

INDICADORES DE CALIDAD EN DOS EXÁMENES FINALES DE METODOLOGÍA DE LA INVESTIGACIÓN Y ESTADÍSTICA. 2010-2012.

Autores

- Dr. Agustín Rodríguez Soto. *
- Dra. María Elena Fernández Collazo.**

*Especialista de 1er Grado en Bioestadística y Estomatología General Integral. Master en Educación Médica Superior. Profesor Asistente. Facultad de Estomatología de la Habana. Edif. 3 Apto 7 Zona 1 Alamar. Habana del Este. La Habana. Telef. 052303579 E-mail. arsoto@infomded.sld.cu

**Especialista de 2do Grado en Estomatología General Integral. Master en Urgencias Estomatológicas. Profesor Auxiliar de la Universidad de Ciencias Médicas de La Habana. Mariae.collazo@infomed.sld.cu

Resumen.

Objetivo: Identificar elementos de calidad en los exámenes finales de la asignatura Metodología de la Investigación y Estadística en los cursos 2009-2010 y 2010-2011 en la Facultad de Estomatología de la Habana.

Método: Como objeto de estudio los exámenes finales de la asignatura Metodología de la Investigación y Estadística correspondientes a los dos cursos, realizados y aplicados a los estudiantes del segundo año de la carrera de Estomatología; el universo para el desarrollo de la investigación lo constituyó la totalidad de los exámenes realizados en ordinario, para un total de 201 en el primero y 173 en el segundo curso. Se calcularon los índices de dificultad, de discriminación y el Alfa de Cronbach para medir la confiabilidad.

Resultados: Encontrándose que las preguntas con mayor grado de dificultad presentan un índice de 0,68; que una sola pregunta discrimina por debajo de 0,20 y que la confiabilidad es de 0,50 en el segundo y de 0,86 en el primer curso que se evalúa.

Conclusiones: Los exámenes finales de la asignatura Metodología de la Investigación y Estadística clasifican como fácil, que la discriminación de los exámenes es buena y se presentan con marcada diferencia en la confiabilidad. Palabras claves: Instrumento evaluativo, calidad y Evaluación

Summary.

Objective: Identifying elements of quality in the finals of the subject of study Methodology of Investigation and Estadística at courses 2009-2010 and 2010-2011 in Estomatology's Faculty of the Habana.

Method: As I object of study the finals of the subject of study correspondent Methodology of Investigation and Estadística to two courses, accomplished and applied to the students of the second year of Estomatology's race; The investigation's developmental universe was constituted by the realized

exámenes's totality in ordinary, for a total of 201 in the first and 173 at the second course. The índices of difficulty calculated themselves, discriminating and Cronbach's Alpha to measure the reliability.

Results: Meeting that questions with bigger degree of difficulty show one of 0.68 indexes; That an only question discriminates de 0,20 underneath and that reliability comes from 0.50 in the second one and of 0.86 at the first course that it is evaluated.

Conclusions: The finals of the subject of study Methodology of Investigation and Estadística classify like easy, that the exámenes's discrimination is good and they show up with marked difference in reliability.

Key words: I orchestrate evaluativo, quality and Evaluation

INTRODUCCION.

La evaluación de la docencia universitaria se reconoce como una actividad compleja, entre otras razones, debido a que múltiples factores inciden en el proceso enseñanza-aprendizaje. Su relevancia estriba en la posibilidad de mejorar el desempeño de los docentes para la formación profesional de los estudiantes. Dos de los elementos que se aprecian como indispensables para quienes realizan la evaluación son la capacidad técnica, y la autoridad y legitimidad académica.¹

La evaluación es un proceso que involucra la elaboración, aplicación y análisis de los instrumentos de evaluación. La evaluación según Salas Perea², constituye un componente esencial de cualquier esfuerzo que se haga para producir algo de valor, así como para poder adoptar una buena decisión.

Uno de los puntos más complicados del proceso de enseñanza – aprendizaje, sea bajo la modalidad presencial o bien con el uso de las Nuevas Tecnologías de Información y Comunicación, es la evaluación. Dicha parte del proceso resulta traumática y ansiógena en los educandos, aunque las nuevas corrientes educativas ven en este proceso el único elemento adecuado y eficiente para la evaluación de la efectividad y validez del método utilizado para la transferencia de conocimientos.³

En el desarrollo de las técnicas de evaluación es necesario que se considere la misión de la institución a la que se pertenece. La misión es amplia y de ella se desprenden los objetivos generales que se espera que los estudiantes que pertenecen a determinada institución vayan alcanzando durante su formación académica.⁴

En la literatura aparecen diversos criterios para evaluar la calidad de los instrumentos evaluativos, sin embargo existen indicadores de fácil manejo que pueden ser utilizados a nivel de base por los colectivos de profesores, máxime cuando el desarrollo alcanzado en los Software, disponibles en cualquier unidad educativa permite su cálculo.

A través de los indicadores de calidad como la discriminación el examinador descubre preguntas mal formuladas, inútiles (no discrimina), errores en las (alternativas vacías) y cuando el resultado final se aleja francamente de la medida estadística uno puede sospechar que han ocurrido otros fenómenos no inherentes al instrumento en sí.

Otro es el índice de dificultad del examen, que no es más que la proporción de personas que responden correctamente un reactivo de la prueba, entre mayor sea esta proporción menor será el grado de dificultad.

Gronlund⁶, (2006) y Klecker⁷ (2005) coinciden en la importancia que tiene la validez y la confiabilidad de los resultados obtenidos de algún instrumento de recopilación de datos en la evaluación del aprendizaje, los autores de este trabajo proponen el análisis de la confiabilidad entre los indicadores a tratar en la base por el colectivo de profesores, teniendo en cuenta que la evaluación de la validez en su integridad necesita de recursos externos.

Según McMillan⁸ (2004) dice que la confiabilidad es alta o buena si el coeficiente de correlación tiene un valor entre .75 o .85. Los resultados se consideran poco confiables cuando el valor del coeficiente es menor de .60

En la asignatura Metodología de la Investigación y Estadística se integran las dos temáticas que aparecen en su nombre y que se convierten en poderosas herramientas para el estudiante durante toda su formación así como al convertirse en egresado para dar solución a diferentes problemas propios de la profesión. Esta asignatura prepara al estudiante para actuar como miembro activo en un equipo de investigación, interpretar los artículos científicos del campo en el que se desempeña, conocer cuándo necesita ayuda de un especialista, actuar según las recomendaciones recibidas por este, interpretar los resultados producto del procesamiento estadístico básico. Los contenidos de esta asignatura constituyen el sustento metodológico de la estrategia curricular de investigación y las habilidades desarrolladas en esta se refuerzan mediante la ejecución de la estrategia, según consta en la fundamentación del programa.

El presente estudio propone identificar elementos de calidad en los exámenes finales de la asignatura con el propósito de que se confeccionen instrumentos con una calidad cada vez mayor y que redunde en mejoría en el aprovechamiento estudiantil, cumpliendo con el principio ético de beneficencia, utilizando indicadores de fácil manejo, que se encuentran en programas accesibles y que no necesitan de criterios externos al de los colectivos de asignatura o disciplina.

Métodos.

La investigación es de tipo descriptivo y se desarrollará en el campo de la evaluación del aprendizaje, teniendo como objeto de estudio el examen final de la asignatura Metodología de la Investigación y Estadística correspondiente al curso 2010-2011 realizado y aplicado a los estudiantes del segundo año de la carrera de Estomatología de la Facultad de Estomatología de la Habana.

El universo para el desarrollo de la investigación lo constituyó la totalidad de los exámenes realizados en el examen ordinario, donde no se incluyen los alumnos eximidos de examen, ni los que no tuvieron derecho al mismo, para un total de 201.

El índice de dificultad del examen, se conceptualiza como la proporción de personas que responden correctamente una pregunta de la prueba. Mientras mayor sea esta proporción menor será el grado de dificultad. Para su cálculo se dividió el número de residentes que contestaron correctamente el ítem, entre el número total de examinados. Usualmente a esta proporción se le denota con

la letra p (dificultad del ítem) según señala⁹ (Backhoff E) y utilizan Argudín¹⁰ Abreu¹¹ y Moreno¹² en sus trabajos.
Se calculó con la siguiente fórmula:

$$p_i = \frac{A_i}{N_i}$$

Donde: p_i = Índice de dificultad de la pregunta.
 A_i = Numero de aciertos en la pregunta.
 N_i = Numero de aciertos más número de errores en la pregunta.

Para ello se asumen los siguientes parámetros:

- Altamente difíciles ($p < 0.32$)
- Medianamente difíciles (de 0.33 a 0.52)
- Dificultad media (de 0.53 a 0.73)
- Medianamente fáciles (de 0.74 a 0.86)
- Altamente fáciles ($p > 0.86$).

Los resultados de este indicador pueden oscilar en:

- 5 % para preguntas fáciles.
- 20 % para preguntas medianamente fáciles.
- 50 % para preguntas de dificultad media.
- 20 % para preguntas medianamente difíciles.
- 5 % difíciles.

El índice de discriminación. Aunque hay varias maneras equivalentes de calcular el índice de discriminación, en este trabajo se utilizó la siguiente fórmula:

$$D_i = \frac{GA \text{ aciertos} - GB \text{ aciertos}}{N \text{ grupo mayor}}$$

Donde:

- D_i = Índice de discriminación del reactivo i
- GA aciertos** = Número de aciertos en el reactivo i del 27% de personas con las puntuaciones mas altas en el test.
- GB aciertos** = Número de aciertos en el reactivo 27% de personas con las puntuaciones mas bajas en el test.
- N grupo mayor** = Número de personas en el grupo mas numeroso (GA o GB). Entre mas alto es el índice de discriminación, el reactivo diferenciara mejor a las personas con altas y bajas calificaciones. Si todas las personas del GA contestan correctamente un reactivo y todas las personas del GB 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).

D =	Calidad	Recomendaciones
> 0.39	Excelente	Conservar
0.30 - 0.39	Buena	Posibilidades de mejorar
0.20 - 0.29	Regular	Necesidad de revisar
0.00 - 0.20	Pobre	Descartar o revisar a profundidad
< -0.01	Pésima	Descartar definitivamente

Y para la confiabilidad se utilizó el coeficiente α de Cronbach:

$$\hat{\alpha} = \frac{k}{k-1} \left[1 - \frac{\sum_{i=1}^k s_i^2}{s_x^2} \right]$$

donde k = cantidad de preguntas del instrumento evaluativo

s_x^2 = varianza de las calificaciones del instrumento evaluativo

s_i^2 = varianza de la i -ésima pregunta del instrumento evaluativo

Se utilizó una computadora Pentium IV con ambiente Windows XP 2003. Los textos se procesaron con Word XP 2003, las tablas de recolección se realizaron con Excel XP 2003 y a través del sistema estadístico SPSS en su versión 13 se procesaron los datos, se analizaron los resultados y se presentaron en tablas.

En los requerimientos éticos de la investigación que se presenta se consideró la voluntariedad de participación de los implicados a los cuales les fue dicho que las informaciones obtenidas solo serían utilizadas con fines científicos y que la identidad se mantendría en el anonimato; se obtuvo el consentimiento tanto de los docentes que participaron en la planificación del examen incluyendo la clave de calificación como el de los que participaron en la calificación.

Resultados y Discusión..

El grado de dificultad en todas las preguntas se mantiene por encima de 0.80 por lo que todas las preguntas clasifican en las categorías de fáciles, las preguntas 1 y 3 como medianamente fáciles y las preguntas 2, 4 y 5 como altamente fáciles para el curso 2010-2011 lo cual favorece el resultado de los estudiantes, sin embargo para el curso 2011-2012 a pesar de que el formato, objetivo y contenido de las preguntas se mantiene, solo se cambian los datos, se observa que aparecen dos preguntas con dificultad media, la 3 y la 4, dos medianamente fáciles y solo una clasifica como muy fácil, de forma general ambos exámenes clasifican como fáciles, similares a los resultados obtenidos por Moreno,¹² con un indicador muy por encima de los que se considera por los estándares internacionales aún cuando los exámenes son diferentes en cuanto a su alcance.

El índice de discriminación se presenta en el primer curso que se analiza todas las preguntas con valores positivos y de ellos las preguntas 2 y 4 clasifican con una calidad de discriminación regular para lo que se recomienda necesidad de revisar, hay que destacar que para ese intervalo la pregunta 2 ocupa el menor valor mientras que la 4 ocupa el valor mayor; la pregunta 5 clasifica como buena y dentro de las recomendaciones se dice que tiene posibilidades de mejorar, las preguntas 1 y 3 discriminan de forma excelente y se recomienda conservar, esto no sucede igual para el curso 2011-2012 donde aparecen una pregunta que presenta pobre discriminación, una regular, dos preguntas con buena discriminación y una como excelente. Aquí se puede ver que el comportamiento es bastante parecido salvo que en las preguntas 2 y 5 en el último curso las preguntas discriminan en una escala inferior mientras que en la pregunta 3 se invierte el proceso; resultados similares para un examen de alcance similar obtuvo Carrazana Lee para Morfofisiología.¹³

En la tabla 3 se expresan los resultados del coeficiente α de Cronbach, para los dos cursos. El valor total 0.86 muestra que el primer instrumento tiene una confiabilidad en un rango acorde al que se propone como aceptable para este tipo de examen y por encima de los encontrados por González Zamora¹⁴ y por Backhoff⁹; este resultado indica la constancia de los resultados en el tiempo, según los tipos de preguntas empleadas y de acuerdo con los criterios de los evaluadores, sin embargo esto no sucede en el segundo donde el valor del coeficiente está por debajo de lo que se considera aceptable y muy por debajo del primero, aún cuando los instrumentos tienen mucha similitud, por lo que pudiera estar influyendo algún componente que provoque esta variación.

Conclusiones.

1. Los exámenes finales de la asignatura Metodología de la Investigación y Estadística clasifican como fácil.
2. Los exámenes finales de la asignatura Metodología de la Investigación y Estadística clasifican como exámenes de buena discriminación.

3. Los exámenes finales de la asignatura Metodología de la Investigación y Estadística se presentan con marcada diferencia en la confiabilidad.

Bibliografía

1. Luna Serrano, E. Evaluación de la docencia: paradojas de un proceso institucional. Revista de la Educación Superior.Vol. XXXII(3), No. 127, Julio-Septiembre de 2003. ISSN: 0185-2760
2. Salas Perea RS. La evaluación en la educación superior contemporánea. Biblioteca de Medicina Volumen XXIV, 1998.Educación en salud. Competencia y desempeño profesionales. La Habana: Editorial Ciencias Médicas; 1999. En CD-ROM Maestría Educación Médica. ENSAP; 2005.
3. Córlica José Luis y otros. Propuesta de indicadores de calidad para la evaluación de la calidad de las preguntas en los exámenes de opción múltiple. Primer Congreso Virtual Latinoamericano de Educación a Distancia. LatinEduca 2004.com
4. Ortiz Ramos Evelyn F. Modelo de evaluación del aprendizaje. Revista Paideia Puertorriqueña Universidad de Puerto Rico. Recinto de Río Piedras. Vol.1, Num.2 Año 2006. <http://paideia.uprrp.edu>.
5. Gronlund, N. (2006). *Assessment of student achievement* (8^{va}. ed.). Boston, MA, EE.UU.: Pearson.
6. Gronlund, N. (2006). *Assessment of student achievement* (8^{va}. ed.). Boston, MA, EE.UU.: Pearson.
7. Klecker, B. (2005, marzo). *Assessing learning online: The top ten list*. Paper presented at the annual meeting of the society for information technology and teacher education (SITE), Phoenix, AZ, EE.UU.
8. McMillan, J. (2004). *Educational research: Fundamentals for the consumer* (4^{ta} ed.). Boston, MA: Allyn and Bacon.
9. Backhoff E, Larrazolo N y Rosas N. Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) Rev. Electrónica de Inv Educ. 2000; 2(1):11-28
10. Argudín Somonte Eiglys, Díaz Rojas Pedro, Leyva Sánchez Elizabeth. Índice de Dificultad del examen de Morfofisiología Humana I. Educ Med Super [revista en la Internet]. 2011 Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000200007&lng=es.
11. Abreu Reyes David Rafael. Caracterización del examen teórico final de Morfofisiología Humana del perfil Laboratorio Clínico. Educ Med Super [revista en la Internet]. 2011 Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000100003&lng=es.
12. Moreno Montañez Magalys, Quintana Regalado Grecia. Caracterización del examen estatal escrito en la especialidad de Medicina General Integral: characterization. Educ Med Super [revista en la Internet]. 2009 Sep [citado 2012 Abr 13] ; 23(3): . Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412009000300003&lng=es.

13. Carrazana Lee Armando, Salas Perea Ramón Syr, Ruiz Salvador Ana Karella. Nivel de dificultad y poder de discriminación del examen diagnóstico de la asignatura Morfofisiología Humana I. Educ Med Super [revista en la Internet]. 2011 Mar [citado 2012 Abr 17] ; 25(1): 103-114. Disponible en: <http://scielo.sld.cu/scielo.php?script=sciarttext&pid=S0864-21412011000100010&lng=es>.
14. González Zamora. Tipos de preguntas cerradas abiertas: ventajas y desventajas. Eduteka [en línea] 2004 febrero 17. <http://www2.ual.es/JMC/Webens/288.html>.

Anexos

Tabla 1. Grado de dificultad de los reactivos de los exámenes finales de Metodología de la Investigación y Estadística, 2010-2011 y 2011-2012.

Grado de dificultad	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5
Curso 2010-2011	0,83	0,94	0,82	0,88	0,90
Curso 2011-2012	0,86	0,93	0,68	0,68	0,84

Tabla 2. Índice de Discriminación de los reactivos de los exámenes finales de Metodología de la Investigación y Estadística, 2010-2011 y 2011-2012.

Índice de Discriminación	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5
Curso 2010-2011	0,47	0,2	0,42	0,29	0,36

Curso 2011-2012	0,53	0,15	0,32	0,32	0,2
-----------------	------	------	------	------	-----

Tabla 3. Confiabilidad el examen de los reactivos de los exámenes finales de Metodología de la Investigación y Estadística, 2010-2011 y 2011-2012.

	No. examinados	α Cronbach
Curso 2010-2011	201	0,86
Curso 2011-2012	173	0,50