Descargado el: 5-09-2026 ISSN 1727-897X

ARTÍCULO ORIGINAL DE INVESTIGACIÓN PEDAGÓGICA

Cuestionarios diferentes en la evaluación escrita. A propósito de los resultados en un trabajo de control parcial de Medicina Interna

Different questionnaires in written evaluation. Regarding the results of an internal medicine mid term test

Luis Alberto Corona Martínez[1] Mercedes Fonseca Hernández[2] Yenisey Quintero Méndez[1] Caridad Hernández Abreu[1] Orestes Fresneda Quintana[1]

[1] Hospital General Universitario Dr. Gustavo Aldereguia Lima, Cienfuegos, Cuba
[2] Hospital Pediátrico Universitario Paquito González Cueto, Cienfuegos, Cuba

Cómo citar este artículo:
Corona-Martínez L, Fonseca-Hernández M, Quintero-Méndez Y, Hernández-Abreu C, Fresneda-Quintana O. Cuestionarios diferentes en la evaluación escrita. A propósito de los resultados en un trabajo de control parcial de Medicina Interna. **Medisur** [revista en Internet]. 2017 [citado 2026 Sep 5]; 15(5):[aprox. 9 p.]. Disponible en: https://medisur.sld.cu/index.php/medisur/article/view/3703

Resumen

Fundamento: la preocupación por la calidad de los instrumentos de evaluación ha sido compartida por varios autores. Algunos han hecho alusión a la repercusión de las deficiencias en dichos instrumentos en los resultados académicos.
Objetivo: determinar la posible influencia de diferencias (de contenido y de formato) entre los instrumentos aplicados en una evaluación parcial, en las calificaciones obtenidas por los estudiantes.
Métodos: en cada instrumento se revisó el contenido explorado y el formato de las preguntas utilizadas. Se determinó la distribución de las calificaciones de forma comparativa en cada uno de los cuestionarios, tanto en el examen en su conjunto como en las diferentes preguntas de forma independiente. Los resultados fueron contrastados con los del mismo trabajo de control del curso precedente.
Resultados: fueron constatadas diferencias significativas estadísticamente en las calificaciones entre ambos instrumentos del curso 2016-2017, con un franco predominio de las calificaciones de mejor calidad en el cuestionario A (58 % vs 30 %); la probabilidad de desaprobar el examen también fue significativamente mayor en el cuestionario B. En cambio, se comprobó la existencia de escasas diferencias (no significativas) en la distribución de las calificaciones del primer trabajo de control del curso 2015-2016.
Conclusiones: la realización de exámenes en los cuales son aplicados cuestionarios que difieren entre sí, ya sea en el formato de las preguntas o en su contenido, puede determinar la ocurrencia de resultados diferentes en las calificaciones, favoreciendo a unos estudiantes sobre otros. Se sugiere incorporar la cualidad “justeza” a las características generales (ya reconocidas en la literatura) que debe tener todo instrumento de evaluación.

Palabras clave: evaluación, estudiantes de medicina, medicina interna

Abstract

Foundation: the concern for the quality of evaluation instruments has been shared by several authors. Some have alluded to the impact of the deficiencies in such instruments on academic outcomes.
Objective: to determine the possible influence of differences (content and format) among the instruments applied in a mid term test, in the scores obtained by the students.
Methods: in each instrument the content explored and the format of the questions used were reviewed. The distribution of the scores was determined in a comparative way in each one of the batteries, both in the examination as a whole and in the different questions independently. The results were contrasted with those of the same control work of the previous course.
Results: statistically significant differences were found in the scores between the two instruments of the 2016-2017 academic course, with a frank predominance of the best quality scores in Questionnaire A (58 % vs 30 %); the likelihood of failing the test was also significantly higher in Questionnaire B. On the other hand, there was little (non-significant) difference in the distribution of the scores of the first mid term test of the 2015-2016 academic year.
Conclusions: conducting exams in which different questionnaires are applied, either in questions format or in their contents, can determine the occurrence of different results in scoring, favoring some students over others. It is suggested to incorporate the "fairness" quality to the general characteristics (already recognized in literature) which any evaluation instrument should have.

Key words: evaluation, students, medical, internal medicine

Aprobado: 2017-09-27 08:43:55

Correspondencia: Luis Alberto Corona Martínez. Hospital General Universitario Dr. Gustavo Aldereguia Lima. Cienfuegos. luis.corona@gal.sld.cu

INTRODUCCIÓN

El proceso docente educativo está constituido por diversos componentes relacionados entre sí en forma de un sistema; entre estos componentes se encuentra uno que hace referencia a los resultados de este proceso.[1-3]

Expresado de otra manera, el proceso docente posee un componente cuya principal función es informar a estudiantes y profesores acera de la medida en que se cumplen los objetivos formativos definidos para los diferentes niveles de sistematización de dicho proceso. Este componente se concreta en la denominada evaluación del aprendizaje.[1-4] La evaluación constituye un subsistema dentro del sistema mayor: el proceso docente educativo.

Para muchas disciplinas y asignaturas de la carrera de Medicina dicho subsistema se encuentra configurado en diferentes niveles, en correspondencia a su vez con los niveles de generalización de la categoría rectora del proceso docente educativo: los objetivos de aprendizaje. Según el planteamiento anterior, la evaluación puede ser frecuente, parcial o final.[5]

En el caso concreto de la evaluación parcial, esta evaluación se relaciona con el grado de cumplimiento no de objetivos específicos o generales (con lo cual se relacionan la evaluación frecuente y la evaluación final, respectivamente), sino con la manera en que el estudiante va adquiriendo la capacidad de integrar contenidos correspondientes a varios temas de la asignatura, y que implica a su vez, la consecución integrada de un conjunto de objetivos específicos.

En la asignatura Medicina Interna, enmarcada en el sexto semestre de la carrera, la evaluación parcial se lleva a cabo a través de los llamados “trabajos de control”. Este trabajo de control consiste concretamente en un examen escrito que se aplica al unísono a la totalidad de los estudiantes, en el cual es evaluado el aprendizaje de los contenidos (específica y casi exclusivamente la base teórica) derivados de los objetivos de un grupo de temas de la asignatura.

El origen de este trabajo emana del análisis realizado en la Cátedra de Clínica Médica sobre los resultados del primer trabajo de control de Medicina Interna del curso 2016-2017; en este análisis se puso de manifiesto la existencia de evidentes diferencias en la distribución de las calificaciones entre los dos instrumentos (cuestionarios) utilizados. A diferencia del curso precedente, ambos cuestionarios difirieron, en alguna medida, en cuanto al contenido explorado y al formato de las preguntas empleadas. El objetivo del trabajo radica en la determinación de la posible influencia de las diferencias mencionadas entre los instrumentos aplicados, en las calificaciones obtenidas por los estudiantes.

MÉTODOS

Estudio realizado por profesores de la cátedra de Medicina Interna del Hospital Gustavo Aldereguía de Cienfuegos.

En cada instrumento se revisó el contenido explorado y el formato de las preguntas utilizadas; para esto último fue utilizado un instructivo metodológico acerca de la construcción de instrumentos evaluativos escritos.[6]

Posteriormente se determinó la distribución de las calificaciones según el sistema de símbolos establecido en la educación superior cubana: 5 (excelente), 4 (bien), 3 (regular) y 2 (mal), de forma comparativa en cada uno de los instrumentos; dicha distribución fue analizada tanto en el examen en su conjunto como en las diferentes preguntas de forma independiente. Para este análisis fueron revisados los exámenes individuales de la totalidad de los alumnos que participaron en la actividad docente (284, distribuidos a la mitad en cada batería).

Con la intención de facilitar la interpretación de las diferencias entre los cuestionarios del primer trabajo de control de Medicina Interna del curso 2016-2017, fueron comparadas las calificaciones del primer trabajo de control de Medicina Interna del curso precedente (2015-2016). Este examen también contó con dos cuestionarios, pero con la diferencia de que ambos exploraron el mismo contenido y con el mismo formato de las preguntas, por lo que las diferencias solo fueron en el orden de los incisos en las preguntas tipo “respuestas cortas” y “de apareamiento”. En este último examen participaron 211 estudiantes.

Los resultados son expuestos en gráficos que expresan la distribución porcentual de las frecuencias de las distintas calificaciones obtenidas por los estudiantes. En el análisis estadístico fue utilizado el estadígrafo Chi^2 para diferencias de proporciones y, en aquellas

situaciones que lo permitieron, el Odds Ratio (OR) y su intervalo de confianza; en todos los casos se aceptó un nivel de confianza de 95 % (p <0.05).

La investigación fue aprobada por el consejo científico de la institución.

RESULTADOS

Descripción de los instrumentos utilizados en el curso 2016-2017

Ambos instrumentos estuvieron conformados por tres preguntas. La primera abordó los contenidos relacionados con las intoxicaciones exógenas, mientras que la segunda y tercera exploraron el tema de las enfermedades infecciosas y parasitarias. Las diferencias entre los instrumentos A y B fueron las siguientes:

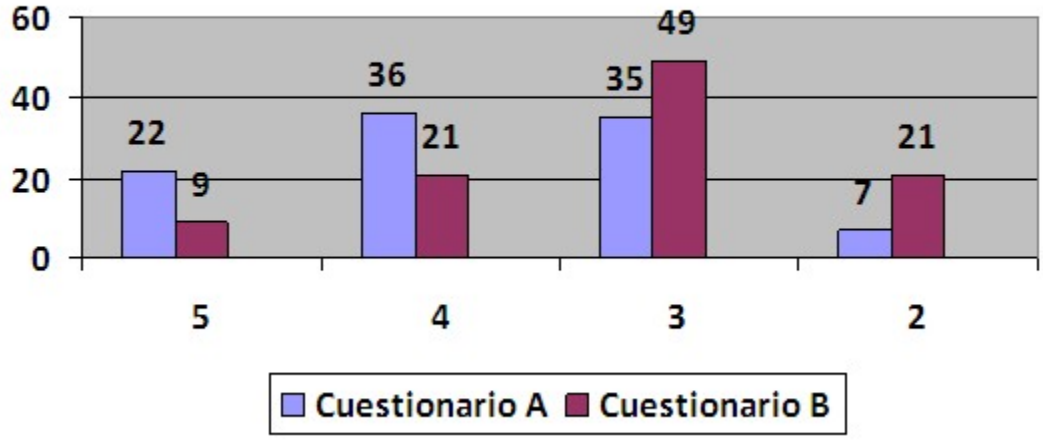
- La pregunta 1 en el cuestionario A fue elaborada con un formato correspondiente a "pregunta de respuestas cortas", mientras que en el cuestionario B se utilizó el formato de "pregunta de apareamiento".
- La pregunta 2 abordó contenidos diferentes: paludismo y dengue respectivamente.
- La pregunta 3 en el cuestionario A fue elaborada con un formato correspondiente a "pregunta de apareamiento", mientras que en el cuestionario B se utilizó el formato de "pregunta de respuestas cortas".

Como se puede observar las tres preguntas difirieron, ya sea en su formato (preguntas 1 y 3) o en su contenido (pregunta 2). ¿Condicionarían estas diferencias a la ocurrencia de resultados diferentes en las calificaciones de los estudiantes en los instrumentos?

Distribución comparativa de las calificaciones en los cuestionarios A y B, curso 2016-2017.

El analizar las calificaciones en el trabajo de control según cuestionario, se pudo observar la existencia de evidentes diferencias en las calificaciones entre ambos instrumentos, con un franco predominio de las calificaciones de mejor calidad en el cuestionario A (58 % vs 30 %) y las de menor calidad en el cuestionario B (70 % vs 42 %); diferencias que fueron significativas desde el punto de vista estadístico (OR= 3,15[1,88;5,29]; X^2=21,73, p<0.05). La probabilidad de desaprobar el examen también fue significativamente mayor en el cuestionario B (OR= 3,19[1,45;7,12]; X^2=10,29, p<0.05). (Gráfico 1).

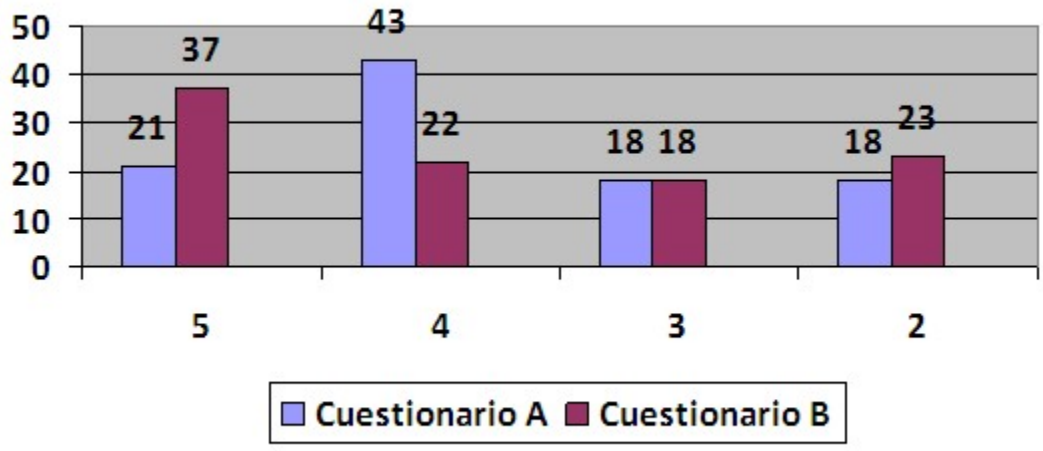
Gráfico 1. Distribución porcentual de las calificaciones en el trabajo de control según cuestionario. Curso 2016-2017



(Cuestionario A: n=142; Cuestionario B: n=142)

El patrón descrito anteriormente para el examen como un todo, consistente en diferencias evidentes en las calificaciones entre ambos cuestionarios, fue observado también en cada una de las tres preguntas de dicho examen. En el caso de la pregunta uno, aunque las diferencias entre los cuestionarios fueron significativas desde el punto de vista estadístico, no quedó claro un sentido definido de esas diferencias, pues si bien en el cuestionario B el porcentaje de estudiantes con 5 fue mayor, también este cuestionario mostró un porcentaje más elevado de desaprobados. (Gráfico 2).

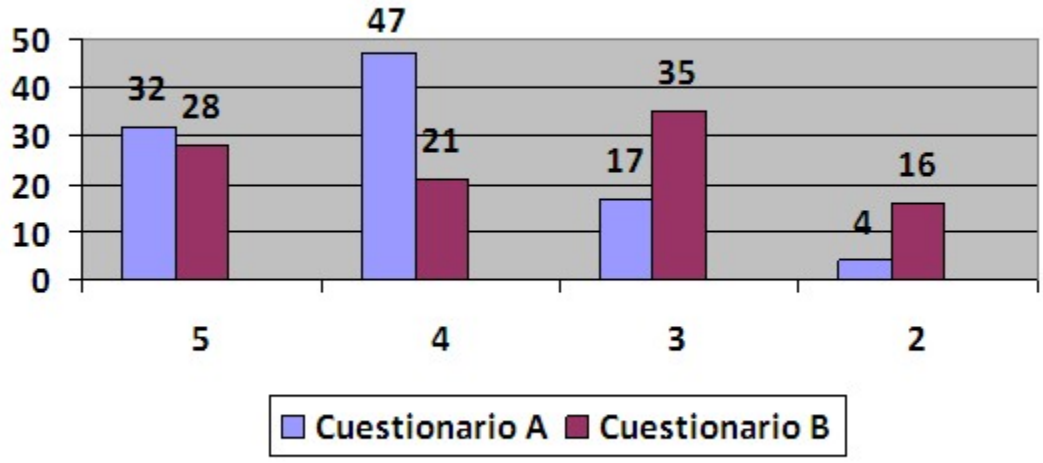
Gráfico 2. Distribución porcentual de las calificaciones en la pregunta 1 del trabajo de control según cuestionario. Curso 2016-2017



(Cuestionario A: n=142; Cuestionario B: n=142)
X^2=17,28; p<0.05

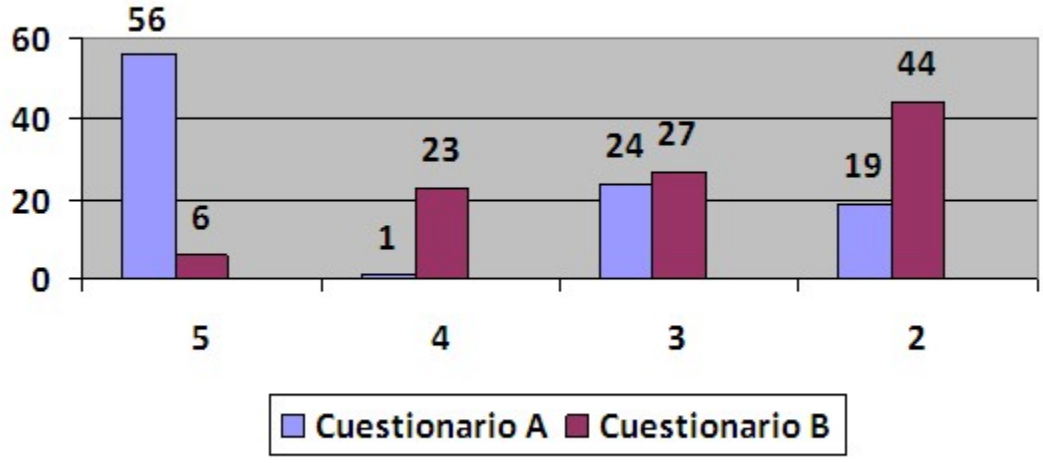
La mayor probabilidad de calificaciones de mejor calidad y la menor probabilidad de desaprobar fueron claramente observadas en el cuestionario A, tanto en la pregunta 2 (Gráfico 3) como en la pregunta 3 (gráfico 4). En ambos casos las diferencias fueron significativas desde el punto de vista estadístico (pregunta 2: OR= 3,95[2,28;6,88]; X^2=28,17, p<0.05 para la calidad de la calificación y OR= 4,38[1,62;12,46]; X^2=11,1, p<0.05 para la probabilidad de desaprobar; pregunta 3: OR= 3,27[1,94;5,52]; X^2 =22,99, p<0.05 para la calidad de la calificación y OR= 3,3[1,87;5,84]; X^2=20,05, p<0.05 para la probabilidad de desaprobar). (Gráficos 3 y 4)

Gráfico 3. Distribución porcentual de las calificaciones en la pregunta 2 del trabajo de control según cuestionario. Curso 2016-2017



(Cuestionario A: n=142; Cuestionario B: n=142)
X^2=34,4; p<0.05

Gráfico 4. Distribución porcentual de las calificaciones en la pregunta 3 del trabajo de control según cuestionario. Curso 2016-2017



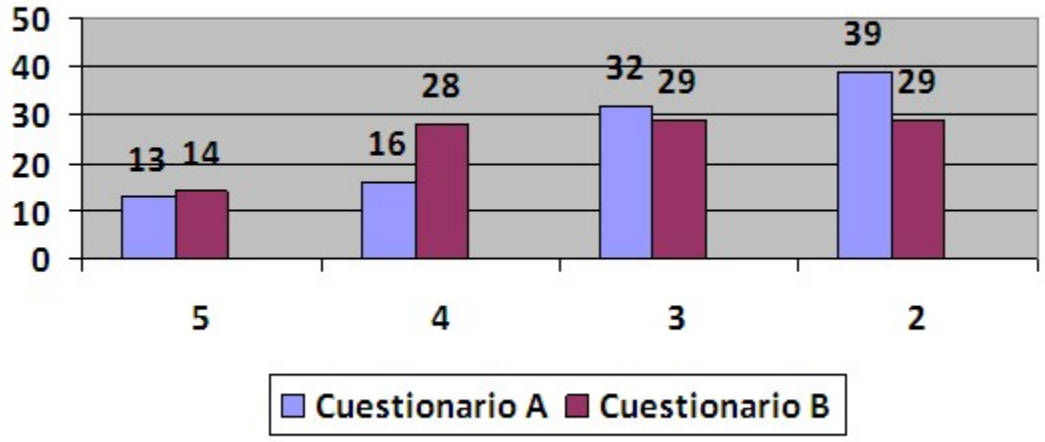
(Cuestionario A: n=142; Cuestionario B: n=142)
X^2=99,51; p<0.05

Distribución comparativa de las calificaciones en los cuestionarios A y B, curso 2015-2016.

Se comprobó la existencia de pocas diferencias en la distribución de las calificaciones entre los cuestionarios del primer trabajo de control del

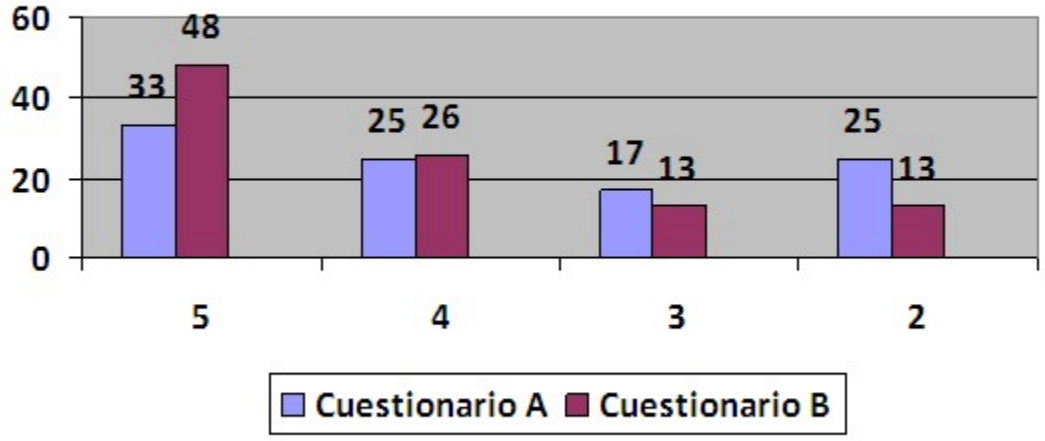
curso 2015-2016, aunque siempre ligeramente mejores en el cuestionario B; estas diferencias no mostraron significación estadística, ni para la calificación del instrumento en su conjunto (Gráfico 5), ni en el análisis de cada una de las tres preguntas que lo constituyeron. (Gráficos 6, 7 y 8).

Gráfico 5. Distribución porcentual de las calificaciones en el trabajo de control según cuestionario. Curso 2015-2016



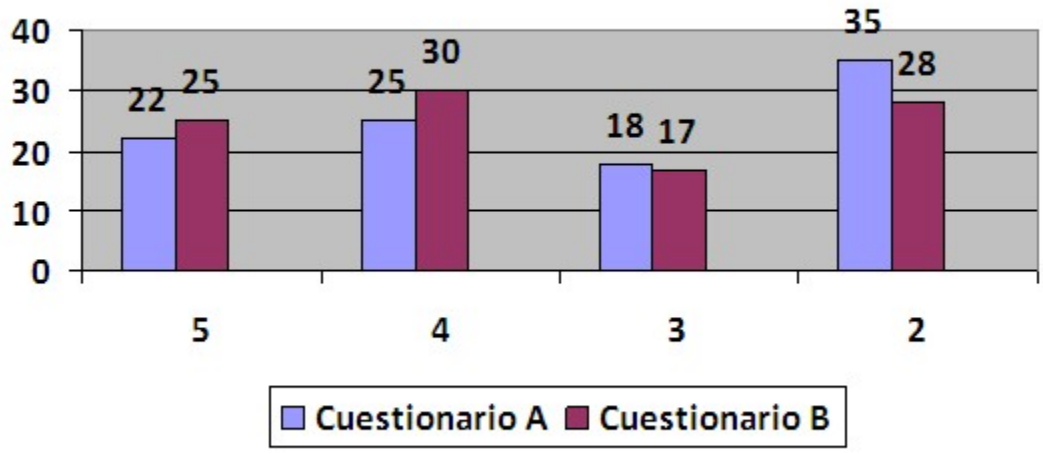
(Cuestionario A: n=106; Cuestionario B: n=105)
p>0.05

Gráfico 6. Distribución porcentual de las calificaciones en la pregunta 1 del trabajo de control según cuestionario. Curso 2015-2016



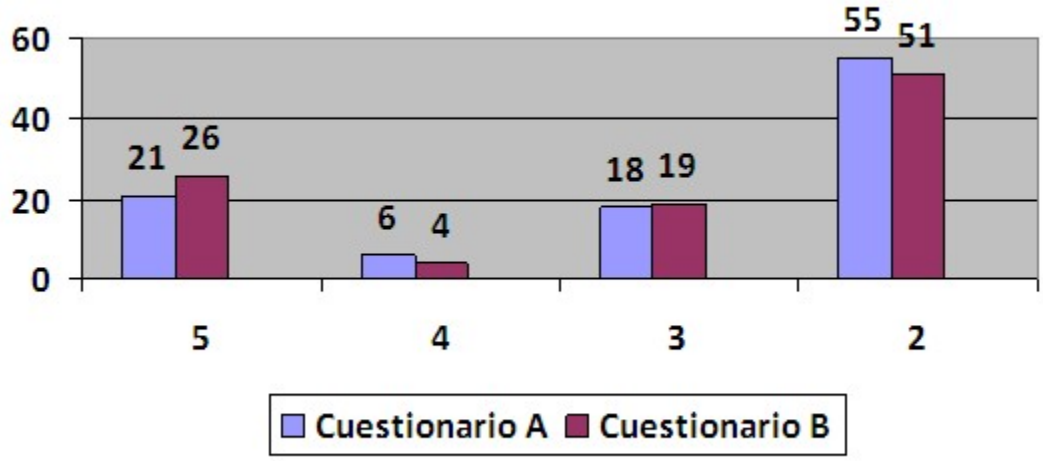
(Cuestionario A: n=106; Cuestionario B: n=105)
p>0.05

Gráfico 7. Distribución porcentual de las calificaciones en la pregunta 2 del trabajo de control según cuestionario. Curso 2015-2016



(Cuestionario A: n=106; Cuestionario B: n=105)
p>0.05

Gráfico 8. Distribución porcentual de las calificaciones en la pregunta 3 del trabajo de control según cuestionario. Curso 2015-2016



(Cuestionario A: n=106; Cuestionario B: b=105)
p>0.05

DISCUSIÓN

La preocupación por la calidad de los instrumentos de evaluación ha sido compartida por varios autores.[7-13] Igualmente, otros han hecho alusión a la repercusión de las deficiencias en dichos instrumentos en los resultados académicos.[14,15] Estos elementos han determinado el reconocimiento de necesidades en el aprendizaje sobre la elaboración de instrumentos evaluativos.[16,17]

El desigual comportamiento de las diferencias en las calificaciones de los estudiantes entre los cuestionarios, en los cursos 2015-2016 y 2016-2017, amerita un comentario. Las diferencias intercuestionarios en el primer curso fueron mínimas, sin alcanzar significación estadística; en cambio, en el curso 2016-2017 las diferencias fueron evidentes, hasta el punto de ser estadísticamente significativas en los distintos análisis realizados.

En el caso de los resultados del primer curso, el hecho de ser los cuestionarios iguales "en contenido" y "en formato de preguntas" conduce a interpretar las pequeñas diferencias intre ellos como consecuencia de las inevitables diferencias individuales entre los estudiantes que conformaron al azar cada uno de los dos subgrupos de cuestionarios. En cambio, dada su magnitud, las diferencias entre los cuestionarios A y B en el curso académico 2016-2017 no pueden ser explicadas solamente por la mencionada razón.

Es lógico pensar que explorar contenidos diferentes pudiera constituir una potencial fuente de diferencias en las calificaciones, aún en presencia de un mismo formato de la pregunta en ambos cuestionarios del instrumento evaluativo. "Un contenido" pudiera ser de más complejidad que el "otro contenido", en cuanto al grado de dominio por el estudiante (de dichos contenidos) que se espera para un momento específico del proceso docente educativo. O el nivel de profundidad en la exploración, concretado en la clave de calificación, también pudiera ser diferente entre los cuestionarios para preguntas de un mismo tema que abordan contenidos diferentes. También la norma para la calificación de una pregunta pudiera ser más exigente que la norma para la otra pregunta.

De cualquier manera, independientemente de la explicación concreta, este fenómeno parece haber estado presente en nuestro objeto de estudio.

En el curso 2015-2016, la pregunta de desarrollo del trabajo de control fue dedicada al tema de las enfermedades infecciosas y dentro de estas al contenido correspondiente a leptospirosis; en ambas baterías se aplicó la misma pregunta, y los resultados fueron muy similares (gráfico 7). En cambio, en el curso 2016-2017 fueron abordados contenidos diferentes en esa misma pregunta: paludismo y dengue; las diferencias en las calificaciones entre los cuestionarios fueron ostensibles (gráfico 3).

De igual manera, se pudiera asumir que un mismo contenido explorado de forma diferente también puede conllevar a desigualdades en las calificaciones.

Si tomamos como ejemplo la pregunta 3 del trabajo de control del curso 2016-2017 (gráfico 4), se pone de manifiesto una "desventaja" para los estudiantes cuando se utiliza el formato correspondiente a las "preguntas de respuestas cortas" en comparación con el formato correspondiente a las "preguntas de apareamiento". En las primeras el estudiante tiene que generar la respuesta a partir de la información que se le ofrece; en las segundas el estudiante debe reconocer la respuesta en un listado de opciones. Realmente, "generar" y "reconocer" no expresan un mismo nivel del grado de dominio de un contenido.

Contribuye a reforzar el análisis anterior la observación de los resultados en las preguntas 1 y 3 del examen correspondiente al curso 2015-2016, las cuales contaron con el mismo formato en ambos cuestionarios ("respuestas cortas" para la pregunta 1, "de apareamiento" para la pregunta 3): las diferencias entre los cuestionarios en la distribución de las calificaciones fueron menores, sin significación estadística (gráficos 6 y 8).

A partir de estos elementos se puede inferir que la realización de exámenes en los cuales son aplicados cuestionarios que difieren entre sí, ya sea en el formato de las preguntas o en su contenido, puede determinar la ocurrencia de resultados diferentes en las calificaciones, favoreciendo a unos estudiantes y en detrimento de otros. En nuestro estudio, este planteamiento queda fundamentado en los desiguales resultados mostrados en los gráficos 1 y 5 correspondientes a las calificaciones en el primer trabajo de control de ambos cursos académicos.

Es precisamente en este punto donde la anterior inferencia conduce a una interrogante de naturaleza ética para la cual parece improbable encontrar respuesta en la teoría didáctica: ¿por qué son necesarios dos instrumentos para examinar a una misma cohorte de estudiantes?

En opinión de los autores, la evaluación parcial (y también la final) sustentada en exámenes escritos debe garantizar, adicionalmente a los atributos de validez, confiabilidad, objetividad, amplitud, utilidad, practicabilidad (factibilidad) y pertinencia ya conocidos,[6] la igualdad de oportunidades en los estudiantes para demostrar el nivel de dominio de los contenidos, de forma tal que las desigualdades que se obtengan en las calificaciones sean atribuibles a las reales o verdaderas diferencias individuales en el grado de consecusión de un objetivo o, al menos, que dichas diferencias estén "descontaminadas" de la posible variabilidad determinada por las diferencias en los instrumentos aplicados, emanadas estas del proceso de elaboración de los mismos.

De esta manera se lograría una evaluación más "justa"; asumida dicha cualidad desde la perspectiva de los estudiantes en cuanto a la interpretación de los resultados individuales; perspectiva que si bien no es esencial para el cumplimiento de las funciones más relevantes de la evaluación como componente del proceso docente educativo, no debería ser ignorada al tener en cuenta el lugar protagónico del estudiante en este proceso.

REFERENCIAS BIBLIOGRÁFICAS

1. Álvarez de Zayas CM. La escuela en la vida. Didáctica. La Habana: Editorial Pueblo y Educación; 1999.

2. Compendio de Pedagogía. La Habana: Editorial Pueblo y Educación; 2002.

3. Álvarez de Zayas CM. Fundamentos teóricos de la dirección del proceso docente educativo en la Educación Superior Cubana. La Habana: Ministerio de Educación Superior; 1990.

4. Salas Perea RS. Educación en salud. Competencia y desempeño profesionales. La Habana: Editorial Ciencias Médicas; 1999.

5. Ministerio de Salud Pública. Normas de trabajo metodológico de la Educación Médica Superior (V.A.D.3/90). La Habana: MINSAP; 1990.

6. García Hernández I, Salas Perea R. La construcción de instrumentos evaluativos escritos. Elementos para mejorar su calidad: Carpeta No. 2. La Habana: MINSAP; 1997.

7. Damiani Cavero JS, Vicedo Tomey AG, Sierra Figueredo S, Fernández Asan A, Pernas Gómez M, Blanco Aspiazú MÁ, et al. Diseño y validación de un instrumento para valorar las funciones especiales en estudiantes de sexto año de la carrera de Medicina. Educ Med Super [revista en Internet]. 2015 [cited 5 Jun 2017] ; 29 (1): [aprox. 17p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412015000100007.

8. Blanco Aspiazú MÁ, Rodríguez Collar T, Blanco Aspiazú O, Hernández Díaz L. Enfermedades de la evaluación. Educ Med Super [revista en Internet]. 2013 [cited 5 Jun 2017] ; 27 (2): [aprox. 9p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412013000200012.

9. Díaz Rojas PA, Leyva Sánchez E. Metodología para determinar la calidad de los instrumentos de evaluación. Educ Med Super [revista en Internet]. 2013 [cited 5 Jun 2017] ; 27 (2): [aprox. 17p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412013000200014.

10. Pérez Reyes R, Otero Ceballo M. Evaluación de la competencia en médicos intensivistas. Rev Cub Med Mil [revista en Internet]. 2003 [cited 5 Jun 2017] ; 32 (2): [aprox. 7p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-65572003000200010.

11. López Fernández R, Chou Rodríguez R, Vázquez Cedeño S, Avello Martínez R, Gómez Ortega E, Sánchez Gálvez S. Transformación a los instrumentos de evaluación a pequeños grupos de estudiantes. Medisur [revista en Internet]. 2015 [cited 5 Jun 2017] ; 13 (1): [aprox. 8p]. Available from: http://www.medisur.sld.cu/index.php/medisur/article/view/2938.

12. Trujillo Saínz C, Guerra Pando JA, Henríquez Trujillo D. Percepción de profesores y directivos de morfofisiología sobre esencialidades de la confección de instrumentos evaluativos finales. Rev Ciencias Médicas [revista en Internet]. 2013 [cited 5 Jun 2017] ; 17 (4): [aprox. 9p]. Available

from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-31942013000400016&lng=es.

13. Sánchez Hernández E, Medina Pavón M, Rodríguez García M, Vega Van Der Meer L, de la Torre Vega G. Indicadores de calidad para un examen teórico de la especialidad de medicina general integral. MEDISAN [revista en Internet]. 2015 [cited 5 Jun 2017] ; 19 (2): [aprox. 7p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192015000200002.

14. Ortiz Romero GM, Rodríguez Neyra ME, Díaz Rojas PA, Cuenca Font K. Perfeccionamiento de la calidad de los instrumentos evaluativos en la asignatura Metodología de la Investigación y Estadística. CCM [revista en Internet]. 2014 [cited 5 Jun 2017] ; 18 (4): [aprox. 11p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1560-43812014000400012.

15. Abreu Reyes DR. Caracterización del examen teórico final de Morfofisiología Humana del perfil Laboratorio Clínico. Educ Med Super [revista en Internet]. 2011 [cited 5 Jun 2017] ; 25 (1): [aprox. 18p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000100003.

16. López Espinosa GJ, Quintana Mugica R, Rodríguez Cruz O, Gómez López L, Pérez de Armas A, Aparicio Manresa G. El profesor principal y su preparación para diseñar instrumentos de evaluación escritos. EDUMECENTRO [revista en Internet]. 2014 [cited 5 Jun 2017] ; 6 (2): [aprox. 10p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742014000200007.

17. Cordero Miranda Y. Necesidades de aprendizaje sobre la elaboración de instrumentos evaluativos escritos. Educ Med Super [revista en Internet]. 2013 [cited 5 Jun 2017] ; 27 (2): [aprox. 7p]. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412013000200007.