentan en orden de dificultad creciente, y que una distribución al azar de las preguntas perjudica a los estudiantes, mientras que otras dicen que la secuencia de las preguntas no incide en el rendimiento (8-11)

En el curso Bianual de Cardiología que se desarrolla en la Sociedad Argentina de Cardiología (SAC), anualmente se diseñan y aplican 8 exámenes de 50 preguntas cada uno, lo que representa una tarea de

elaboración y revisión de más de 400 preguntas por año que son respondidas por unos 300 residentes que se están formando como especialistas en Cardiología en programas de residencias médicas en distintos hospitales.

A partir de la pandemia por COVID -19 y del aislamiento social obligatorio, las actividades de enseñanza pasaron a la modalidad virtual y los exámenes fueron respondidos *online*. La disponibilidad actual de ciertos recursos (*softwares*) para la administración y corrección de cuestionarios permite construir varias versiones de un mismo examen simplemente mezclando las preguntas sin respetar ningún criterio de secuencia lógica o temática, totalmente al azar. El armado de exámenes individualizados, con preguntas aleatorias apareció como una alternativa para hacer difícil un potencial fraude utilizando dispositivos electrónicos móviles.-

Este trabajo se propuso estudiar si el ordenamiento al azar de las preguntas tiene algún efecto en los resultados de los exámenes rendidos por los residentes que estaban participando en el Curso Bianual de Cardiología.

MATERIAL Y MÉTODOS

Los docentes de primero y de segundo año del Curso Bianual de Cardiología de la SAC diseñaron las preguntas para dos exámenes: Prevención de la Enfermedad Cardiovascular (1° año) y Valvulopatías (2° año). Cada examen era de 50 preguntas de selección múltiple con 3 opciones: una respuesta correcta y dos distractores. A continuación, se presentan 4 preguntas a modo de ejemplo. El asterisco (*) señala en cada caso la respuesta correcta

Enunciado/ pregunta	Preguntas de Prevención	Opciones
<ul style="list-style-type: none"> - ¿Cuál es la conducta a seguir si un paciente refiere dolor muscular y se sospecha que es un efecto adverso del tratamiento con estatinas? - Paciente masculino de 41 años, deportista, con LDL de 195 mg/dl y riesgo a diez años bajo. Además de dieta ¿qué tratamiento inicia porque ha demostrado mayor beneficio? 	<ul style="list-style-type: none"> a) Suspender el tratamiento con estatinas y solicitar dosaje de CPK y TSH * b) Continuar con el tratamiento sin hacer ningún cambio c) Continuar con estatinas hasta recibir los resultados de CPK y TSH a) AAS 100 mg/d b) Rosuvastatina 40 mg/día * c) Ezetimibe 10 mg/día 	
Preguntas de Valvulopatías		
<ul style="list-style-type: none"> - ¿Qué mecanismos participan en el desarrollo de una insuficiencia mitral funcional o secundaria? - Consulta una paciente de 34 años por palpitaciones. Ritmo sinusal en la consulta. Trae un Holter con 8 episodios de taquiarritmia supraventricular de breve duración, asintomáticos. Ecocardiograma Doppler: ventrículo izquierdo con diámetros conservados, aurícula izquierda con volumen marcadamente aumentado de 58 ml/m² y presión pulmonar estimada en 30 mmHg. Leve calcificación comisural, sin afectación del aparato subvalvular mitral y buena movilidad valvar. Área valvular mitral de 1,2 cm². La paciente refiere deseo de embarazo. ¿Cuál es la conducta a seguir? 	<ul style="list-style-type: none"> a) Dilatación anular mitral con desplazamiento del punto de coaptación.* b) Engrosamiento progresivo de las valvas con afinamiento cordal. c) Ruptura cordal con válvula flail que prolapsa hacia la aurícula izquierda en sístole. a) Indicar que no se embarace, presenta contraindicación absoluta por riesgo materno fetal inaceptable. b) Explicar que su patología no empeorará con el embarazo y que debe realizar controles cardiológicos trimestrales. c) Realizar un ecoestrés con ejercicio y determinar eventual necesidad de valvuloplastia pre-embarazo.* 	

Los autores de este trabajo revisaron las preguntas, mejoraron su redacción, consensuaron la respuesta correcta y prepararon las dos versiones de cada examen: una versión con las preguntas ordenadas por nivel de dificultad creciente y agrupadas según subtemas, y la otra con las preguntas al azar.

El examen de Prevención fue rendido por los alumnos de primer año: el grupo A rindió la versión ordenada y el grupo B la versión al azar.

El examen de Valvulopatías fue rendido por los alumnos de segundo año: el grupo C rindió la versión con preguntas ordenadas y el grupo D rindió la versión con preguntas aleatorias.

El índice de dificultad de cada pregunta se define como el porcentaje de respuestas correctas. De acuerdo con ese índice de dificultad (o de facilidad) las preguntas se pueden clasificar en 5 niveles de dificultad. (12)

- difícil: respondida correctamente por hasta el 15% de los alumnos.

- relativamente difícil: respondida correctamente por 16-31%.
- dificultad media: respondida correctamente por 32-68%.
- relativamente fácil: respondida correctamente por 69-84%.
- muy fácil: respondida correctamente entre el 85 y el 100% de los alumnos

Los resultados de cada versión del examen se expresaron en puntajes obtenidos (rango) y en cantidad de preguntas en cada nivel de dificultad.

Para el análisis de la media aritmética y DS se aplicó la prueba de Student, para la mediana y rango intercuaril la prueba Mann-Whitney-Wilcoxon (Tablas 1 y 2).

La diferencia entre porcentajes de preguntas en cada nivel de dificultad se estimó con el test de chi cuadrado o el test de Fisher si correspondiera. (Tablas 3 y 4).

Tabla 1. Puntajes obtenidos. Examen Prevención de la Enfermedad Cardiovascular: puntaje máximo posible: 50. Respondieron 155 residentes, grupos A y B

	Grupo A Versión ordenada n = 78	Grupo B Versión al azar n = 77	P
Rango de puntajes obtenidos	28-46	28-46	
Media	40,78	40,43	0,57
Desvío estándar	3,66	4,14	
Mediana	40	41	0,69
Percentilo 25-75	38-43	37-43	

Tabla 2. Puntajes obtenidos. Examen Valvulopatías: puntaje máximo posible: 50. Respondieron 129 residentes, grupos C y D

	Grupo C Versión ordenada n = 64	Grupo D Versión al azar n = 65	P
Rango de puntajes obtenidos	26-46	26-48	
Media	38,36	39,58	0,21
Desvío estándar	5,04	5,99	
Mediana	39	42	0,07
Percentilo 25-75	35-42	34,5-44	

Tabla 3. Nivel de dificultad del examen de Prevención de la Enfermedad Cardiovascular. Según cantidad de preguntas en cada una de las 5 categorías

Nivel de dificultad de las preguntas	Cantidad de preguntas Versión A Ordenada	Cantidad de preguntas Versión B Al azar	p
Difícil	1 (2%)	0 (0%)	0,802
Relativamente difícil	1 (2%)	1 (2%)	
Dificultad media	9 (18%)	10 (20%)	
Relativamente fácil	11 (22 %)	13 (26%)	
Muy fácil	28 (56%)	26 (52%)	
Total de preguntas del examen	50 (100%)	50 (100 %)	

Tabla 4. Nivel de dificultad del examen de Valvulopatías. Según cantidad de preguntas en cada una de las 5 categorías

Nivel de dificultad de las preguntas	Cantidad de preguntas Versión C Ordenada	Cantidad de preguntas Versión D Al azar	p
Difícil	0 (0%)	0 (0%)	0,345
Relativamente difícil	2 (4%)	0 (0%)	
Dificultad media	13 (26%)	12 (24%)	
Relativamente fácil	16 (32%)	15 (30%)	
Muy fácil	19 (38%)	23 (46%)	
Total de preguntas del examen	50 (100%)	50 (100%)	

RESULTADOS

Durante la primera semana de agosto de 2020, se administraron los exámenes utilizando la plataforma Moodle. El cuestionario estuvo disponible durante 75 minutos. Respondieron 284 residentes de 42 servicios de Cardiología del área metropolitana (ciudad de Buenos Aires y conurbano bonaerense).

El examen de Prevención resultó una prueba fácil ya que más del 50% de las preguntas, en ambas versiones, se ubicó en la categoría de muy fácil.

El examen de Valvulopatías también resultó una prueba fácil porque entre el 70% y 76% de las preguntas resultaron relativamente fáciles y muy fáciles.

DISCUSIÓN

Algunos autores sostienen que en el ensamblaje/ armado de un examen las preguntas o reactivos se deben ordenar según tema, subtema y saber específico y que se deben secuenciar de acuerdo con el grado de dificultad. Tradicionalmente se ha sostenido este criterio en el ámbito educativo (2) y también se han publicado algunos trabajos que muestran que el ordenamiento no afecta el rendimiento de los estudiantes. (8-11)

Los resultados obtenidos en este trabajo son consistentes con los de las investigaciones que no encontraron diferencias en los resultados de pruebas con distintos ordenamientos de las preguntas. Cabe señalar que ninguno de esos estudios fue realizado con estudiantes universitarios del área de ciencias de la salud.

Un factor a considerar es la especificidad de los contenidos o disciplina que se evalúan en el examen y también el perfil, nivel de conocimientos previos y/o experiencia profesional de los sujetos que responden. En el caso particular de los residentes, médicos con 3 años de experiencia profesional, es probable que ya hayan desarrollado la habilidad de enfrentar y resolver problemas clínicos de diferente orden, es decir sin una secuencia lógica, porque la práctica clínica implica recibir pacientes con distintas patologías y demandas en forma aleatoria. Un cuestionario que presenta distintos casos, estudios o contextos, sin orden alguno estaría reproduciendo, de alguna manera, la práctica cotidiana. Para llegar a comprobar esta posible explicación sería necesario comparar el rendimiento de los residentes con el rendimiento de estudiantes de medicina de los últimos años de la carrera y ver si a los estudiantes, novatos, los afecta un cuestionario desordenado.

Por otra parte, hay que considerar si la cantidad de preguntas fue suficiente como para lograr “una mezcla” de temas, teniendo en cuenta que estaban limitadas a entidades clínicas específicas y acotadas. Para revisar esta cuestión sería necesario diseñar un cuestionario más extenso, con distintos temas, distintas patologías y situaciones clínicas diversas. ¿Cuáles serían los resultados si el cuestionario explorara distintos temas y tuviera mayor cantidad de preguntas? En un examen

final de carrera o en un examen de admisión al sistema de residencias médicas ¿las preguntas ordenadas/ agrupadas por tema resultarían más fáciles de responder y las aleatorias serían más difíciles?

Limitaciones. Una importante debilidad de este trabajo es que las preguntas incluidas en los exámenes no fueron validadas en una prueba piloto. Y la secuencia de más fácil a más difícil fue establecida con base al criterio de los docentes, pero los resultados muestran que la gran mayoría de las preguntas eran fáciles y muy fáciles y por lo tanto esa secuenciación no funcionó. Es ampliamente conocido que las preguntas “buenas” deben tener un Índice de dificultad media para lograr un nivel de discriminación aceptable.

Se requieren más estudios y experiencias para encontrar formas de evaluación que puedan implementarse en tiempos de distanciamiento social y cuyos resultados sean válidos y confiables.

CONCLUSIONES

La pregunta con la que se inició este estudio ha quedado sin responder porque sus resultados no tienen la potencia suficiente para fundamentar una u otra forma de presentar las preguntas en un examen de ciencias clínicas.

Se plantea la conveniencia de hacer otro estudio con un cuestionario con mayor cantidad de preguntas y distintos temas que se administre a estudiantes a punto de graduarse y a residentes con algunos años de práctica profesional.

En esta experiencia no se logró ningún dato que permita hacer alguna consideración respecto a la utilización de exámenes con preguntas distribuidas al azar como una forma de evitar el fraude en las evaluaciones *online*.

Declaración de conflictos de intereses

Los autores declaran no tener ningún tipo de conflicto de intereses.

(Véase formulario de conflicto de intereses de los autores en la web / Material suplementario).

Agradecimientos

A la Dra Marina Khoury que revisó el manuscrito en general y los resultados en particular.

A los residentes de Cardiología alumnos del Curso BIANUAL de Cardiología cohorte 2019-2020.

BIBLIOGRAFÍA

1. Asociación de Facultades de Ciencias Médicas de la República Argentina (AFACIMERA). Evaluación Educativa. Módulo 2. Construcción y aplicación de exámenes. Buenos Aires. 2002
2. Mahias Finger P, Polloni Erazo M. Desarrollo de instrumentos de evaluación: pruebas. Instituto Nacional para la evaluación de la educación de México y Centro UC.MIDE. Chile. 2019. Disponible en: <https://www.inee.edu.mx/wp-content/uploads/2019/08/P2A354.pdf>
3. Palés-Argullós J. ¿Cómo elaborar correctamente preguntas de elección múltiple? Educ Méd 2010;13:149-55.

4. Epstein RM. Assessment in medical education. *N Engl J Med* 2007;56:387-96. [https://doi.org/ 10.1186/1472-6920-9-40](https://doi.org/10.1186/1472-6920-9-40)
5. Case S, Swanson D. National Board of medical examiners. *Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas*. NBME 2016. 4ª ed. Disponible en: https://www.nbme.org/sites/default/files/2020-01/DownloadingtheGoldBook_ES.pdf
6. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Med Educ* 2009;9:40. <https://doi.org/10.1186/1472-6920-9-40>.
7. Haladyna TM, Downing SM. How Many Options is enough for a Multiple-Choice Test Item? *Educational and Psychological Measurement* 1993. <https://doi.org/10.1177/0013164493053004013>
8. Marso R. N. Test item arrangement, testing time, and performance. *J Educ Measure* 1970; 7:113-8. <https://doi.org/10.1111/j.1745-3984.1970.tb00704.x>
9. Opara IM, Uwah I V. Effect of test item arrangement on performance in mathematics among junior secondary school students in obio/akpor local government area of rivers state Nigeria. *Br J Educ* 2017;5:1-9
10. BLEPT (Board Licensure Examination for Professional Teachers). Guidelines for teachers in arranging test items. 2015 Disponible en: <http://blept.blogspot.com>.
11. Süleyman NŞ. Does difficulty-based item order matter in multiple-choice exams? (Empirical evidence from university students). *Studies in Educational Evaluation* 2020;64: 100812. <https://doi.org/10.1016/j.stueduc.2019.100812>
12. Guilbert, Jean-Jacques, Rodríguez Torres, A, Carreres Quevedo, J, Feroso García, J, Franch Valverde, J. et al. *Guía pedagógica para el personal de salud / J.-J. Guilbert ; edición en español a cargo de A. Rodríguez Torres ... [et al.]*, 6a ed. Madrid : Instituto de Ciencias de la Educación, Universidad de Valladolid, 1994 <https://apps.who.int/iris/handle/10665/40678>