

## ARTÍCULO ORIGINAL

### Indicadores de calidad para un examen teórico de la especialidad de medicina general integral

#### Quality indicators for a theoretical exam of the Comprehensive General Medicine speciality

MsC. Ernesto Sánchez Hernández,<sup>1</sup> MsC. Marianela Medina Pavón,<sup>II</sup> MsC. Mislav Rodríguez García,<sup>1</sup> MsC. Liudmila Vega Van Der Meer<sup>1</sup> y Gertrudis de la Torre Vega<sup>III</sup>

<sup>1</sup> Hospital Clínicoquirúrgico Docente "Dr. Joaquín Castillo Duany", Santiago de Cuba, Cuba.

<sup>II</sup> Universidad de Ciencias Médicas, Santiago de Cuba, Cuba.

<sup>III</sup> Centro Provincial de Información de Ciencias Médicas, Santiago de Cuba, Cuba.

#### RESUMEN

Se realizó un estudio descriptivo y transversal combinado con enfoques cualitativos, a fin de evaluar la calidad de los exámenes de control parcial aplicados a 30 médicos de primero y segundo años residentes en medicina general integral, que laboraban en el Hospital Clínicoquirúrgico Docente "Dr. Joaquín Castillo Duany" de Santiago de Cuba, durante el año 2014. Se emplearon métodos teóricos, empíricos y de procesamiento estadístico en el análisis de los instrumentos de evaluación. Ambos exámenes teóricos presentaron diferencias y similitudes constatadas a través de indicadores, tales como número de horas asignadas al contenido y cantidad de incisos empleados en la exploración, índice de dificultad por temáticas y preguntas, así como nivel de discriminación y consistencia interna del instrumento, todo lo cual ofreció validez y confiabilidad para el sistema de evaluación del aprendizaje.

**Palabras clave:** evaluación del aprendizaje, calidad de instrumentos de evaluación, índice de dificultad, poder de discriminación, atención secundaria de salud.

#### ABSTRACT

A descriptive and cross-sectional study with qualitative approaches was carried out, in order to evaluate the quality of the exams of partial control applied 30 doctors of first and second years of residency in Comprehensive General Medicine, who worked in "Dr. Joaquín Castillo Duany" Teaching Clinical Surgical Hospital from Santiago de Cuba, during the year 2014. Theoretical, empiric methods were used as well as those of statistical processing in the analysis of the evaluation tools. Both theoretical exams presented differences and similarities verified through indicators, such as number of hours assigned to the content and quantity of subsections used in the exploration, index of difficulty by theme and question, as well as discrimination level and internal consistency of the tool, all of which offered validity and reliability for the evaluation system of learning.

**Key words:** evaluation of learning, quality of evaluation tools, index of difficulty, discrimination power, secondary health care.

## INTRODUCCIÓN

En el proceso de especialización en medicina general integral (MGI), el aprendizaje se desarrolla unido a la actividad laboral, en condiciones reales, y con una alta independencia del residente, lo cual requiere de la intervención de profesores y tutores, a fin de valorar el cumplimiento de los objetivos propuestos en el programa formativo.

Ahora bien, en dependencia del propósito y momento en que se producen, el sistema establece 3 tipos de evaluaciones: de curso, promoción y graduación. Particularmente, la primera, permite hacer una valoración integral del desarrollo alcanzado por el residente; y con el trabajo de control parcial, como una de sus modalidades, se pretende la verificación del grado de cumplimiento y calidad de los objetivos propuestos en cada módulo, mediante la aplicación de ejercicios teórico-prácticos o teóricos, de lo cual debe existir constancia en el grupo básico de trabajo hasta el egreso del residente.<sup>1</sup>

Cuando se crean instrumentos de evaluación del aprendizaje, tales como los trabajos de control parcial, es necesario conocer los indicadores que definen su calidad, con el fin de establecer criterios uniformes que disminuyan la influencia del factor subjetivo en las calificaciones, así como asegurar validez y confiabilidad en la medición. Mientras que, en países como Estados Unidos es obligatorio que estos criterios de calidad se satisfagan, para otros, es inexistente esta norma.<sup>2-4</sup>

Sin lugar a dudas, Cuba enfrenta estos retos y tendencias en la educación médica superior con un sistema exigente en la formación académica, técnica, humanística y social de sus recursos humanos.<sup>5</sup>

Sobre la base de estas observaciones, donde la validez y confiabilidad de los instrumentos de evaluación diseñados pueden presentar limitaciones para su medición sobre bases científicas, se decidió realizar este estudio con la finalidad de evaluar la calidad de los exámenes de control parcial aplicados a residentes de medicina general integral, pertenecientes a las instituciones armadas en Santiago de Cuba.

## MÉTODOS

Se realizó un estudio descriptivo y transversal, combinado con enfoques cualitativos, a fin de evaluar la calidad de 30 exámenes de control parcial aplicados a los residentes de primer y segundo años de medicina general integral, que laboraban en el Hospital Clínicoquirúrgico Docente "Dr. Joaquín Castillo Duany" de Santiago de Cuba, desde abril hasta mayo del 2014.

Fue necesaria la consulta a expertos para alcanzar, por consenso, una clasificación del índice de dificultad que debía caracterizar las preguntas del examen.

El instrumento evaluativo se definió como un examen de respuesta estructurada con preguntas de tipo verdadero o falso, respuestas cortas, de emparejamiento y opción múltiple. La información primaria se obtuvo de la planificación docente para cada año de residencia, otros documentos oficiales de la especialidad y de los 30 exámenes de control parcial.

Se valoró la correspondencia entre el fondo de tiempo asignado en la planificación docente a los contenidos de los módulos y la cantidad de incisos que lo exploraban. Se determinó el coeficiente de correlación de Spearman<sup>6</sup> que expresa el grado de asociación entre 2 variables, según el sentido de la relación de estas en términos de aumento o disminución. Para comparar los resultados y clasificar los ítems según el índice de dificultad, se aceptaron los intervalos contenidos en la propuesta de Bachkoff.<sup>3</sup>

Fueron clasificados los incisos reales y calculados sus porcentajes, lo cual permitió la comparación con los ítems esperados, cuyos valores se obtuvieron a partir de la clasificación contenida en la bibliografía consultada.<sup>2</sup>

#### Indicadores de validez del instrumento

1. Índice de dificultad: la dificultad de un ítem se entendió como la proporción de personas que respondieron correctamente un reactivo (inciso, temática o pregunta) del examen; o sea, a mayor dificultad del ítem menor fue su índice.
2. Índice y coeficiente de discriminación: expresan la medida en que un ítem o pregunta discrimina (distingue) entre aquellos que obtuvieron buenas calificaciones en el trabajo de control parcial y los que alcanzaron bajas calificaciones.<sup>7</sup>

#### Índice de discriminación (D)

Mientras más alto es el valor, el reactivo diferenciará mejor a los alumnos con altas y bajas calificaciones. Si todos los residentes que pertenecen al grupo que posee notas más altas contestan correctamente un inciso, temática o pregunta y aquellos del grupo que tiene notas más bajas contestan incorrectamente, entonces  $D=1$  (valor máximo de este indicador); si sucede lo contrario,  $D=-1$  (valor máximo negativo); si ambos grupos contestan por igual,  $D=0$  (valor mínimo de discriminación).

Coeficiente de discriminación ( $r_{pbis}$ ): la correlación del punto biserial ( $r_{pbis}$ ) se utiliza para saber si las personas "adecuadas" son las que obtienen las respuestas correctas, el poder predictivo del inciso, temática o pregunta, y cómo puede contribuir a las predicciones.

3. Confiabilidad del examen por preguntas: el coeficiente alfa de Cronbach toma un valor positivo. Indica que la pregunta afecta la consistencia interna y disocia los resultados del examen. Los resultados deben tabularse según el alfa del examen y de cada pregunta. El alfa se calcula sin tener en cuenta los datos de las preguntas.<sup>2</sup>

## RESULTADOS

La tabla 1 muestra que en la planificación docente se asignó a todos igual número de horas para el aprendizaje; sin embargo, en el módulo 6 se concentró 34 % de todos los incisos elaborados en el instrumento evaluativo, seguido en orden de frecuencia por los módulos 4 (19,5 %) y 22 (17,0 %), respectivamente. El coeficiente de correlación de Spearman permitió asegurar que los valores de ambas variables tuvieron una correlación entre fuerte y perfecta.

**Tabla 1.** Fondo de tiempo asignado al contenido de primer año y número de ítems para su exploración

No.	Contenido primer año	Fondo de tiempo asignado	%	Número de ítems que exploran el contenido	%
1	Módulo 3	8 horas	16,6	4	10,0
2	Módulo 4	8 horas	16,6	8	19,5
3	Módulo 6	8 horas	16,6	14	34,0
4	Módulo 7	8 horas	16,6	4	10,0
5	Módulo 9	8 horas	16,6	4	10,0
6	Módulo 22	8 horas	16,6	7	17,0
7	Total	48 horas	100,0	41	100,0

Coefficiente de correlación de Spearman= 0,99

Para los residentes de segundo año, el mayor fondo de tiempo se concentró con similar frecuencia en los módulos 10, 12, 19 y 21 (14,0 %); no así la cantidad de incisivos por preguntas, donde 11,4 % (como valor máximo) estuvo presente en el contenido explorado de los módulos 11, 12, 16 y 21. Al aplicar el coeficiente de correlación de Spearman el valor obtenido fue de 0, por lo cual se desestimó la posibilidad de correlación (tabla 2).

**Tabla 2.** Fondo de tiempo asignado al contenido de segundo año y número de ítems para su exploración

No.	Contenido segundo año	Fondo de tiempo asignado	%	Número de ítems que exploran el contenido	%
1	Módulo 10	8 horas	14,0	4	9,0
2	Módulo 11	4 horas	7,0	5	11,4
3	Módulo 12	8 horas	14,0	5	11,4
4	Módulo 15	4 horas	7,0	4	9,0
5	Módulo 14	4 horas	7,0	4	9,0
6	Módulo 16	4 horas	7,0	5	11,4
7	Módulo 17	4 horas	7,0	4	9,0
8	Módulo 18	4 horas	7,0	4	9,0
9	Módulo 19	8 horas	14,0	4	9,0
10	Módulo 21	8 horas	14,0	5	11,4
	Total	56 horas	100,0	44	100,0

Coefficiente de correlación de Spearman= 0,00

Al analizar la tabla 3 se observó, que para el primer año aumentó el número de incisivos fáciles (58,5%), sobre lo esperado (5,0 %); y hubo ausencia de los difíciles en comparación con la estimación deseada (5,0 %). Situación semejante ocurrió en el examen de segundo año al identificarse 50,0 % de incisivos fáciles en comparación con el valor de referencia esperado, pero hubo correlación entre los ítems medianamente fáciles (20,0 %) y difíciles (5,0 %) con su par esperado, respectivamente.

**Tabla 3.** Incisos reales y su índice de dificultad esperada en el examen

Incisos según su índice de dificultad	Incisos esperados		% de incisos esperados	Incisos reales		% de incisos reales	
	Primer año	Segundo año	Primero y segundo años	Primer año	Segundo año	Primer año	Segundo año
Fáciles	2	2	5,0	24	22	58,5	50,0
Medianamente fáciles	8	9	20,0	7	9	17,1	20,4
Dificultad media	20	22	50,0	7	11	17,1	25,0
Medianamente difíciles	8	9	20,0	3		7,3	
Difíciles	2	2	5,0		2		4,5
Total	40	44	100,0	41	44	100,0	100,0

Como se aprecia en la tabla 4, para los residentes de primer año el índice de dificultad para las temáticas de urgencias médicas (0,6), envejecimiento (0,8) y crisis no transitorias de la familia (0,8) fue el más bajo, y las preguntas se clasificaron en el rango de dificultad media. Según los tipos de preguntas fueron más frecuentes las de verdadero o falso y de emparejamiento.

**Tabla 4.** Examen de 1er año por preguntas, incisos y contenidos temáticos según tipo de pregunta e índice de dificultad

Pregunta	Número de incisos	%	Contenidos temáticos	Tipo de preguntas	Índice de dificultad
1	4	10	Análisis de la situación de salud	Opción múltiple	0,9
2	4	10	La familia y etapas del ciclo vital	Emparejamiento	1,0
3	5	12	Alimentación y nutrición en el embarazo	Verdadero o falso	0,9
4	4	10	Grandes problemas geriátricos	Respuesta corta	1,0
5	4	10	Lesiones traumáticas del sistema osteomioarticular (SOMA)	Emparejamiento	0,9
6	4	10	Respuesta a las urgencias médicas	Verdadero o falso	0,6
7	3	7	Cetoacidosis diabética	Opción múltiple	0,9
8	4	10	Cambios fisiológicos en el envejecimiento	Respuesta corta	0,8
9	5	12	Nutrición en el niño	Verdadero o falso	0,9
10	4	10	La familia y crisis no transitorias	Emparejamiento	0,8
Total	41	100,0			

Respecto al segundo año de la especialidad, las interrogaciones con mayores problemas fueron aquellas cuyos contenidos se relacionaron con tuberculosis, parasitismo intestinal, enfermedad cerebrovascular y suicidio, con valores inferiores a 0,9, por lo cual se evaluaron en el rango de dificultad media. Asimismo, hubo un predominio de las preguntas de respuesta corta (tabla 5).

**Tabla 5.** Examen de segundo año por preguntas, incisos y contenidos temáticos según tipo de pregunta e índice de dificultad

Preguntas	Número de incisos	%	Contenidos temáticos	Tipo de preguntas	Índice de dificultad
1	4	9	Tuberculosis	Verdadero o falso	0,8
2	5	11,4	Insuficiencia cardiaca	Respuesta corta	1,0
3	5	11,4	Parasitismo intestinal	Emparejamiento	0,7
4	4	9,0	Diabetes Mellitus	Verdadero o falso	0,9
5	4	9,0	Vulvovaginitis infecciosa	Emparejamiento	0,9
6	5	11,4	Anemia	Respuesta corta	0,9
7	4	9,0	Enfermedad cerebrovascular	Verdadero o falso	0,6
8	4	9,0	Dengue	Respuesta corta	1,0
9	4	9,0	Suicidio	Respuesta corta	0,6
10	5	11,4	Conjuntivitis	Emparejamiento	0,9
Total	44	100,0			

Teniendo en cuenta los valores del índice de discriminación, en el primer año de la especialidad, sobre la base de su poder discriminativo, las temáticas exploradas pudieron agruparse en 3 grupos. En el primero se encontraban los reactivos que discriminaron de forma excelente, con valores superiores a 0,39 (análisis de la situación de salud, la familia y etapas del ciclo vital, alimentación y nutrición en el embarazo, entre otras); en el segundo, el reactivo al discriminar con valor igual a 0,20 sugirió la necesidad de revisar (cetoacidosis diabética), y el último, tenía valores inferiores al anterior (temáticas restantes) que discriminaban pobremente e indicaban descartarlos o su revisión profunda. Por otra parte, los valores  $r_{pbis}$ , aseguraron para 9 de los reactivos un excelente o buen poder de discriminación con cifras superiores a 0,26, y un reactivo con un débil poder discriminativo con valor igual a 0 (lesiones traumáticas del SOMA).

En el segundo año de la residencia predominaron los reactivos que discriminaron excelentemente al superar el valor de referencia 0,39. Igualmente, en orden de frecuencia (D 0,30-0,39) se situaron las temáticas exploradas en las preguntas 4 y 8 (diabetes mellitus y dengue), donde a pesar de que fue buena la calidad se sugirió mejorarla. El reactivo 2 (insuficiencia cardiaca) presentó un pobre poder discriminativo. Todos los valores  $r_{pbis}$  superaron la referencia 0,35 con excelente nivel de discriminación.

## DISCUSIÓN

Los participantes en este estudio consideraron que la relación entre fondo de tiempo e ítem del examen del primer año de la residencia fue muy adecuada en comparación con el segundo año. El coeficiente de correlación de Spearman demostró eficiencia en el grado de variación entre las variables, pues a medida que aumentan las horas asignadas al contenido debe suceder igual con el número de incisos para su exploración. Contrario a lo deseado para el instrumento del segundo año de la especialidad y en concordancia con los resultados obtenidos por Díaz *et al*,<sup>2</sup> quienes incluso con un coeficiente de correlación de 0,55 aseguran que no siempre se obtienen los resultados esperados.

Se observó además, que hubo una alta frecuencia de incisos fáciles en el examen teórico de control parcial para ambos años de la especialidad. Resulta importante señalar que el factor intencional nunca primó en el proceso, pues para los autores las diferencias entre lo real y esperado generó asombro, así como advertencias futuras para mejorar la calidad de la exploración del conocimiento. Los resultados de otros autores<sup>8-10</sup> muestran similitud con lo encontrado en esta investigación.

Para ambos exámenes, los valores del índice de dificultad en la mayoría de las preguntas fueron expresión de una adecuada proporción de alumnos que contestaron correctamente. Ello pudiera relacionarse con el hecho de que más de la mitad de los incisos fueron clasificados como fáciles; sin embargo, resultó llamativo que en preguntas de respuesta corta o de verdadero o falso, con 4 incisos para su exploración, en temas tales como urgencias médicas, enfermedad cerebrovascular y contenido de enfermedades psiquiátricas como el suicidio, el índice de dificultad obtenido haya sido el más bajo, lo cual mostró que fue difícil resolver la pregunta. De la misma manera, en otros estudios<sup>7,8</sup> se constató la variabilidad de este indicador.

Al valorar los resultados del índice de discriminación se interpretó que en las temáticas sobre problemas geriátricos, lesiones traumáticas del SOMA e insuficiencia cardiaca hubo pobre distinción entre los residentes con mayor y menor rendimiento académico, a pesar de que los primeros, a juzgar por el coeficiente de discriminación, obtuvieron el mayor número de respuestas correctas, lo que finalmente sugiere la necesidad de revisar estas preguntas. No obstante, en la comparación con el  $\alpha_{\text{examen}} = 1$ , cada pregunta aportó consistencia interna a su trabajo de control parcial.

Los indicadores empleados permitieron evaluar la calidad del trabajo de control parcial en los 2 años de la residencia, en los cuales se hallaron dificultades y fortalezas con respecto a su planeación. Se recomendó para las próximas convocatorias, elevar el nivel de profundidad en la evaluación del sistema de conocimientos, hábitos, habilidades y modos de actuación del residente.

## REFERENCIAS BIBLIOGRÁFICAS

1. Cuba. Ministerio de Salud Pública. Reglamento del régimen de residencia en ciencias de la salud. Resolución 108. La Habana: MINSAP; 2004.
2. Díaz Rojas PA, Leyva Sánchez E. Metodología para determinar la calidad de los instrumentos de evaluación. *Educ Med Super*. 2013 [citado 2 Ago 2014]; 27(2).
3. Escudero B, Larrazolo Reyna N, Rosas Morales M. Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Rev Electrónica de Inv Educ*. 2000 [citado 12 Jul 2014]; 2(1).
4. Cuba. Ministerio de Educación Superior. Reglamento de Trabajo Docente y Metodológico. Resolución 210. La Habana: Gaceta Oficial de la República de Cuba; 2007.
5. Escobar Yéndez NV, Almaguer Delgado AJ, Plasencia Asorey CG. Periodización de la formación de posgrado del médico en etapa de especialización en medicina general integral: dimensiones, variables y claves hermenéuticas. *MEDISAN*. 2010 [citado 2014 Ago 22]; 14(2).
6. Martínez Ortega RM, Tuya Pendás LC, Martínez Ortega M, Pérez Abreu A, Cánovas AM. El coeficiente de correlación de los rangos de spearman. Caracterización. *Rev haban cienc méd*. 2009 [citado 12 Jul 2014]; 8(2).

7. Sánchez González MG, Tapia Pérez GG. Descripción del nivel de facilidad y poder de discriminación del examen de inferencia estadística en métodos estadísticos en Medicina veterinaria y Zootecnia. Jornadas de Educación Médica. 2010 [citado 2 Ago 2014].
8. Abreu Reyes DR. Caracterización del examen teórico final de Morfofisiología Humana del perfil Laboratorio Clínico. Educ Med Super. 2011 [citado 22 Ago 2014]; 25(1).
9. Moreno Montañez M, Quintana Regalado G. Caracterización del examen estatal escrito en la especialidad de Medicina General Integral. Educ Med Super. 2009 [citado 22 Ago 2014]; 23(3).
10. González Pérez M. La evaluación del aprendizaje: tendencias y reflexión crítica. Educ Med Super. 2001 [citado 22 Ago 2014]; 15(1).

Recibido: 2 de octubre del 2014.

Aprobado: 5 de noviembre del 2014.

*Ernesto Sánchez Hernández.* Hospital Clínicoquirúrgico Docente "Dr. Joaquín Castillo Duany", Punta Blanca s/n, Santiago de Cuba, Cuba.  
Correo electrónico: [Esanchez@hmscu.scu.sld](mailto:Esanchez@hmscu.scu.sld)