

Effect of Multiple-choice Questions Sequence on the Results of Cardiology Exams

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

AMANDA GALLI¹, MIRTA DANIELA LOBIANCO¹, SANDRA SWIESZKOWSKI¹, LUCÍA KAZELIAN¹, GRACIELA GIMENO,
SUSANA LAPRESA¹, MARISA PAGES¹, HUGO GRANCELLO¹

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

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 Bianual de Cardiología. Dos grupos de 1º año (tema Prevención) y dos grupos de 2º 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

INTRODUCTION

The Learning Assessment literature, (1-4) specifically on the topic of structured testing, gives instructions to ensure that the assessment instrument provides valid and reliable results.

The National Board of Medical Examiners (NBME) has published many recommendations on how to build exams in the area of Basic and Clinical Sciences. It insists that relevant content should be explored more than the recollection of specific data and that questions should be asked about conducts to follow in clinical situations. It suggests all answer options be homogeneous. (5)

Several authors have explored the number of options a multiple-choice question should have, and have shown that, in the area of health sciences, 3 options are enough. (6,7)

Various investigations have evaluated the impact of question order on student performance. Some publications claim that students perform better when questions are presented in order of increasing difficulty, and that a random distribution of questions harms students, while others state that the sequence of questions does not affect performance (8-11)

In the Biannual course of Cardiology developed in the Argentine Society of Cardiology (SAC), 8 exams of 50 questions each are designed and administered annually. This represents the generation and review of

more than 400 questions per year that are answered by about 300 residents who are training as specialists in Cardiology in medical residency programs in different hospitals.

Since the COVID -19 pandemic and due to mandatory social isolation, teaching has shifted to virtual activities and exams have been answered online. The current availability of certain resources (software) for the administration and correction of questionnaires allows the construction of several versions of the same exam simply by mixing the questions without respecting any criteria of logical or topic sequence, totally at random. The assembly of individualized exams, with random questions, appeared as an alternative to make potential fraud difficult when using portable electronic devices.

This work aimed to study whether random question order had any effect on the results of the exams taken by the residents who were participating in the Biannual Course of Cardiology.

METHODS

The first and second-year teachers of the SAC Biannual Course of Cardiology designed the questions for two exams: Prevention of Cardiovascular Disease (1st year) and Heart Valve Diseases (2nd year). Each exam consisted of 50 multiple choice questions with 3 options: one correct answer and two distractors. Here are 4 sample questions. The asterisk (*) marks the correct answer in each case.

Prevention Questions

Statement/question	Options
- What is the course of action to follow if a patient reports muscle pain, suspected to be an adverse effect of statin treatment?	a) Suspend treatment with statins and request CPK and TSH measurement * b) Continue treatment without making any changes c) Continue with statins until CPK and TSH results are received
- 41-year-old male patient, athlete, with LDL cholesterol of 195 mg/dl and low ten-year risk. In addition to diet, what treatment do you start because it has shown greater benefit?	a) ASA 100 mg/d b) Rosuvastatin 40 mg/day * c) Ezetimibe 10 mg/day

Valve Disease Questions

- What mechanisms are involved in the development of functional or secondary mitral regurgitation?	a) Mitral annular dilatation with displacement of the coaptation point. * b) Progressive thickening of the valves with chordal thinning. c) Chordal rupture with flail valve that prolapses towards the left atrium in systole.
- A 34-year-old patient consults for palpitations. Sinus rhythm in consultation. She brings an ECG Holter with 8 episodes of brief asymptomatic supraventricular tachyarrhythmia. Doppler echocardiogram: left ventricle with preserved diameters, left atrium with a markedly increased volume of 58 ml/m ² and pulmonary pressure estimated at 30 mmHg. Mild commissural calcification, without involvement of the mitral subvalvular apparatus and good valve mobility. Mitral valve area of 1.2 cm ² . The patient reports a desire for pregnancy. What is the conduct to follow?	a) Indicate not to get pregnant, it presents an absolute contraindication due to unacceptable maternal-fetal risk. b) Explain that her pathology will not worsen with pregnancy and that she should carry out quarterly cardiology controls c) Carry out a stress echocardiography with exercise and determine the possible need for a pre-pregnancy valvuloplasty. *

The authors of this work reviewed the questions, improved their wording, agreed on the correct answer and prepared the two versions of each exam: one version with the questions ordered by level of increasing difficulty and grouped according to subtopics, and the other with the questions placed randomly.

The Prevention exam was taken by first-year students: group A took the ordered version and group B the random version.

The Heart Valve Disease exam was taken by the second-year students: group C took the version with ordered questions and group D the version with random questions.

The results of each exam version were expressed as scores obtained (range and central tendency) and number of questions for each difficulty level.

The difficulty index of each question is defined as the percentage of correct answers. According to this difficulty (or ease) index the questions can be classified into 5 levels

of difficulty. (12)

- difficult: answered correctly by up to 15% of students.
- relatively difficult: answered correctly by 16-31%.
- medium difficulty: answered correctly by 32-68%.
- relatively easy: answered correctly by 69-84%.
- very easy: answered correctly by 85-100% of students.

The results of each version of the exam were expressed in scores obtained (range) and in the number of questions at each level of difficulty. For the analysis of the arithmetic mean and SD, the Student test was applied, and the Mann-Whitney-Wilcoxon test for the median and interquartile range (Tables 1 and 2).

The difference between the percentages of questions at each level of difficulty was estimated using the chi-square test or Fisher's test, if applicable. (Tables 3 and 4).

Table 1. Scores obtained in the Cardiovascular Disease Prevention Examination: maximum possible score: 50. A total of 155 residents belonging to groups A and B responded to the questionnaire

	Group A Ordered version n=78	Group B Random version n=77	p
Range of scores obtained	28-46	28-46	
Mean	40.78	40.43	0.57
Standard Deviation	3.66	4.14	
Median	40	41	0.69
Interquartile range	38-43	37-43	

Table 2. Scores obtained in the Heart Valve Diseases examination: maximum possible score: 50. A total of 129 residents belonging to groups C and D responded to the questionnaire.

	Group C Ordered version n= 64	Group D Random version n= 65	p
Range of scores obtained	26-46	26-48	
Mean	38.36	39.58	0.21
Standard Deviation	5.04	5.99	
Median	39	42	0.07
Interquartile range	35-42	34,5-44	

Table 3. Difficulty level of the Cardiovascular Disease Prevention exam according to the number of questions in each of the 5 categories

Difficulty level of the questions	Number of questions		p
	Version A	Version B	
Difficult	1 (2%)	0 (0%)	0,802
Relatively difficult	1 (2%)	1 (2%)	
Medium difficulty	9 (18%)	10 (20%)	
Relatively easy	11 (22 %)	13 (26%)	
Very easy	28 (56%)	26 (52%)	
Total exam questions	50 (100%)	50 (100%)	

The Prevention exam was an easy test since more than 50% of the questions, in both versions, were in the very easy category.

Table 4. Difficulty level of the Heart Valve Diseases exam according to the number of questions in each of the 5 categories

Difficulty level of the questions	Number of questions		p
	Version C	Version D	
Difficult	0 (0%)	0 (0%)	0.345
Relatively difficult	2 (4%)	0 (0%)	
Medium difficulty	13 (26%)	12 (24%)	
Relatively easy	16 (32%)	15 (30%)	
Very easy	19 (38%)	23 (46%)	
Total exam questions	50 (100%)	50 (100 %)	

The Heart Valve Diseases exam was also an easy test because between 70% and 76% of the questions were relatively easy to very easy.

RESULTS

During the first week of August 2020, the exams were administered using the Moodle platform. The questionnaire was available for 75 minutes and it was answered by 284 residents of 42 Cardiology services in the metropolitan area (City of Buenos Aires and Greater Buenos Aires)..

DISCUSSION

Some authors sustain that in the assembly of an exam the questions or reagents must be ordered according to topic, sub-topic and specific knowledge, and that they must be sequenced according to the degree of difficulty. This criterion has traditionally been upheld in the educational field (2) and some works that show that the ordering does not affect student performance have also been published. (8-11)

The results obtained in this work are consistent with those of the investigations that did not find differences in the results of tests with different question order. It should be noted that none of these studies was conducted with university students in the area of health sciences.

One factor to consider is the specificity of the content or discipline that is evaluated in the exam and also the profile, level of previous knowledge and/or professional experience of the subjects who respond. In the particular case of residents, physicians with 3 years of professional experience, it is likely that they have already developed the ability to face and solve clinical problems of different order, that is, without a logical sequence, because clinical practice involves randomly receiving patients with different pathologies and demands. A questionnaire that presents different cases, studies or contexts, without any order, would somehow be reproducing daily practice. In order to verify this possible explanation, it would be necessary to compare the performance of residents with that of medical students in the last years of their degree and see if the students, novices, are affected by a disordered questionnaire.

On the other hand, it is necessary to consider whether the number of questions was sufficient to achieve a "mixture" of topics, taking into account that they were restricted to specific and limited clinical entities. To review this issue, it would be necessary to design a more extensive questionnaire, with different topics, different pathologies and diverse clinical situations. What would be the results if the questionnaire explored different topics and had more questions? In a final exam or in an admission exam to the medical residency system, would the questions ordered/grouped by topic be easier to answer and the random ones more difficult?

riminator of body fat or, in fact, when patients without obesity or overweight are compared with other patients, those with underweight could be at greater risk, tipping the scale. Like our findings, other authors have not found any differences in mortality

according to BMI either during hospitalization or at long term follow-up. (17, 18) In our study, only Killip class other than A, age and multi-vessel disease were independent predictors of mortality.

Limitations

An important weakness of this work is that the questions included in the exams were not validated in a pilot test. And the sequence from easiest to most difficult was established based on the professors' criteria, but the results show that the vast majority of the questions were easy and very easy, and therefore that sequencing did not work. It is widely known that "good" questions must have an index of medium difficulty to achieve an acceptable level of discrimination.

More studies and experiences are required to find forms of evaluation that can be implemented in times of social distancing with valid and reliable results.

CONCLUSIONS

The query with which this study began has remained unanswered because its results do not have sufficient power to support one or another way of presenting the questions in a clinical science exam.

The convenience of carrying out another study with a questionnaire having a greater number of questions and different topics that is administered to students about to graduate and to residents with some years of professional practice is proposed.

In this experience, no data was obtained that would allow any consideration to be made regarding the use of exams with randomly distributed questions as a way to avoid fraud in online evaluations.

Conflicts of interest

None declared.

(See authors' conflict of interests forms on the web/Additional material.)

Acknowledgements

To Dr. Marina Khoury who reviewed the manuscript in general and the results in particular.

To the residents of Cardiology, students of the Biannual Course of Cardiology, cohort 2019-2020

REFERENCES

1. Asociación de Facultades de Ciencias Médicas de la República Argentina (AFACIMERA). Evaluación Educacional. 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>
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 func-

tioning 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>