





Calidad del instrumento de evaluación final de la asignatura Célula, tejidos y sistema tegumentario (Original)

Quality of the final evaluation instrument of the subject cell tissues and integumentary system (Original)

Marisela Resco Zequeira. Doctora en Medicina. Especialista de Primer Grado en Histología. Especialista de Primer Grado en MGI. Profesora Asistente. Universidad de Ciencias Médicas de Granma. Granma. Cuba. mresco@infomed.sld.cu 

Rafael Gutiérrez Núñez. Doctor en Medicina. Máster en Educación Médica Superior. Especialista de Primer Grado en Embriología. Profesor Auxiliar. Universidad de Ciencias Médicas de Granma. Granma. Cuba. rafagutierrez@infomed.sld.cu 

Ibis Ofelia Sánchez Rodríguez. Doctora en Medicina. Especialista de Primer Grado en Histología. Especialista de Primer Grado en MGI. Profesora Asistente. Universidad de Ciencias Médicas de Granma. Granma. Cuba. ibisa@infomed.sld.cu 

Yeumicelys González Verdecia. Licenciada. Especialista de Primer Grado en Histología. Profesor Asistente. Universidad de Ciencias Médicas de Granma. Granma. Cuba. yeumicelysgv@infomed.sld.cu 

Recibido: 22-01-2021/ Aceptado: 09-06-2022

Resumen

La evaluación del aprendizaje es un componente esencial del proceso de enseñanza-aprendizaje como vía de retroalimentación para su propia dirección y perfeccionamiento. Se realizó un estudio descriptivo en el campo de la evaluación a 407 exámenes que corresponden al total de estudiantes presentados al Examen Ordinario de la Asignatura Célula, Tejidos y Sistema Tegumentario para evaluar la calidad de dicho examen a través del nivel de dificultad y el poder

de discriminación del instrumento. Los resultados obtenidos mostraron una adecuada calidad a partir de la coherencia entre las preguntas, el nivel de dificultad y el poder de discriminación. Esta investigación permitió al claustro docente profundizar en el diseño de los instrumentos de evaluación y de esta forma realizar acciones que garanticen la calidad en la formación de los estudiantes para el perfeccionamiento del proceso enseñanza-aprendizaje, principalmente el sistema evaluativo.

Palabras clave: Evaluación del Aprendizaje; Índice de Dificultad; Poder de Discriminación; Índice de Discriminación; Coeficiente de Discriminación

Abstract

The learning assessment is an essential component of the teaching-learning process as a means of feedback for its own direction and improvement. A descriptive study was carried out in the field of evaluation of 407 exams that correspond to the total number of students presented to the Ordinary Examination of the Cell Tissue and Integumentary System Subject to evaluate the quality of said exam through the level of difficulty and the power of discrimination of the instrument. The results obtained showed adequate quality from the coherence between the questions, the level of difficulty and the power of discrimination. This research allowed the teaching staff to deepen the design of the evaluation instruments and thus carry out actions that guarantee the quality of the students' training for the improvement of the teaching-learning process, mainly the evaluation system.

Keywords: Learning Assessment; Difficulty Index; Discrimination Power; Discrimination Index; Discrimination coefficient

Introducción

La enseñanza constituye el proceso de organización y dirección de la actividad cognoscitiva en el cual interactúan el profesor y el estudiante y por ello se denomina proceso de enseñanza-aprendizaje. El mismo constituye un sistema en el que cada uno de sus componentes se interrelacionan, siendo la evaluación el mecanismo regulador del sistema y puede considerarse como el instrumento de control de la calidad del producto resultante del proceso. En este sentido la evaluación del aprendizaje es un componente esencial del proceso de enseñanza-aprendizaje como vía de retroalimentación para su propia dirección y perfeccionamiento (Carrazana et al., 2011; Chaviano et al., 2016; Salas & Salas, 2017).

Estos autores consideran que es imprescindible que a lo largo de toda la carrera se compruebe en qué medida están preparados los estudiantes para enfrentarse exitosamente a su responsabilidad social y si su aprendizaje se corresponde con los objetivos de cada asignatura, disciplina, año de estudio y la carrera en general.

De acuerdo a Rivera et al. (2017) cuando se utilizan instrumentos de alto impacto, como es el caso de los exámenes finales de las asignaturas, es necesario conocer los indicadores técnicos que definen la calidad del instrumento evaluativo que se emplea. Por su dimensión y por el poderoso impacto social que tienen, la elaboración de exámenes debe ajustarse a rigurosos estándares de calidad, pues los defectos en la elaboración de los mismos tienen efectos negativos en el currículo, el proceso docente y los estudiantes.

Reviste gran importancia que la calidad de los exámenes aplicados sea verificada, pues sirven como instrumento de medición educativa, para inferir capacidades de las personas, ofrecer información correcta para la adopción de decisiones y permitir evaluar la marcha de procesos docentes (Carrazana et al., 2011; Sánchez et al., 2015).

Diversos estudios realizados tanto en Cuba como en otros países reportan defectos en la elaboración de instrumentos evaluativos escritos. Rodríguez et al. (2018) consideran que esto repercute de manera perjudicial en las características psicométricas y por tanto en la validez de dichos instrumentos.

Es necesario que el sistema de evaluación sea lo más sólido posible, se impone que los instrumentos que se empleen tengan la calidad requerida y se pongan a prueba. Entre los indicadores de calidad de exámenes escritos que con mayor frecuencia se trabajan están la validez, confiabilidad, dificultad y discriminación, los dos últimos fueron los estudiados en esta investigación. Estos indicadores aportan información útil a estudiantes y profesores sobre el grado de dificultad que presenta un examen o contenido evaluado para trabajar en la superación de los problemas encontrados, además permite diferenciar los estudiantes que dominan un conocimiento o habilidad de los que no lo han logrado aún, valorar el nivel de autopreparación alcanzado y orientar mejor el trabajo diferenciado; referencias necesarias para la planificación, organización y ejecución de las actividades para mejorar lo relacionado con la evaluación y con ello el proceso enseñanza-aprendizaje (Carrazana et al., 2011).

La implementación del Plan E en la carrera de Medicina comienza a partir del curso 2019-2020 e incluye en su currículo como una de sus disciplinas a Bases Biológicas de la Medicina, la misma es integradora y sigue la tendencia actual de las ciencias básicas biomédicas en la unificación interdisciplinaria para tratar los problemas comunes o afines de varias ramas científicas. La asignatura Célula, Tejidos y Sistema Tegumentario se imparte en el primer semestre de primer año e incluye los contenidos referentes a la célula eucariota, tejidos básicos y sistema tegumentario.

De modo que se toma como problema científico el desconocimiento de la calidad (dificultad y discriminación) del examen ordinario de la asignatura Célula, Tejidos y Sistema Tegumentario aplicado al primer año de la carrera de Medicina en la Facultad de Ciencias Médicas de Granma.

Entre los objetivos concebidos, se persigue a modo general evaluar la calidad del Examen Ordinario de la Asignatura Célula, Tejidos y Sistema Tegumentario a través del nivel de dificultad y el poder de discriminación del instrumento. Específicamente se busca describir la estructura del instrumento evaluativo utilizado teniendo en cuenta el número de preguntas, procedimientos y técnicas empleados en la elaboración de cada pregunta, y determinar la frecuencia de errores, el nivel de dificultad y el poder de discriminación por preguntas y del instrumento evaluativo en general.

Población y muestra

Se realizó un estudio descriptivo en el campo de la evaluación al Examen Ordinario de la Asignatura Célula, Tejidos y Sistema Tegumentario, aplicado al primer año de la Carrera de Medicina en la Facultad de Ciencias Médicas de Granma en el curso 2019- 2020. El universo estuvo integrado por 407 exámenes que corresponden al total de estudiantes presentados al examen de la asignatura.

Se clasificaron las preguntas teniendo en cuenta procedimientos y técnicas empleados para su elaboración y se determinó la frecuencia de errores de cada una. Los métodos estadísticos se utilizaron para la identificación del nivel de dificultad y del poder de discriminación de cada pregunta y del instrumento como un todo.

El índice de dificultad y de discriminación se calcularon a través del software de hoja de cálculo Excel de Windows XP, junto con una calculadora científica. Se trabajó con números

enteros, porciento e índices para la presentación de los resultados en tablas. Los cálculos se realizaron por la fórmula de Backhoff (Backhoff et al., 2000).

Nivel de dificultad: Proporción de personas que responden correctamente dentro del total de examinados. Es inversamente proporcional a la dificultad. Bajo una perspectiva estadística se expresa como índice de dificultad (p) y como cualquier índice, debe estar cercano a una proporción de 0.5-0.6. Se distribuyeron los valores de p en: altamente difícil menos de 0,32; medianamente difícil de 0,32-0,52; dificultad media de 0,53-0,73; medianamente fácil de 0,74-0,86 y altamente fácil más de 0,86.

Poder de discriminación: Permite distinguir a los estudiantes de altos y bajos rendimientos, además aporta criterio de predicción sobre los resultados del estudiante, para identificarlo se utilizan dos indicadores: índice de discriminación y coeficiente de discriminación.

Índice de discriminación: permite distinguir los estudiantes de altos y bajos rendimientos, solo se valora 54% (27% más alto y 27% más bajo) de los estudiantes evaluados, para su cálculo se organizaron los resultados de los exámenes en orden decreciente, se calculó 27% de los resultados más altos y de los más bajos, se calculó el número que representa 46% restante.

El Di se distribuyó de la siguiente forma: mayor de 0,39 calidad excelente (conservar), de 0,30 – 0,39 buena calidad (posibilidades de mejorar), de 0,20 – 0,29 calidad regular (necesidad de revisar), de 0,00 – 0,20 pobre calidad (descartar o revisar a profundidad) y < -0.01 pésima calidad (descartar definitivamente).

Coeficiente de discriminación: Tiene en cuenta a todos los estudiantes evaluados, permite predecir sobre los resultados del estudiante. Según estándares internacionales la distribución de los resultados del coeficiente discriminación punto de correlación biserial (rpbis) es: menor de 0

discriminan negativamente, de 0 - 0,14 discriminan pobremente, de 0,15- 0,25 discriminan regular, de 0,26 -0,35 buen poder discriminativo, mayor de 0,35 excelente poder de discriminación. Los resultados del coeficiente de discriminación se distribuyeron en adecuado (coeficiente positivo) e inadecuado (coeficiente negativo).

Análisis de los resultados

Tabla 1. Procedimientos y Técnicas utilizados en el Examen Ordinario de Células, Tejidos y Sistema Tegumentario

Preguntas	Procedimiento	Técnica
1	Test Objetivo	Verdadero y Falso
2	Test Objetivo	Enlace o Apareamiento
3	Test Objetivo	Selección Múltiple
4	Test Objetivo	Complemento Simple
5	Ensayo	Respuesta Larga
6	Test Objetivo	Enlace o Apareamiento
7	Ensayo	Respuesta Corta

En el presente estudio el examen ordinario estuvo compuesto por siete preguntas según se muestra en la Tabla 1. Predominaron las preguntas de tipo Test Objetivo o Respuesta Estructurada con un total de 5 que representa el 71,42% y dentro de estas las de enlace o apareamiento. En estas preguntas el educando elige su respuesta dentro de varias, no se puede negar que son fáciles y rápidas en su calificación, permiten evaluar un número mayor de estudiantes y un número mayor de contenido, pero deben ser elaboradas con rigurosidad.

En las revisiones realizadas, autores como Carrazana et al. (2011) analizaron instrumentos evaluativos con formato mixto, resultados que evidencian la tendencia de elaboración de exámenes en este formato en ciencias médicas. Zayas (2002) refiere que las preguntas más empleadas son las de test objetivo, pero contrariamente las preguntas de mayor preferencia por los estudiantes son las de ensayo. Las autoras consideran que lo correcto sería utilizar exámenes de formato mixto, utilizando ambos tipos de preguntas y de esta forma aprovechar las ventajas de cada una y minimizar sus desventajas.

Tabla 2. Frecuencia de errores calculada al Examen Ordinario de Célula, Tejido y Sistema Tegumentario

Pregunta	Frecuencia de Errores	%	Aprobados	%
1	38	9,33%	369	90,67%
2	22	5,40%	385	94,60%
3	64	15,72%	343	84,28%
4	70	17,19%	337	82,81%
5	152	37,34%	255	62,66%
6	112	27,51%	295	72,49%
7	31	7,61%	376	92,39%
Examen	67	16,46%	340	83,53%

Al determinar la frecuencia de errores apreciable en la Tabla 2 se observó que la pregunta 2 tuvo mayor número de aciertos con un 94,60%, o sea, fue respondida correctamente por más examinados mientras que la pregunta 5 tuvo el mayor número de errores o menor número de aciertos con el 37,34%. Para el grupo de examinados la pregunta 5 es más difícil que la pregunta 2. Coincidiendo esto con el índice de dificultad calculado a estas preguntas. Resultaron aprobados en el examen 340 estudiantes que representa el 83,53% y desaprobados 67 estudiantes para un 16,46%.

Tabla 3. Dificultad calculada al examen de Célula, Tejido y Sistema Tegumentario

Pregunta	p	Dificultad
1	0,76	Medianamente Fácil
2	0,93	Altamente Fácil
3	0,54	Dificultad Media
4	0,53	Dificultad Media
5	0,30	Altamente Difícil
6	0,63	Dificultad Media
7	0,88	Altamente Fácil
Examen	0,69	Dificultad Media

Leyenda P: Índice de Dificultad

En todas las preguntas se determinó el índice de dificultad, mostrado en la Tabla 3, acorde con la curva de distribución de frecuencias según las respuestas correctas dadas por los estudiantes, un 28,57% considera fácil el examen, un 14,28% medianamente fácil, el 42,85% lo considera de dificultad media, y difícil lo considera el 14,28%. El examen de forma general

calificó de dificultad media con un índice de 0,69 lo que indicó que más del 80% de los estudiantes lograron responder el examen satisfactoriamente. Estos resultados difieren con lo que se desea obtener para un examen, pues según el criterio de varios autores, como promedio deben oscilar en un rango de 0,5 y 0,6, aunque otros extienden el límite superior de este rango hasta 0,73 considerando a estos de dificultad media a medianamente fácil, criterio asumido por las autoras (Carrazana et al., 2011; Backhoff et al., 2000; Argudín et al., 2011).

Lo deseado para un examen según Backhoff et al. (2000) es que un 5% considere fácil el examen, 20% medianamente fácil, 50% con dificultad media, 20% medianamente difícil y un 5% difícil. A consideración de las autoras no se realizó con la profundidad requerida la planificación del examen y esto pudo ser la causa de la diferencia de los valores porcentuales esperados y observados de los incisos fáciles y difíciles.

Es consenso para diferentes autores que han publicado sobre el tema que un examen que posea dificultad media, es superior, en confiabilidad de resultados, a aquellos que tengan mucha dificultad o que resulten muy fáciles, debido a que aumenta el criterio para diferenciar los estudiantes que dominan un contenido de los que no lo hacen aún. Es decir, exámenes muy fáciles o muy difíciles dan poco margen de discriminación, debido a que los estudiantes son igualados, en un examen muy fácil la mayoría tiene altos resultados, en un examen muy difícil la mayoría tiene muy bajos resultados, estas preguntas no contribuyen a la fiabilidad (Morales, 2009; Blanco et al., 2015).

La Tabla 4 muestra el índice de discriminación el cual osciló entre 0,41 y 0,89, se aprecia que 5 de las preguntas que representa el 71,42% presentaron una discriminación excelente y un mínimo tuvo pobre discriminación. Estas cifras muestran que el instrumento presenta buena eficacia para diferenciar los estudiantes que tienen buenos resultados de aquellos con resultados

bajos, o sea, define sin dificultad aquellos estudiantes que tienen el conocimiento correcto de aquellos que no lo tienen. Otros estudios afines mostraron resultados diferentes, entre ellos Ortiz et al. (2015) y Martínez et al. (2019).

Tabla 4. Poder de discriminación calculado a cada pregunta y al examen general

Pregunta	Poder de Discriminación			
	Di	Discriminación	rpbis	Discriminación
1	0,51	Excelente	0,69	Excelente
2	0,86	Excelente	0,60	Excelente
3	0,10	Pobre	0,66	Excelente
4	0,10	Pobre	0,66	Excelente
5	0,41	Excelente	0,64	Excelente
6	0,86	Excelente	0,63	Excelente
7	0,89	Excelente	0,63	Excelente
Examen	0,69	Excelente	0,62	Excelente

Leyenda: Di: Índice de Discriminación rpbis: Coeficiente de Discriminación

La excelente discriminación general del instrumento con un índice de 0,69 concuerda con la frecuencia de errores del examen, lo cual se corresponde con lo esperado en exámenes con una dificultad media. Blanco et al. (2015) encontraron resultados similares en su investigación.

El coeficiente de discriminación del punto biserial mostró que las 7 preguntas y el examen como un todo discriminaron de forma excelente, con valores entre 0,62 a 0,69. Este siempre fue positivo por lo que se catalogó como adecuado.

Las dos preguntas con índice discriminativo pobre (28, 27%) coincidieron en ser clasificadas de dificultad media según su índice de dificultad, fueron elaborados de tipo test objetivo, con poca profundidad y sus resultados no permitieron predecir que los estudiantes que obtuvieron altas puntuaciones en el examen deben haber obtenido altas calificaciones en las preguntas.

Es criterio de las autoras que en las preguntas de dificultad media que no discriminaron adecuadamente, debe valorarse la necesidad de descartar o revisar con profundidad para lograr

reelaborarlas con mayor complejidad, y puedan predecir que el estudiante que salió bien en el examen obtuvo buenos resultados en las preguntas.

El coeficiente de discriminación, el cual tiene en cuenta la totalidad de los estudiantes, mostró cifras positivas en todas las preguntas y en el examen de forma general, lo cual revela que todas las preguntas fueron adecuadas, estos resultados coinciden con las cifras promedio que se buscan en un examen cuando se quiere valorar la capacidad de predicción de un instrumento.

Se puede inferir que las preguntas miden el mismo conocimiento o habilidad que está midiendo el examen total, es decir, las preguntas permitieron predecir que el estudiante que salió bien en el examen debe haber salido bien en cada una de las preguntas, lo cual coincide con diferentes estudios publicados, aunque se diferencia de los resultados de Backhoff et al. (2000) que reportó al menos una pregunta que no discrimina ni permite predecir.

Coincidiendo con otros autores se debe tener presente que cuando se revise un examen hay que verificar si las preguntas responden a los objetivos generales a evaluar, si otras respuestas predominantes serían igualmente correctas, si existen problemas de redacción que hagan que el estudiante interprete diferente a la respuesta correcta y si constan preguntas con respuestas muy elementales o muy difíciles que disminuyan el poder de discriminación que se persigue (Carrazana et al., 2011; Backhoff et al., 2000).

Conclusiones

- 1- El examen estuvo compuesto por siete preguntas predominando las preguntas de tipo test objetivo o respuesta estructurada y la técnica más empleada fue la de enlace o apareamiento.
- 2- La calidad del examen fue adecuada a partir del análisis de la coherencia entre las preguntas, el nivel de dificultad y el poder de discriminación.

Referencias bibliográficas

- Argudín, E., Díaz, P., & Leyva, E. (2011). Índice de Dificultad del examen de Morfofisiología Humana I. *Educación Médica Superior*, 25(2), 97-106.
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000200007&lng=es&nrm=iso
- Backhoff, E., Larrazolo, N., & Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2(1). <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>
- Blanco, M. E, Martínez, L., González, A., & Jordán, M. (2015). Calidad del examen final teórico de Morfofisiología Humana I en la Facultad de Ciencias Médicas de Matanzas. Cursos 2012-2013 y 2013-2014. *Revista Médica Electrónica*, 37(4).
<http://www.revmedicaelectronica.sld.cu/index.php/rme/article/view/1173>
- Carrazana Lee, A, Álvarez Bustamante, G, Quesada Rodríguez, M. y Hidalgo Cerito, Y. (2018). Dificultad y discriminación de exámenes ordinarios de la asignatura Célula, Tejidos y Sistema Tegumentario en Ciencias Básicas. *Revista habana cienciasmédicas*, 17(2) ,278-289.
- Carrazana, A, Salas, R. S., & Ruiz, A. K. (2011). Nivel de dificultad y poder de discriminación del examen diagnóstico de la asignatura Morfofisiología Humana I. *Educación Médica Superior*, 25(1), 103-114. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000100010&lng=es&tlng=es
- Chaviano, O., Baldomir, T., Coca, O., & Gutiérrez, A. (2016). La evaluación del aprendizaje: nuevas tendencias y retos para el profesor. *Edumecentro*, 8(4), 191-205.
<http://www.revedumecentro.sld.cu/index.php/edumc/article/view/855>

- Martínez, R., Ríos, M., Roque, Y., & Caballero, K. Y. (2019). Evidencias de validez del examen final de la asignatura Biología molecular, carrera de Medicina. *Medimay*, 26(3), 338-48. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=92640>
- Morales, P. (2009). *Análisis de ítems en las pruebas objetivas*. Universidad Pontificia Comillas.
- Ortiz, G. M., Díaz, P. D., Llanos, O. R., Pérez, S. M., & González, K. (2015). Dificultad y discriminación de los ítems del examen de Metodología de la Investigación y Estadística. *Edumecentro*, 7(2). <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=56348>
- Rivera, J., Flores, F., Alpuche, A., & Martínez, A. (2017). Evaluación de reactivos de opción múltiple en medicina. Evidencia de validez de un instrumento. *Investigación en Educación Médica*, 6(21), 8-15. <https://doi.org/10.1016/j.riem.2016.04.005>
- Rodríguez, T., Sondón, M. E., Peña, M., Hernández, M., & Robles, R. (2018). La calidad del examen final de la asignatura Morfofisiología IV. *Correo Científico Médico de Holguín*, 22(3), 446-462. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=84221>
- Sánchez, E., Medina, M., Rodríguez, M., Vega, L., & de la Torre, G. (2015). Indicadores de calidad para un examen teórico de la especialidad de medicina general integral. *Medisan*, 19 (2), 150-157. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=56112>
- Salas, R. S., & Salas, A. (2017). *Modelo formativo del médico cubano. Bases teóricas y metodológicas*. Editorial Ciencias Médicas.

Zayas, M. (2002). *Estudio descriptivo de los instrumentos evaluativos en la asignatura Farmacología* [Trabajo de terminación de Diplomado en Educación Médica Superior] Instituto Superior de Ciencias Médicas “Serafín Ruíz de Zarate Ruiz.