



I Congreso INTERNACIONAL



Sociedad Cubana de Ciencias Estomatológicas. Capítulo Santiago de Cuba

La ciencia en función de una eterna sonrisa

Calidad del examen final de la asignatura Sangre y Sistema Nervioso en la carrera Estomatología

Quality of final test from Blood and Nervous System subject in Dentistry's career

Oscar Rodríguez Reyes¹, Martha Fajardo Puig², Margarita Hernández Cunill³.

¹ Estomatólogo, Doctor en Ciencias Médicas, Máster en Atención de Urgencias Estomatológicas, Especialista de Primer y Segundo Grado en Fisiología Normal y Patológica, Profesor Titular e Investigador Auxiliar, Universidad de Ciencias Médicas de Santiago de Cuba, Facultad de Estomatología, Santiago de Cuba, Cuba, E-mail: orreyes7807@infomed.sld.cu, ORCID: <https://orcid.org/0000-0002-3042-9153>

² Estomatóloga, Especialista de Primer y Segundo Grado en Fisiología Normal y Patológica, Profesor Auxiliar, Universidad de Ciencias Médicas de Santiago de Cuba, Facultad de Estomatología, Santiago de Cuba, Cuba, E-mail: martha.fajardo@infomed.sld.cu, ORCID: <https://orcid.org/0000-0003-2437-6996>

³ Licenciada, Máster en Medicina Natural y Tradicional, Especialista de Primer Grado en Histología, Profesora Auxiliar, Universidad de Ciencias Médicas de Santiago de Cuba, Facultad de Estomatología, Santiago de Cuba, Cuba, E-mail: mhernandezcunill2022@gmail.com; ORCID: <https://orcid.org/0000-0002-4959-946X>

Autor para la correspondencia: orreyes7807@infomed.sld.cu

RESUMEN

Introducción: estimar la calidad de cualquier instrumento evaluativo escrito, es requisito esencial no solo en su proceso de construcción, sino también en la obtención de información para la toma correcta de decisiones.

Objetivo: evaluar la calidad del examen final de la asignatura Sangre y Sistema Nervioso en la disciplina Bases Biológicas de la Estomatología

Métodos: estudio observacional analítico y transversal, en 22 exámenes correspondientes al examen final de la asignatura Sangre y Sistema Nervioso, disciplina Bases Biológicas de la Estomatología, carrera de Estomatología, Universidad de Ciencias Médicas de Santiago de Cuba, curso 2022. Se utilizaron como métodos teóricos: análisis y síntesis e inducción-deducción, y empíricos: análisis documental y entrevistas a expertos. Se analizó la correspondencia del fondo de tiempo y la cantidad de ítems explorados en el examen, se calcularon índices de dificultad y de discriminación de incisos y preguntas, así como de consistencia interna.

Resultados: existió correspondencia adecuada entre el fondo de tiempo asignado a los contenidos de las temáticas y la cantidad de ítems explorados en el examen. Predominaron en un 60 % los reactivos con índice de dificultad media y buen poder de discriminación. La consistencia interna de las preguntas y del examen general se evaluó de excelente.

Conclusiones: la calidad del examen final aplicado en la asignatura Sangre y Sistema Nervioso de la disciplina Bases Biológicas de la Estomatología, carrera de Estomatología, Universidad de Ciencias Médicas de Santiago de Cuba, curso 2022, fue adecuada.

Palabras clave: dificultad; discriminación; confiabilidad; reactivos

ABSTRACT

Introduction: to estimate the quality of any written evaluative instrument is an essential requirement not only in its construction process, but also in obtaining information for correct decision-making.

Objective: to evaluate the quality of final test from Blood and Nervous System subject in Biological Bases of Dentistry discipline.

Methods: analytical and cross-sectional observational study, in 22 tests corresponding to the final test of Blood and Nervous System subject, Biological Bases of Dentistry discipline, Dentistry career, Santiago de Cuba Medical Sciences University, course 2022. As theoretical methods were used: analysis and synthesis and induction-deduction, and empirical: documentary analysis and interviews with experts. The correspondence from the background of time and the number of items explored in the test were analyzed, indices of difficulty and discrimination of items and questions, and internal consistency were calculated.

Results: there was an adequate correspondence between the amount of time assigned to the contents of the themes and the number of items explored in the test. Items with a medium

difficulty index and good discrimination power prevailed in 60%. The internal consistency of the questions and general test was evaluated as excellent.

Conclusions: the quality of the final test from Blood and Nervous System subject, Biological Bases of Dentistry discipline, Dentistry career, Santiago de Cuba Medical Sciences University, course 2022, was adequate.

Key words: difficulty; discrimination; reliability; reagents

INTRODUCCIÓN

La evaluación constituye el mecanismo regulador del proceso enseñanza-aprendizaje, al posibilitar su retroalimentación, dirección y perfeccionamiento. Como uno de los aspectos más polémicos y de mayor interés en la práctica pedagógica, permite en la educación médica superior cubana, el control y la valoración de los conocimientos, habilidades, hábitos y modos de actuación que los estudiantes van adquiriendo sobre los objetivos de cada disciplina, asignatura y del plan de estudio. Según Blanco Pereira y otros,⁽¹⁾ para que la evaluación sea eficaz debe de estar vinculada con la competencia de los profesores, los métodos y medios empleados, así como de la calidad del currículo.

Al concluir la confección de cualquier instrumento evaluativo escrito, sea frecuente, parcial o final, se presenta el problema de determinar en qué medida la información obtenida a través de su aplicación, refleja el nivel de competencia del examinado. Desde esta perspectiva, estimar la calidad del mismo es requisito esencial no solo en su proceso de construcción, sino también en la obtención de información para la toma correcta de decisiones.⁽²⁾

Múltiples investigadores como Villamarín Guevara,⁽³⁾ Díaz Rojas y Leyva Sánchez⁽⁴⁾, coinciden al afirmar que dentro de los indicadores de calidad de los exámenes escritos que con mayor frecuencia se estudian, están la validez, confiabilidad, discriminación y dificultad. Aseguran que el análisis que se realiza en el colectivo de profesores durante el proceso de planeación y de elaboración de las preguntas e ítems, la forma de definir la clave de calificación y de determinar el éxito y el fracaso; así como, la experiencia del claustro docente en la temática evaluación, es fundamental para el logro de un instrumento válido y confiable.

La carrera de Estomatología en Cuba, incluye en su currículo base del plan de estudio E la disciplina Bases Biológicas de la Estomatología (BBE). La misma es integradora y sigue la tendencia actual de las ciencias biomédicas en la unificación interdisciplinaria, cuyo sustento

se materializa en la identificación de los núcleos básicos de los contenidos, sus esencialidades, su relación con los problemas de salud bucal que los futuros egresados deben ser capaces de enfrentar y/o resolver, así como las tareas que en interacción con la familia, la comunidad y el medio ambiente, deben desarrollar para darles solución y que definen su futura competencia.⁽⁵⁾

Sangre y Sistema Nervioso, es una de las asignaturas de mayor complejidad dentro de las BBE. Se imparte en el primer año (segundo período) del currículo base, e incluye el estudio de los sistemas Hemolinfopoyético y Nervioso, temas de vital importancia para el futuro estomatólogo general que se pretende formar. Su evaluación final abarca los contenidos de ambos sistemas y los resultados docentes de modo habitual, no siempre han sido los esperados por el claustro de profesores de las ciencias básicas biomédicas de la facultad de Estomatología, Universidad de Ciencias Médicas de Santiago de Cuba.

La carencia de una estimación científica sobre la calidad del examen final de la asignatura Sangre y Sistema Nervioso, constituye el problema científico de la presente investigación, la cual tiene como objetivo, evaluar la calidad de dicho instrumento evaluativo durante el curso académico 2022, a través del análisis de los indicadores dificultad, discriminación y confiabilidad.

MÉTODOS

Características generales de la investigación

Se realizó un estudio observacional analítico y transversal, en la facultad de Estomatología, Universidad de Ciencias Médicas de Santiago de Cuba, entre los meses de septiembre a noviembre de 2022. El universo de estudio estuvo constituido por 22 exámenes, correspondientes a la prueba final de la asignatura Sangre y Sistema Nervioso, disciplina Bases Biológicas de la Estomatología, plan de estudio E, carrera Estomatología, aplicada a los estudiantes del primer año del curso académico 2022.

Métodos, técnicas y procedimientos

Se emplearon métodos teóricos de análisis y síntesis e inducción-deducción para los referentes relacionados con la evaluación del aprendizaje, calidad de los instrumentos de evaluación y para la interpretación de los resultados obtenidos. Dentro de los métodos empíricos se utilizó el análisis documental y la entrevista a expertos, lo que facilitó la revisión del programa de la asignatura, documentos normativos de la disciplina, así como caracterizar

el examen según formato, contenidos que evalúa y criterio de respuestas correctas en los ítems de tipo ensayo y test objetivo. Para ello, se seleccionaron cinco profesores de las ciencias básicas biomédicas de la Universidad de Ciencias Médicas de Santiago de Cuba, con más de 15 años de experiencia en la docencia y con categoría docente principal de Auxiliar y Titular.

En una hoja de cálculo de Microsoft Excel, se tabularon las notas de cada pregunta y la general del examen objeto de estudio, se incluyó además los resultados de cada ítem del instrumento categorizados en aprobado (1) y desaprobado (0), lo que garantizó su posterior procesamiento estadístico.

La existencia o no de correspondencia entre la cantidad de horas por temas y el número de ítems explorados en el examen final, se determinó a través del coeficiente de correlación por rangos de Spearman.⁽⁶⁾

Para dar salida al objetivo propuesto, se definieron y operacionalizaron los siguientes índices y coeficientes:

1. Índice de dificultad del ítem (D_i): número de estudiantes examinados que respondió correctamente el ítem dividido entre el total de examinados.
2. Índice de dificultad de la pregunta (D_p): se sumó el total de estudiantes que respondieron correctamente cada uno de los ítems que corresponden a la pregunta, y se dividió entre el número de ítems de la misma. Luego se calculó la media de respuestas correctas para la pregunta y se dividió por el total de alumnos examinados.
3. Índice de dificultad del examen (D_e): se sumó el total de estudiantes que respondieron correctamente cada ítem del examen, se dividió entre el número de ítems, y a continuación entre el número de examinados.

El índice de dificultad para todos los casos se clasificó en las categorías: ^(7, 8)

- Difícil: ($< 0,32$)
- Medianamente difícil: ($0,32 - 0,52$)
- Dificultad media: ($0,53 - 0,73$)
- Medianamente fácil: ($0,74 - 0,86$)
- Fácil: ($> 0,86$)

4. Índice de discriminación (I_d): para poder aplicar este indicador, en una columna de la hoja de cálculo se colocaron los resultados de la pregunta y en la columna siguiente los resultados

del ítem. A continuación se ordenó la columna de la pregunta de la nota mayor a la menor (aceptando la opción “ampliar selección” de la aplicación Excel).

Con esta organización se seleccionó el número de estudiantes que representó el 33 % de los examinados (9) y se procedió a seleccionar el grupo de mejores notas (en sentido descendente a partir de la más alta) y el grupo de notas más bajas (en sentido ascendente a partir de la más baja). Luego para el cálculo del Id se empleó la siguiente fórmula:

$$Id = \frac{GA(\text{aciertos}) - GB(\text{aciertos})}{N (\text{grupo mayor})}$$

Id: índice de discriminación del reactivo

GA: número de aciertos en el reactivo i del 33 % de personas con las puntuaciones más altas en el test.

GB: número de aciertos en el reactivo i del 33 % de personas con las puntuaciones más bajas en el test.

N: número de personas en el grupo más numeroso (GA o GB).⁽⁴⁾

Para la evaluación de la calidad de los ítems y preguntas en términos de índice de discriminación, se empleó la siguiente escala utilizada por *Martínez Pérez y otros*:⁽⁷⁾

- Excelente: mayor o igual de 0,40 (conservar)
- Buena: entre 0,30 – 0,39 (posibilidades de mejorar)
- Regular: entre 0,20 – 0,29 (necesidad de revisar)
- Pobre: entre 0,00 – 0,19 (descartar o revisar a profundidad)
- Pésimo: < 0,00 (descartar definitivamente)

5. Relación discriminativa (Rd): indicador que permite determinar si un reactivo o pregunta es aceptable o no a partir de su índice de discriminación. Está dado por el cociente entre el índice de discriminación del ítem -Id (i)- y su norma discriminativa (ND).⁽⁴⁾

$$Rd = \frac{Id(i)}{ND}$$

El valor de la norma discriminativa se obtuvo a partir de la siguiente fórmula:

$$ND = 0,3 * (Di) \text{ si } 33 \% \leq (Di) \leq 67 \% \text{ o } ND = 1,00 - (Di) \text{ si } 67 \% \leq (Di) \leq 100 \%$$

Si Rd mayor que 1 – se considera el ítem aceptable

Si Rd (0,60-0,99) – se recomienda analizar el ítem

Si Rd menor de 0,6 – se recomienda rechazar el ítem

6. Consistencia interna de reactivos (Cir): ⁽³⁾ se determinó a través del cálculo del coeficiente KR20 o de Küder-Richardson mediante la siguiente fórmula:

$$KR20 = \frac{k}{k-1} * \left[1 - \frac{\sum p * q}{Vt} \right]$$

KR20 = coeficiente de confiabilidad

k = número total de ítems en el instrumento

Vt = varianza total

$\sum p*q$ = sumatoria de la varianza de los ítems

p = TRC/N : total de respuestas correctas (TRC) entre el número de sujetos participantes (N)

q = 1-p

7. Consistencia interna del examen (Cie): ⁽⁴⁾ se determinó a través del cálculo del coeficiente de confiabilidad alfa de Cronbach empleando el paquete estadístico SPSS v.22.

La consistencia interna de reactivos y del examen en general se evaluó de la siguiente forma:

- menor o igual a 0,53 – confiabilidad nula
- 0,54 – 0,59 – confiabilidad baja
- 0,60 – 0,65 – confiable
- 0,66 – 0,71 – muy confiable
- 0,72 – 0,99 – excelente confiabilidad
- 1,00 – confiabilidad perfecta

El nivel de significación empleado para todas las pruebas estadísticas fue de $\alpha = 0,05$.

Conforme a los aspectos bioéticos y legales, se solicitó a las autoridades académicas correspondientes de la facultad de Estomatología, Universidad de Ciencias Médicas de Santiago de Cuba, la autorización para el desarrollo de la investigación. Se mantuvo durante todo el estudio, la confidencialidad e integridad de los datos obtenidos.

RESULTADOS

La tabla 1 muestra el fondo de tiempo de los contenidos y su presencia en el examen final. Como se aprecia, existió un balance equilibrado de horas según temas abordados. La relación entre fondo de tiempo e ítems del instrumento fue positiva perfecta, con un coeficiente de correlación de Spearman de 1,00.

Tabla 1. Fondo de tiempo de los contenidos de los temas según programa de la asignatura y presencia de los incisos en el examen

Contenidos de la asignatura	Fondo de tiempo		Debían planearse		Presencia en el examen	
	Horas	%	Incisos	%	Incisos	%
Hemolinfopoyético	16	20,5	10	20	10	20
Sistema nervioso	62	79,5	40	80	40	80

Rho de Spearman: 1,00; $p=0,000$, $p<0,05$

La correspondencia entre la cantidad de incisos según su índice de dificultad esperada con los reales del examen se exhiben en la tabla 2.

Tabla 2. Incisos de la prueba parcial de Sangre y Sistema Nervioso según índice de dificultad esperado y real

Incisos según índice de dificultad	Incisos esperados	% de Incisos esperados	Incisos reales	% de Incisos reales
Fáciles	3	6	1	2
Medianamente fáciles	10	20	8	16
De dificultad media	25	50	19	38
Medianamente difíciles	9	18	17	34
Difíciles	3	6	5	10
Totales	50	100	50	100

Rho de Spearman: 0,87; $p=0,054$, $p>0,05$

Obsérvese como en el examen existió un incremento en un 16 % de los incisos medianamente difíciles y una disminución en un 12 % de los de dificultad media con respecto a lo esperado. El resto de los reactivos se comportó bastante cercano a lo deseado. El coeficiente de correlación de Spearman, evidenció una correlación positiva considerable entre los incisos esperados y los reales obtenidos.

La dificultad por preguntas, su tipología y por temáticas, se resume en la tabla 3. Predominaron en un 60 % del instrumento evaluativo las preguntas con índice de dificultad media. Las temáticas relacionadas con aspectos morfofuncionales del dolor y del sistema motor somático,

fueron las de mayor complejidad para los estudiantes al exhibir el menor índice de dificultad. El examen en sentido general se catalogó de dificultad media.

Tabla 3. Índice de dificultad de las preguntas, su tipología, temáticas y el examen

Preguntas	Temáticas	Tipo de pregunta	Índice de dificultad
1	Sistema hemolinfopoyético	Selección múltiple complemento simple	0,64
2	Sistema nervioso central	Test de verdadero o falso	0,80
3	Sistemas somatosensoriales (dolor)	Situación clínica simulada	0,42
4	Sistema motor somático	Complementación	0,54
5	Sistema nervioso periférico	Apareamiento o asociación	0,72
Examen			0,62

Los resultados obtenidos en relación a la discriminación de los ítems de las preguntas se agrupan en la tabla 4.

Tabla 4. Índice de discriminación de los incisos según preguntas

Preguntas	Tématicas	< - 0,01	0,00 a 0,19	0,20 a 0,29	0,30 a 0,39	> 0,39	Total
1	Sistema hemolinfopoyético	1	1	1	0	7	10
2	Sistema nervioso central	1	4	2	0	3	10
3	Sistemas somatosensoriales (dolor)	0	1	1	0	8	10
4	Sistema motor	0	0	1	0	9	10
5	Sistema nervioso periférico	0	3	3	0	4	10
Total		2	9	8	0	31	50

Nótese que más de la mitad de los reactivos del examen (31) discriminan de forma adecuada para un 62 % del total. Las preguntas de hemolinfopoyético y sistema nervioso central deben ser revisadas, entre ambas dos ítems deben ser descartados de forma definitiva (4 %), cinco revisados a profundidad (10 %) y tres se sugiere revisar (6 %).

La tabla 5 proporciona una visión general sobre el índice de dificultad, de discriminación y la relación discriminativa de las preguntas.

Tabla 5. Índice de dificultad, índice de discriminación y relación discriminativa por preguntas

Preguntas	Tipo de pregunta	Índice de dificultad	Índice de discriminación	Relación discriminativa
1	Selección múltiple complemento simple	0,64	0,33	1,65
2	Test de verdadero o falso	0,80	0,23	1,64
3	Situación clínica simulada	0,42	0,30	2,14
4	Complementación	0,54	0,61	4,07
5	Apareamiento o asociación	0,72	0,31	2,07

De igual manera que en el caso de los ítems, la pregunta 2 debe revisarse a profundidad. En las restantes se aprecia buena correspondencia entre los tres indicadores y el tipo de pregunta. Se sugiere que la pregunta 5 aunque posee buena calidad, tiene posibilidades de mejorarse para futuros exámenes parciales de la asignatura. La relación discriminativa habla a favor de las preguntas 3 y 4.

La consistencia interna de los reactivos (Cir) y del examen en general (Cie) mostró una excelente confiabilidad ($KR20 = 0,92$, α de Cronbach = $0,89$).

DISCUSIÓN

El examen final de Sangre y Sistema Nervioso se estructuró de 5 preguntas con 50 ítems, según el formato de ellas se clasificó como un examen mixto, con un predominio de ítems test objetivo. Los expertos consideraron que es común en esta asignatura la elaboración de exámenes finales de este tipo, criterio compartido por los autores de la investigación.

A pesar que existió una correspondencia adecuada entre el fondo de tiempo asignado en la planificación docente a cada temática abordada, con la cantidad de incisos explorados en el examen, autores como Martínez Pérez y otros,⁽⁷⁾ manifiestan que cuando esta correspondencia falla, se resta validez al instrumento, pues el contenido del mismo debe ser una muestra representativa del constructo o dominio definido en el programa de la asignatura. Añaden además que la selección de temas a evaluar en el examen, se centra en la relevancia

y la representatividad del contenido objeto de la evaluación, por tanto, ha de reflejar fielmente los núcleos básicos de los contenidos y sus esencialidades.

En relación a la correspondencia entre la cantidad de incisos según su índice de dificultad esperada con los del examen, los resultados obtenidos coinciden con los reportados por Moreno Montañez y otros,⁽⁹⁾ quienes en el análisis de calidad de un examen final escrito de Medicina Comunitaria con 41 ítems en la facultad de Medicina de Cabinda, Angola (2019), obtuvieron un predominio de incisos con nivel de dificultad media y ausencia de ítems difíciles. Hallazgos diferentes son reportados por Saldaña, Delgadillo y Méndez citados por Gómez López,⁽²⁾ quienes evaluaron el índice de dificultad y la discriminación de un examen parcial de Bioquímica, y encontraron que el 4 % de los reactivos fueron difíciles, el 17 % medianamente difíciles, el 43 % de dificultad media y un 20 % fáciles.

Las temáticas relacionadas con el dolor y el sistema motor somático, exhibieron los más bajos índices de dificultad (0,42 y 0,54 respectivamente). A juicio de los autores, resultaron ser las de mayor complejidad por ser de tipo problémica y de complementación, por lo que exigen un nivel de asimilación aplicativo por parte de los estudiantes. En este sentido Minte Münzenmayer y otros,⁽¹⁰⁾ al reflexionar sobre el tema en su artículo “Evaluación en Educación Superior: características y demandas cognitivas de preguntas escritas,” expresan que los procedimientos evaluativos aplicados en las universidades promueven de manera general, niveles de razonamiento y dimensiones del proceso cognitivo básicos. Solo una reducida fracción de los procedimientos evaluativos logra ubicarse en los niveles superiores de razonamiento y dimensión del proceso cognitivo, fenómeno observado tanto en carreras profesionales como técnicas, lo que ha arrojado como consecuencia que los educandos desarrollen bajos niveles de asociación, produzcan respuestas elementales y no se potencien sus capacidades intelectuales. En otras palabras, los alumnos no son evaluados para lograr aprendizajes, son evaluados bajo parámetros tradicionales, centrados en la repetición, en la memoria, no en el desarrollo de competencias intelectuales de orden superior.

Rodríguez Graña,⁽¹¹⁾ Camacho Saavedra⁽¹²⁾ y González Machado,⁽¹³⁾ exponen que el nivel medio de dificultad de un examen escrito debe oscilar entre 0,5 y 0,6, lo que concuerda con los resultados de la investigación. Según dichos autores, un examen de dificultad media es aquel que más de la mitad de los educandos puede aprobar y es superior en confiabilidad que los muy fáciles o aquellos con mucha dificultad.

El índice de discriminación (Id) permitió identificar que solo la pregunta 2, de tipología test de verdadero o falso, y que abarcó los contenidos relacionados con la temática sistema nervioso central, discriminó pobremente entre los estudiantes con altas y bajas calificaciones, al obtener un índice de dificultad catalogado como medianamente fácil (0,80). Este resultado sugiere como proponen Villamarín Guevara⁽¹⁴⁾ y Menacho Chiok⁽¹⁵⁾ que debe ser revisada a profundidad por el claustro de profesores de la asignatura o descartarla de forma definitiva para futuros exámenes finales.

Los valores del coeficiente α de Cronbach y el KR20 demuestran que el instrumento tiene una excelente confiabilidad o consistencia interna, lo que concuerda con los estándares propuestos por el Colegio de Médicos de Canadá, donde se consideran excelentes entre 0,72 y 0,99; y los establecidos por la Comisión Nacional de Evaluación de la Competencia y el Desempeño del Ministerio de Salud Pública, con valores superiores a 0,60.^(16, 17)

Se concluye que la calidad del examen final aplicado en la asignatura Sangre y Sistema Nervioso de la disciplina Bases Biológicas de la Estomatología, carrera de Estomatología, Universidad de Ciencias Médicas de Santiago de Cuba, curso 2022, fue adecuada, al presentar un índice de dificultad media, buena discriminación de sus preguntas y una excelente consistencia interna.

REFERENCIAS BIBLIOGRÁFICAS

1. Blanco Pereira ME, Martínez L, González Gil A, Jordán Padrón M. Calidad del examen final teórico de Morfofisiología I en la facultad de Ciencias Médicas de Matanzas. Cursos 2012-2013 y 2013-2014. Revista Médica Electrónica. 2015 [acceso: 11/10/2022]; 37(4):323-332. Disponible en: <http://scielo.sld.cu/pdf/rme/v37n4/rme030415.pdf>
2. Gómez López VM, Rosales Gracia S, García Galaviz JL, Berrones Sánchez KI, Berrones Sánchez CM. Índice de dificultad y discriminación de ítems para la evaluación en las asignaturas básicas de medicina. Revista Educación Médica Superior. 2020 [acceso: 11/10/2022]; 34(1):e1727. Disponible en: <http://www.ems.sld.cu/index.php/ems/article/view/1727/1015>
3. Villamarín Guevara RS. Indicadores de calidad de los instrumentos de evaluación y el rendimiento académico [Tesis doctoral]. Perú: Universidad Nacional de San Marcos; 2017 [acceso: 11/10/2022]. Disponible en:

https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/6636/Villamarin_gr.pdf?sequence=3&isAllowed=y

4. Díaz Rojas PA, Leyva Sánchez E. Metodología para determinar la calidad de los instrumentos de evaluación. *Educ Med Super.* 2013 [acceso: 11/10/2022]; 27(2):269-286. Disponible en: <http://scielo.sld.cu/pdf/ems/v27n2/ems14213.pdf>
5. Sánchez Fernández NV, Sánchez Clemente A, Thomas Ferrer F, Tabera García ME. Programa de la Disciplina Estomatología Integral. Análisis del plan de estudio "D". *Revista Información Científica (RIC).* 2017 [acceso: 11/10/2022]; 96(2):280-287. Disponible en: <https://www.redalyc.org/journal/5517/551764114014/html/>
6. Ortiz Romero GM, Díaz Rojas PA, Llanos Domínguez OR, Pérez Pérez SM, González Sapsin K. Dificultad y discriminación de los ítems del examen de Metodología de la Investigación y Estadística. *EDUMECENTRO.* 2015 [acceso: 09/09/2022]; 7(2):19-35. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742015000200003
7. Martínez Pérez R, Ríos Rodríguez M, Roque Marrero Y, Caballero Pedrón KY. Evidencias de validez del examen final de la asignatura Biología molecular, carrera de Medicina. *Revista Electrónica Medimay.* 2019 [acceso: 12/9/2021];26(3):338-348. Disponible en: <http://www.medimay.sld.cu/index.php/rcmh/article/view/1483/1677>
8. Baladrón J, Sánchez-Lasheras F, Romero Ladrero JM, Curbelo J, Villacampa Menéndez P, Jiménez-Fonseca P. Evolución de los parámetros dificultad y discriminación en el ejercicio de examen MIR. Análisis de las convocatorias de 2009 a 2017. *FEM: Revista de la Fundación de Educación Médica.* 2018 [acceso: 16/10/2022]; 21(4):181-193. Disponible en: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S2014-98322018000400003
9. Moreno Montañez M, Lardoeyt Ferrer M, Ramos Arencibia AL. Calidad del examen final escrito de Medicina Comunitaria en la facultad de Medicina Cabinda, Angola. *Revista Conrado.* 2020 [acceso: 13/10/2022];16(73):214-220. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442020000200214
10. Minte Münzenmayera A, Sepúlveda Obreque H, Jaramillo Alvarado R, Díaz-Levicoy D. Evaluación en Educación Superior: características y demandas cognitivas de preguntas escritas. *Revista de Estudios y Experiencias en Educación (REXE).* 2021 [acceso: 11/10/2022]; 20(44):43-52. Disponible en: <https://revistas.ucsc.cl/index.php/rexe/article/view/1002/752>

11. Rodríguez Graña T, Sondón Fernández ME, Peña González M, Hernández Tamayo M, Robles García R. La calidad del examen final de la asignatura Morfofisiología IV. CCM. 2018 [acceso: 09/10/2022]; 22(3):446-462. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S156043812018000300009&lng=es
12. Camacho-Saavedra LA, Huamán-Saavedra JJ, Plasencia Alvarez JO. Eficiencia de preguntas de opción múltiple con tres alternativas. Rev méd Trujillo. 2020 [acceso: 16/10/2022]; 15(4):163-9. Disponible en: <https://revistas.unitru.edu.pe/index.php/RMT/article/view/3214>
13. González Machado EC, Reyes Piñuelas EP, López Ortega M. Construcción de una prueba para evaluar aprendizajes en educación superior. En: Congreso Nacional de Investigación Educativa; 2017; San Luis Potosí: Universidad Autónoma de Baja California [acceso: 16/10/2022]. Disponible en: <http://www.comie.org.mx/congreso/memoriaelectronica/v14/doc/2300.pdf>
14. Villamarín Guevara RS. Índices de la calidad de los instrumentos de evaluación. Revista redipe. 2017 [acceso: 14/10/2022];6(5):238-245. Disponible en: <https://revista.redipe.org/index.php/1/article/view/274>
15. Menacho Chiok CH, Cano Alva Trinidad JM. Evaluación de pruebas informatizadas aplicando la teoría clásica de los test y la teoría de respuesta al ítem. Anales Científicos. 2020 [acceso: 15/10/2022];81(2):278-288. Disponible en: https://revistas.lamolina.edu.pe/index.php/acu/article/view/1638/pdf_270
16. Durán-Pérez FB, Lara-Abad GE. Aplicación del coeficiente de confiabilidad de Kuder Richardson en una escala para la revisión y prevención de los efectos de las rutinas formadas durante el periodo de confinamiento a partir de la identificación del seguimiento de medidas de seguridad, de comida y de descanso. Boletín Científico de la Escuela Superior Atotonilco de Tula. 2021 [acceso: 16/10/2022]; 8(15):51-55. Disponible en: <https://repository.uaeh.edu.mx/revistas/index.php/atotonilco/article/view/6693>
17. Ríos Hidalgo N, Rivero Machado IC, Ordoñez DJ, Sitchao Suárez NM, López Vergara JM, Garcés Sigas JR. La retención de contenidos biomédicos relacionados con la enseñanza desarrolladora en Anatomía Patológica. Escuela Latinoamericana de Medicina. Panorama

Cuba y Salud. 2017 [acceso: 13/10/2022];12(2):15-23. Disponible en:
<http://www.revpanorama.sld.cu/index.php/panorama/article/view/643>