

## Exámenes de selección para postulantes a residencias en cardiología. Variables que se asocian con los resultados

ALBERTO E. ALVES DE LIMA<sup>1†</sup>, FERNANDO BOTTO<sup>1</sup>, FABIAN AZZARI<sup>1</sup>, AMANDA GALLI<sup>2</sup>, LUIS GUZMAN<sup>1\*</sup>, MARCELO TRIVI<sup>1\*</sup>, JORGE ALBERTAL<sup>1\*</sup>, JORGE BELARDI<sup>1\*</sup>

### RESUMEN

#### *Objetivo*

Se realizó este estudio con el propósito de identificar variables predictivas de buen rendimiento en una prueba de tipo opción múltiple diseñada para seleccionar médicos aspirantes a ingresar en un programa de entrenamiento en cardiología.

#### *Población y método*

Se construyó una prueba de 50 preguntas presentadas en dos temas: las preguntas 1 a 25 del tema 1 eran iguales a las preguntas 26 a 50 del tema 2 y viceversa. El total de los aspirantes se dividió en dos grupos según el rendimiento definido por la mediana: bueno (B) (resultado > mediana) y no bueno (NB) (resultado ≤ mediana). Se compararon ambos grupos según las siguientes variables: sexo, edad, universidad que otorgó el título, promedio general de la carrera (PC), experiencia como auxiliar docente en Ciencias Básicas (AD) y experiencia en Medicina Interna (residencia o concurrencia ≥ 1 año).

#### *Resultados*

Noventa y cinco aspirantes realizaron la prueba. La mediana del puntaje obtenido fue 25. De todas las variables analizadas, sólo el PC ( $p = 0,00002$ ) y el antecedente de AD ( $p = 0,04$ ) se asociaron con B en la prueba. Con respecto a los resultados de ambos temas, hubo una tendencia a mejor desempeño en el tema 2 (mediana 28) *versus* el tema 1 (mediana 23) ( $p = 0,07$ ), sin que hubieran diferencias significativas entre los postulantes de ambos temas.

#### *Conclusión*

El antecedente de AD y de PC fueron los únicos predictores de buen rendimiento en esta prueba de selección. La diferencia del rendimiento entre ambos temas pudo deberse a la secuencia en que se presentaron las preguntas. *REV ARGENT CARDIOL* 2001; 69: 327-331.

*Palabras clave* Educación médica - Examen de selección - Residencias médica - Confección de exámenes

### INTRODUCCION

La evaluación en el marco del proceso educacional implica la obtención de información válida y confiable para la comprobación del logro de los objetivos especificados con antelación. De este modo contribuye significativamente a corregir,

mejorar y reorientar los procesos de enseñanza y de aprendizaje.

Constituye además un proceso que comienza mucho antes del día en el que se realiza la prueba, con la definición de los objetivos de aprendizaje y de los propósitos de la prueba y finaliza mucho des-

<sup>1</sup> Instituto Cardiovascular de Buenos Aires. Departamento Educación

<sup>2</sup> Facultad de Medicina. Universidad de Buenos Aires

\* Miembro Titular SAC

† Para optar a Miembro Titular SAC

Trabajo recibido para su publicación: 9/5/00. Aceptado: 1/5/01

Dirección para separatas: Dr. Alberto E. Alves de Lima. Blanco Encalada 1543, (1428) Buenos Aires, Argentina. E-mail: alves@intramed.net ar

pués a través del análisis y la devolución de la información obtenida. (1)

La evaluación del área cognoscitiva a través de pruebas estructurales de tipo opción múltiple es muy utilizada en nuestro medio, incluso como único método para certificar competencia clínica.

Su redacción y confección requieren un proceso de elaboración complejo y meticuloso.

El objetivo del estudio es el de identificar variables predictivas de buen rendimiento en una prueba estructurada de tipo opción múltiple diseñada con el fin de seleccionar médicos para ingresar en un programa de entrenamiento en cardiología.

## POBLACION Y METODO

Los requisitos para participar de la prueba fueron edad menor de 35 años y haber egresado de la universidad en un período no mayor de 5 años.

Se construyó una prueba escrita estructurada de 50 preguntas de tipo opción múltiple con cinco opciones cada una, una respuesta correcta y cuatro distractores. La prueba exploraba conocimientos en Medicina Interna. Se prepararon dos versiones de la prueba: tema 1 y tema 2. Ambas versiones tenían las mismas preguntas, pero presentadas en diferente orden; las preguntas 1 a 25 del tema 1 eran iguales a las preguntas 26 a 50 del tema 2 y viceversa. Los alumnos tuvieron 75 minutos para responder la prueba.

Las pruebas fueron corregidas por dos docentes en forma independiente. Hubo una correlación perfecta entre los puntajes otorgados a cada examen entre los dos docentes. Se analizaron en forma retrospectiva las características basales de la población. El total de los examinados se dividió en dos grupos según el rendimiento definido por la mediana: bueno (B), resultado mayor que la mediana, y no bueno (NB), resultado igual o menor que la mediana. Se compararon ambos grupos en relación con las siguientes variables: sexo, edad, universidad que otorgó el título, promedio general de la carrera, experiencia como auxiliar docente en Ciencias Básicas y experiencia en Medicina Interna: residencia o concurrencia mayor o igual a un año. Se determinó el índice de dificultad de cada pregunta. El índice de dificultad expresa la proporción de sujetos que respondió correctamente el ítem. A mayor cantidad de sujetos que responden correctamente un ítem, menor dificultad.

## Método estadístico

Los datos cuantitativos se expresan a través de su media  $\pm$  1 desvío estándar o su mediana según corresponda por su distribución. Las variables de tipo cualitativas se expresan en valores porcentuales. Las variables cuantitativas se analizaron con la

**Tabla 1**  
Características basales de la población

Variables	
Edad (años $\pm$ DS)	25,8 ( $\pm$ 1,5)
Sexo masculino (%)	62
UBA (%)	53
PC $\pm$ DS	7,75 ( $\pm$ 0,95)
AD (%)	42,3
EMI (%)	13,4

Referencias: UBA: Porcentaje de postulantes egresados de la Universidad de Buenos Aires. PC: Promedio general de la carrera. AD: Antecedentes de actividad docente como auxiliar en Ciencias Básicas. EMI: Experiencia > 1 año en Medicina Interna.

prueba de Student o de Mann Whitney o de Wilcoxon de acuerdo con su distribución y las de tipo cualitativas con la prueba de chi cuadrado o la prueba de Fisher según correspondiera. Se consideró estadísticamente significativo un valor de  $p < 0,05$ .

## RESULTADOS

Los 95 participantes que realizaron la prueba conformaron un grupo joven con una edad promedio de 25,8 ( $\pm$  1,5) años, predominantemente de sexo masculino (62%). El 53% era egresado de la Universidad de Buenos Aires. El promedio general de la carrera fue de 7,75 ( $\pm$  0,95). El 42% tenía antecedentes de actividad docente como auxiliar en Ciencias Básicas y sólo el 13,4% contaba con una experiencia mayor o igual a un año en una residencia o concurrencia en clínica médica (Tabla 1). La media del puntaje obtenido por la población total fue de 24,9/50. La mediana fue de 25. El puntaje máximo obtenido fue de 35/50 y el mínimo de 11/50. De todas las variables analizadas entre el grupo B y el grupo NB, el promedio general de la carrera ( $p = 0,00002$ ) y el antecedente de docencia como auxiliar en Ciencias Básicas (0,04) se asociaron con un buen rendimiento en la prueba (Tabla 2).

Se observó una tendencia a mejor rendimiento en el tema 2 (mediana del puntaje 28) con respecto al tema 1 (mediana del puntaje 23) ( $p = 0,07$ ). Se compararon las características basales de ambos grupos

**Tabla 2**  
Variables evaluadas por grupo según mediana

Variable	Grupo B n = 47	Grupo NB n = 48	p
EMI (%)	14	12,8	0,9
UBA (%)	24,2	25,8	0,9
Edad (años)	25,9	25,6	0,4
sexo masculino (%)	56	68	0,3
AD (%)	53,2	32	0,03
PC (media)	8,20	7,3	0,00003

**Tabla 3**  
Características basales por tema

Variable	Tema 1 n = 47	Tema 2 n = 48	p
EMI (%)	8,5	19,4	0,2
UBA (%)	57,5	51,0	0,5
Edad (años)	25,9	25,7	0,5
PC (media)	7,6	7,8	0,5
Sexo masculino (%)	63,8	59,1	0,8
AD (%)	42,5	40,8	0,9

en relación con las siguientes variables: sexo, edad, universidad que otorgó el título, promedio general de la carrera, experiencia como auxiliar docente en Ciencias Básicas y experiencia en Medicina Interna: residencia o concurrencia mayor de un año.

No se hallaron diferencias significativas en las variables analizadas (Tabla 3).

En el Gráfico 1 se muestra el índice de dificultad por pregunta entre el tema 1 y el tema 2. Allí puede observarse que en la primera mitad del examen (preguntas 1 a 25) ambos temas tienen índices de dificultad por pregunta similares. A partir de la pregunta 26 se observa una tendencia a menor dificultad por pregunta del tema 2 con respecto al tema 1.

De acuerdo con el modo en que se confeccionó la prueba, el bloque de preguntas 26 a 50 del tema 1 fue respondido en la primera mitad como preguntas 1-25 en el tema 2; así existió una diferencia en el momento en que las preguntas fueron contestadas.

## DISCUSION

La evaluación es un proceso complejo que implica búsqueda de información a través de una prueba y el desarrollo de un juicio de valores del significado de la información recogida. Esto lleva fundamentalmente a la toma de decisiones en función de los datos obtenidos y culmina con la comunicación o devolución de los resultados a los individuos involucrados en el proceso educativo. (2)

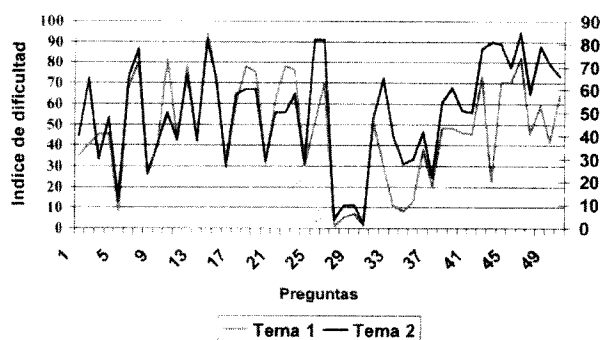


Gráfico 1. Índice de dificultad por pregunta.

El objetivo fundamental es el de mejorar la práctica docente y ayudar a las personas que la realizan. (3)

El análisis de los resultados permite evaluar la calidad del instrumento utilizado para la evaluación (evaluación de la evaluación).

En este trabajo se describe el análisis de los resultados llevado a cabo en un examen de selección teórico de tipo opción múltiple. Se incluyeron médicos jóvenes, la mayoría sin experiencia en Medicina Interna. Debido a que el currículo de entrenamiento en la institución incluye un año de Medicina Interna, la falta de experiencia previa en clínica médica no fue un criterio de exclusión para postularse como residente.

En el análisis de este grupo de aspirantes se observó que el promedio general de la carrera (en este caso superior a 8) y el antecedente de ayudantía previa fueron predictores de buen rendimiento en esta prueba. Con respecto al promedio general de la carrera, existen datos publicados coincidentes con esta observación. Calhoun y colaboradores demostraron que existe correlación positiva entre el promedio general de la carrera y el éxito en la obtención de una posición en un programa de residencia. (4) El promedio general es un indicador de rendimiento académico del estudiante, que se obtiene del resultado de las diferentes evaluaciones de conocimientos desarrolladas durante su carrera. Así, parece lógico pensar que aquellos que tuvieron buen rendimiento académico tengan buen rendimiento en estos exámenes de selección, que en última instancia evalúan conocimientos similares a los dictados durante el período universitario. Con respecto al antecedente de ayudantía previa, no hay datos concluyentes en la literatura, fundamentalmente porque existe gran variabilidad en la definición de ayudantía previa en las diferentes universidades del mundo. Roop y colaboradores observaron una correlación positiva entre el antecedente de actividad docente en los primeros años de la carrera y un buen rendimiento durante el internado rotatorio. (5) Es importante tener en cuenta también que es posible que las variables —los antecedentes de ayudantía y el promedio general— estén relacionadas entre sí. En algunas cátedras existe la necesidad de que se tengan promedios altos como requisito para ingresar en la escuela de ayudantes. De todas maneras, es probable que el tener "habilidades docentes" pueda influir favorablemente en la adquisición de conocimientos y así el rendimiento en la evaluación.

Del trabajo surge además que existió una tendencia a mejor rendimiento en el tema 2 con respecto al tema 1 (mediana de puntaje en el tema 1: 23 versus mediana de puntaje en el tema 2: 28,  $p = 0,07$ ). Esta diferencia pudo deberse al azar. De todas maneras, de acuerdo con los resultados obtenidos (Gráfico 1)

la diferencia en el rendimiento pudo deberse a la secuencia de distribución de las preguntas. Allí es posible observar que el índice de dificultad durante las primeras 25 preguntas fue similar en ambos temas. Durante las segundas 25 preguntas se observó una tendencia a tener mejores índices de dificultad en el tema 2. Cabe destacar que en la elaboración de la prueba se definieron claramente propósitos y objetivos, y se tomó en consideración el equilibrio del temario utilizando una tabla de especificaciones. (3-6) Además, el 30% de las preguntas utilizadas se habían aplicado en otras oportunidades y estaban evaluadas en términos de índice de dificultad y poder discriminativo adecuados. El poder discriminativo de una pregunta permite averiguar si los que respondieron correctamente la pregunta son los alumnos que tienen más conocimientos.

Así es posible especular que la construcción de los temas 1 y 2 del examen a través de la inversión de dos grandes bloques de preguntas podría ser contraproducente. La inversión de grandes bloques podría provocar o exagerar desequilibrios en el temario. Así, variables como el momento en que se responde el bloque, las características de las preguntas del bloque y el cansancio acumulado a lo largo de la prueba podrían influir en los resultados. Una alternativa posible sería ordenar el examen en unidades temáticas y crear los temas 1 y 2 modificando el orden de las preguntas sólo dentro de cada unidad, manteniendo la secuencia original del examen.

El caso analizado en este trabajo pone al descubierto que la secuencia en la que se ordenan las preguntas en una prueba podría ser un factor que influye en los resultados; la secuencia de las preguntas representa o expresa una secuencia lógica en el abordaje del tema; el orden en el que se presentan los reactivos facilita u obstaculiza la respuesta del sujeto. La respuesta del sujeto no se debe sólo a su propio conocimiento de la materia sino que parece, además, condicionada por la forma de preguntar.

En conclusión, el estudio muestra que variables simples pueden predecir buenos resultados en las pruebas de selección para postulantes a programas de entrenamiento en el posgrado. Además surge que la confección de la prueba requiere un proceso de elaboración complejo y meticuloso, ya que detalles mínimos pueden influir en los resultados.

Estos datos nos recuerdan que las pruebas deben satisfacer tres requerimientos básicos: 1. Validez: ¿Mide realmente lo que queremos evaluar? 2. Precisión: ¿Es confiable lo que está midiendo? y 3. Practicable: ¿Es factible en función de tiempo y recursos disponibles? De este modo, un examen es válido si es capaz de medir los objetivos propuestos al inicio

del proyecto, es preciso si es posible reproducir los resultados en diferentes oportunidades y es factible si su costo es adecuado o si el tiempo para obtener los resultados es aceptable. (6-8)

Las estrategias utilizadas para seleccionar alumnos no siempre cumplen estos requisitos básicos. Los exámenes de tipo opción múltiple como los efectuados en este caso constituyen la técnica de elección para certificar competencia clínica, son practicables pero no siempre confiables y su sola aprobación no asegura aptitud para la práctica médica. (7, 8)

#### Limitaciones del estudio

Cabe destacar que se trata de un estudio retrospectivo con un número reducido de casos basado sobre los datos obtenidos en la ficha de inscripción. Es posible que existan otras variables que puedan asociarse con el rendimiento. Esto limita la posibilidad de sacar conclusiones definitivas.

#### SUMMARY

#### SELECTION TEST FOR APPLICANTS TO A RESIDENCY PROGRAM IN CARDIOLOGY: VARIABLES ASSOCIATED TO THE RESULTS

##### *Objective*

To identify predictors of good performance in a multiple choice test designed to select physicians applying to a cardiology-training programme.

##### *Population and methods*

A 50-question test was designed in two different versions: questions 1-25 in version 1 were the same as questions 26-50 in version 2, and vice versa. The whole population of applicants taking the test was divided in two groups, according to the results. Graduate students in group qualified as good were those who obtained a score beyond the median; those with a score below the median were qualified as no good (score  $\leq$  median). Both groups were compared according to different variables: gender, age, university of procedence, final average score at university (FAS), experience in teaching assistance in basic sciences (TA), and experience in internal medicine (residence or concurrence  $>$  1 year).

##### *Results*

Ninety-five applicants took the test. The median score was 25. Of all the analyzed variables, only the FAS ( $p = 0.00002$ ) and the TA were associated with a good performance in the test. There was a tendency to a better performance in those who answered to version 2 (median score 28) versus the responders to version 1 (median score 23) ( $p = 0.07$ ). Nevertheless, there were no significant differences

among the scores of applicants submitted to any of both versions of the same exam.

#### *Conclusion*

The FAS and the experience in TA were the only predictors of a good performance in this selection test. We speculate that the slight (non-significant) difference in performance between those answering version 1 or 2 could be the result of the different order of presentation of the questions.

*Key words* Medical education - Admission exam -  
Medical residence - Exam characteristics

#### BIBLIOGRAFIA

1. Galli A, Castro C. Exámenes orales y escritos. Asociación Amigos de la Facultad de Medicina (ed). Programa de formación de formadores en ciencias de la salud. Buenos Aires, Argentina, La Prensa Médica Argentina 1998; 3: 29-69.
2. Ruelas E, Martini C. La calidad de la educación médica. AFACIMERA; Asociación de Facultades de Ciencias Médicas de la República Argentina (ed). Educación médica. Desafíos y tendencias. Buenos Aires, Argentina, Gráfica Porteña 1999; 3: 63-79.
3. Linn R. Instructional goals and objectives: Foundation for assessment. *En*: Linn R, Gronlund NE (eds). Measurement and assessment in teaching (7<sup>th</sup> ed). NJ, USA, Prentice-Hall Inc 1995; pp 25-46.
4. Calhoun KH, Hokanson JA, Bailey BJ. Predictors of residency performance. A follow-up study. *Otolaryngol Head Neck Surg* 1997; 116: 647-651.
5. Roop SA, Pangaro L. Effect of clinical teaching on student performance during a medicine clerkship. *Am J Med* 2001; 110: 205-209.
6. Ebels RL. The role of measurement and assessment in teaching. *En*: Ebel R, Frisbie DA (eds). Essentials of educational measurement (5<sup>th</sup> ed). NJ, USA, Englewood Cliffs: Prentice Hall 1991; pp 3-23.
7. Newble D. Assessing students. *En*: Newble D, Cannon R (eds). A handbook for medical teachers. Falcon House, Lancaster, UK, MTP Press Limited 1987; pp 88-115.
8. Miller GE. The assessment of clinical skills competence performance. *Acad Med Science* 1990; 65 (Suppl): S63-67.