

## Metodología para determinar la calidad de los instrumentos de evaluación

### Methodology for determining the quality of the evaluation instruments

MSc. Pedro Augusto Díaz Rojas, MSc. Elizabeth Leyva Sánchez

Universidad de Ciencias Médicas de Holguín. Holguín, Cuba.

---

#### RESUMEN

La evaluación del aprendizaje es el mecanismo regulador del proceso enseñanza aprendizaje, favorece la retroalimentación del proceso y permite direccionar las acciones en pos de la calidad del producto resultante. La evaluación del aprendizaje acredita y certifica al estudiante como nuevo profesional y al residente como nuevo especialista. De ahí la importancia de obtener evidencias de la calidad de los instrumentos que se aplican en estos tipos de exámenes. Existe un arsenal de indicadores en la literatura mundial al respecto. En este trabajo se propone una metodología que ha sido validada por más de siete años y que ha permitido el desarrollo de investigaciones en diversas disciplinas y su perfeccionamiento.

**Palabras clave:** evaluación del aprendizaje, calidad de instrumentos de evaluación.

---

#### ABSTRACT

The learning evaluation is the regulating mechanism of the teaching/learning process, favors the feedback on the process and allows directing the actions to the quality of the resulting product. The learning evaluation gives accreditation and certifies a student as a new professional and a resident as a new specialist; hence the importance of obtaining evidence of the quality of those instruments used in this type of evaluations. There is a range of indicators discussed in the world literature about this topic. This paper set forth a methodology that has been validated for 7 years and has made it possible to carry out research in several disciplines and to upgrade it.

**Keywords:** learning evaluation, quality of evaluation, instruments.

## INTRODUCCIÓN

La evaluación del aprendizaje es el mecanismo regulador del proceso enseñanza aprendizaje, nos posibilita su retroalimentación, dirección y perfeccionamiento. Determina los modos de actuación que los estudiantes incorporan en su proceso formativo y acredita y certifica el cumplimiento de los objetivos terminales del futuro profesional y especialista.<sup>1,2</sup>

Salas Perea define la evaluación como la actividad que consiste en el análisis e interpretación de los resultados de diferentes instrumentos de medición sobre la actuación de un educando o candidato ante una actividad y/o tarea, en comparación con un patrón o modelo estandarizado, y que se efectúa con la finalidad de tomar una buena decisión.<sup>3</sup>

La resolución 210/2007 expresa que la evaluación es un proceso consustancial al desarrollo del proceso docente educativo. Tiene como propósito comprobar el grado de cumplimiento de los objetivos formulados en los planes y programas de estudio de la educación superior, mediante la valoración de los conocimientos y habilidades que los estudiantes adquieren y desarrollan, así como por la conducta que manifiestan en el proceso docente educativo. Constituye a su vez, una vía para la retroalimentación y la regulación de dicho proceso.<sup>2</sup>

Su resultado final es emitir un juicio que se expresa en una calificación que lo define. Las calificaciones son formas convencionales establecidas que expresan en categorías el resultado de la evaluación. Su utilización depende del tipo de evaluación: formativa o certificativa.

En el primer caso, informa al estudiante sobre sus insuficiencias y deficiencias y mide el progreso alcanzado, su *tendencia*. Nos permite ajustar las actividades de acuerdo con los avances obtenidos. Se realiza durante los estudios.

En el segundo caso, se diseña para proteger a la sociedad y prevenir la incompetencia, se utiliza para graduar los educandos por méritos, justificar decisiones y fundamentar un título o diploma. Se realiza al concluir los estudios.

Lo esencial en todos los casos radica en que la calificación sea objetiva, que exprese una correcta correspondencia con la calidad de la asimilación y por lo tanto, la necesidad de establecer criterios uniformes, que disminuyan la influencia del factor subjetivo, y que frente a una actuación determinada, con iguales resultados, el grupo de profesores otorguen iguales calificaciones.<sup>1,3</sup>

La aplicación de las diferentes formas de evaluación se puede realizar mediante múltiples métodos, procedimientos y técnicas. El diseño fundamental aplicado en la evaluación certificativa es el examen teórico, práctico y teórico-práctico.

Los instrumentos evaluativos escritos pueden dividirse en dos grandes familias: los exámenes de desarrollo, de respuesta libre, tradicionales o de ensayo y los exámenes de respuesta estructurada o test objetivos.

Puede existir combinación de preguntas de uno u otro tipo y entonces se denominaría examen mixto.

Esta clasificación se basa en la categoría de las preguntas o "ítems", que puedan ser de respuestas libres o de ensayo y de respuestas estructuradas o test objetivos.<sup>3,4</sup>

Un elemento esencial es la planeación de los instrumentos de evaluación, que debe realizarse desde las primeras semanas del curso, se reajusta según los resultados de la evaluación formativa.<sup>5</sup>

Al construir un instrumento evaluativo, este debe ser planificado, con el fin de contribuir a incrementar la efectividad y la eficiencia, en la literatura<sup>5</sup> se plantean las etapas siguientes:

- Definir qué debe ser evaluado.
- Seleccionar los métodos y procedimientos a emplear.
- Cómo calificar el examen.
- Cómo establecer un patrón de aprobado/desaprobado.

En la actualidad, con el incremento del contenido de las ciencias y la tendencia a su integración,<sup>6</sup> podemos definir como pasos para la selección de los contenidos de una evaluación, la identificación de los núcleos básicos de contenidos y sus esencialidades, su relación con los problemas que los educandos deben ser capaces de enfrentar y/o resolver, las tareas que debe desarrollar para darles solución y que definen su futura competencia.

Con estos elementos definidos, podemos elaborar la tabla de contenido del examen y la estructura del futuro instrumento de evaluación, al establecer los niveles de asimilación, las habilidades a explorar, el fondo de tiempo asignado al contenido en el programa, su importancia, su relación con los objetivos generales de las asignaturas y disciplinas y con los objetivos terminales del egresado.

De esta forma logramos que los instrumentos evaluativos reúnan los requisitos de que sus contenidos estén directamente relacionados con los objetivos educacionales y que al aplicarlos demuestren su validez y confiabilidad.<sup>1,3,7</sup>

Hay muchas más propiedades que se exigen a los instrumentos evaluativos los que pueden ser consultados en la literatura publicada.

La validez es la correspondencia entre lo que se pretende verificar por el instrumento, procedimiento o método y lo que realmente se mide o evalúa, y ello abarca tanto a las exigencias propias del control como a la determinación de los índices valorativos que permiten la calificación.<sup>3</sup>

Podemos preguntarnos: ¿Mide el instrumento evaluativo lo que realmente debe medir? ¿Cuál es el grado de correspondencia entre las preguntas del instrumento evaluativo y los contenidos a evaluar?<sup>8</sup>

En la medida en que las actividades de la evaluación y el diseño del instrumento se aproximan a la realidad, o sea, están dirigidas a valorar o medir la aplicación social e independiente del sistema de conocimientos, hábitos, habilidades y modos de actuación por el estudiante, mayor será la validez de dicho control. Mientras más teórico, reproductivo y alejado de la práctica esté un control, mucho menor será su validez.<sup>8</sup>

La confiabilidad es la estabilidad en los resultados de un control, ya sea al repetirlo, o al ser calificado por distintos profesores. Significa que hay constancia en los

resultados obtenidos y que, por tanto, es representativo del grado de aprovechamiento alcanzado por el estudiante en el tipo de control realizado.<sup>3,8,9</sup>

La validez y la confiabilidad están estrechamente relacionadas. Un control que cumpla con las exigencias de la validez, tiene un alto grado de probabilidad de ser confiable, sin embargo, no necesariamente ocurre así a la inversa.

Para lograr estas propiedades debemos tener presente el análisis que se realice en el colectivo de profesores en 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. La experiencia por parte del colectivo de profesores en la temática evaluación es esencial para el logro de un instrumento válido y confiable.

Pero no es hasta que se obtienen los resultados, que el colectivo puede hacer las modificaciones y rectificaciones necesarias, es por ello la importancia de brindar una metodología que permita determinar la calidad de los instrumentos evaluativos, que sirva de guía para el perfeccionamiento de este importante componente del proceso enseñanza aprendizaje.

La metodología que presentamos ha sido utilizada y validada en varias investigaciones de calidad de instrumentos, tanto del pregrado como del posgrado, por más de siete años y ha servido para el perfeccionamiento continuo de diversas asignaturas y como valor agregado ha permitido la elaboración de varias tesinas y tesis de maestría en Educación Médica, muchas de ellas ya publicadas en diferentes medios.<sup>10-15</sup>

## **Metodología propuesta**

### Introducción a la metodología

La posibilidad de contar con un algoritmo lógico para el análisis de la calidad de los instrumentos evaluativos, que permita retroalimentar el proceso a partir de definir las áreas de contenidos con mayores problemas de asimilación por parte de los estudiantes, permitirá que los colectivos docentes puedan intervenir en el proceso docente educativo, haciendo énfasis en aquellos aspectos con dificultades, no solo para el tratamiento a los contenidos y las habilidades sino también para la preparación de la asignatura.

No proponemos matematizar los resultados del análisis de la calidad de un instrumento, sino, favorecer la identificación de las causas que influyen en resultados no satisfactorios y que se tomen las medidas adecuadas en la planeación del futuro instrumento y se retroalimente el proceso.

Nos hemos basado en prestigiosos investigadores en este campo, de donde hemos obtenidos las mejores experiencias.

Las asignaturas y carreras están obligadas a elaborar los instrumentos de evaluación para el examen final y sus variantes de extraordinario y examen de fin de curso, esto implica elaborar varios instrumentos durante un largo período, mientras no se cambie el currículo. De aquí la importancia de contar con una metodología que permita analizar de manera objetiva la calidad de estos instrumentos mediante indicadores de validez y confiabilidad.

### Pasos de la metodología

#### I. Análisis del examen aplicado.

En este acápite se debe tener en cuenta:

- Tipo de examen.
- Cantidad de preguntas del examen y temáticas que abordan.
- Cantidad de ítems por preguntas y área temática.
- Tipos de preguntas que contiene el instrumento.
- Clave de calificación y decisión de éxito o fracaso.
- Resultado obtenido en cada ítem, en términos de correcto e incorrecto.
- Calificaciones por preguntas y del examen en general.

II. Elaboración de la hoja de cálculo con los datos primarios de los resultados de cada estudiante en su examen.

En una hoja de cálculo de Microsoft Excel se tabulan los resultados del examen objeto de estudio.

Se organiza de la siguiente manera:

Las columnas recogerán los ítems de cada pregunta y a continuación la nota otorgada a la pregunta. A los ítems se le asigna la categoría de correcto con C o incorrecto con I. Puede usarse cualquier otro símbolo para clasificar los resultados. Al final, la última columna corresponde a la calificación obtenida en el examen.

Las filas se asignan a los estudiantes, para lo cual es muy importante asignar en la primera columna una clave que permita identificar el grupo de los estudiantes. Ello favorecerá el análisis por los colectivos docentes e incluso por las unidades docentes que intervienen en la impartición de la asignatura.

III. Correspondencia del fondo de tiempo asignado en el P1 a los contenidos y la cantidad de ítems que lo exploran.

Para ello se deben tabular los contenidos con el número de horas que se asignan en el modelo de planificación docente y a continuación la cantidad de ítems que exploran el contenido. En cada caso se obtiene el porcentaje correspondiente.

Este análisis puede hacerse también en función de áreas temáticas agrupando varios temas de interés y de igual manera se puede realizar por ciencias particulares en aquellas asignaturas que son integradas por más de una ciencia.

El resultado que se obtenga nos permitirá establecer la correspondencia entre el tratamiento de contenidos y su exploración en el examen. Debe encontrarse una proporción adecuada de ítems en función del tiempo que se le asignó al tratamiento del contenido.

En las asignaturas integradas por varias ciencias nos daría la proporción de cada una de ellas en el examen y su peso en tiempo.

Para concluir el análisis se puede aplicar el coeficiente de correlación de Spearman.<sup>16</sup>

#### IV. Definición de la escala de clasificación de la dificultad según expertos.

Para poder comparar los resultados y clasificar los ítems, preguntas y el examen, atendiendo al índice de dificultad, nos apoyaremos en los especialistas de la asignatura, con vistas a definir los intervalos según la proporción de estudiantes aprobados para cada componente.

Se utiliza la propuesta de Bachkofft que define que el examen debe clasificarse en ítems o preguntas difíciles, medianamente difíciles, de dificultad media, medianamente fáciles y fáciles.

Un ejemplo pudiera ser:

- Menos de 40 % de aprobados se consideran difíciles.
- Entre 41 y 50 % de aprobados se consideran medianamente difíciles.
- Entre 51 y 80 % de aprobados se consideran de dificultad media.
- Entre 81 y 90 % de aprobados se consideran medianamente fáciles.
- Entre 91 y 100 % de aprobados se consideran fáciles.

Esta escala puede variar atendiendo al criterio de los especialistas de más experiencia de la asignatura, pero debe ser congruente con la calidad que se exige en los resultados de un examen.

Para poder comparar posteriormente con otros instrumentos evaluativos, se debe mantener la misma clasificación.

#### V. Obtención de indicadores de validez del instrumento.

- Índice de dificultad

De cada ítem:

Se determina el índice de dificultad de cada uno de los ítems, se divide el número de estudiantes que contestó correctamente el ítem entre el número total que contestaron el ítem. Esta proporción se le denota como  $D_i$  y se utiliza la siguiente fórmula (fórmula 1):<sup>9,17-20</sup>

$$D_i = \frac{A}{N}$$

Donde:

$D_i$  = Índice de dificultad del ítem.

A = Número de respuestas correctas del ítem.

N = Número total de estudiantes que contestaron el ítem.

- Índice de dificultad de las temáticas

En este caso se suma el total de estudiantes que respondieron correctamente cada uno de los incisos de la temática y se divide entre el número de incisos que explora el tema. Se calcula la media de respuestas correctas para el tema y se divide entre el total de examinados, como se representa en la siguiente fórmula (fórmula 2):

$$Dt = \frac{\sum (A_{t1} + A_{t2} + \dots + A_{tn})}{\frac{n_{tt}}{Nt}}$$

Donde:

Dt = Índice de dificultad de la temática.

A<sub>t</sub> = Número de respuestas correctas de los incisos de la temática.

n<sub>tt</sub> = Número total de incisos que exploran la temática.

Nt = Número total de presentados en el examen.

- Índice de dificultad de las preguntas

Se suman el total de estudiantes que respondieron correctamente cada uno de los ítems que corresponden a la pregunta y se divide entre el número de ítems de la pregunta. Se calcula la media de respuestas correctas para la pregunta y se divide por el total de examinados. Se utiliza la fórmula anterior aplicada a la pregunta como se muestra a continuación en la fórmula (fórmula 3):

$$Dp = \frac{\sum (A_{p1} + A_{p2} + \dots + A_{pp})}{\frac{n_{tp}}{Nt}}$$

Donde:

Dp = Índice de dificultad de la pregunta.

A<sub>p</sub> = Número de respuestas correctas de los incisos de la pregunta.

n<sub>tp</sub> = Número total de incisos de la pregunta.

Nt = Número total de presentados en el examen.

- Índice de dificultad de las ciencias integradas

Se suman el total de estudiantes que respondieron correctamente cada uno de los ítems que corresponden a los contenidos particulares de una ciencia explorada en el examen y se divide entre el número de ítems de esa ciencia presente en el examen. Se calcula la media de respuestas correctas para la ciencia y se divide por el total de

examinados. Se utiliza la fórmula anterior aplicada a la pregunta como se muestra a continuación en la siguiente fórmula (fórmula 4):

$$Dc = \frac{\sum (A_{c1} + A_{c2} + \dots + A_{cp})}{\frac{n_{tc}}{Nt}}$$

Donde:

Dc = Índice de dificultad del contenido de la ciencia.

Ac = Número de respuestas correctas de los incisos que exploran la ciencia.

n<sub>tc</sub> = Número total de incisos que exploran la ciencia.

Nt = Número total de presentados en el examen.

Comparación de los resultados según índice de dificultad obtenido (real) con la dificultad esperada:

A partir de la obtención de los índices de dificultad, estamos en condiciones de clasificar cada ítem y pregunta o temática, según la escala de clasificación definida en el colectivo. Estos resultados se tabulan. Para ello se confeccionó una tabla donde se expresaron los resultados que debíamos esperar con los obtenidos en realidad, según la clasificación.

Escala de clasificación de los ítems según la dificultad esperada, considerada en la literatura sobre esta temática es como sigue:<sup>9</sup>

5 % para ítems fáciles.

20 % para ítems medianamente fáciles.

50 % para ítems de dificultad media.

20 % para ítems medianamente difíciles.

5 % para ítems difíciles.

Para ratificar el comportamiento de los ítems y preguntas podemos utilizar la correlación por rangos de Spearman como test no paramétrico que es menos restrictivo.<sup>22</sup>

- Análisis de la discriminación de los ítems y preguntas

La discriminación, conocida también como poder de discriminación, expresa la medida en que el ítem o pregunta discrimina entre los buenos estudiantes y los peores estudiantes. Se entiende que los buenos estudiantes deben haber respondido en su mayoría correctamente el ítem y los estudiantes malos respondieron incorrectamente el ítem.



Este indicador nos permite además predecir el comportamiento del ítem o pregunta dentro del examen, para la elaboración de nuevos instrumentos con estructura bastante similar.

Para el estudio de este indicador podemos utilizar el Índice de discriminación y el Coeficiente de discriminación.<sup>9,17-19, 21</sup>

Índice de discriminación:

Para poder aplicar este indicador debemos organizar los datos, de la siguiente forma, en una columna se ponen los resultados de la pregunta y en la columna siguiente los resultados del ítem. A continuación se ordena la columna de la pregunta de la nota mayor a la nota menor, siempre se debe aceptar la opción "ampliar la selección" de la aplicación Excel. Con esta organización determinamos qué número de estudiantes representa el 27 % de los examinados.

Se procede entonces 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).

La fórmula para su cálculo es la siguiente fórmula (fórmula 5):

$$D_i = \frac{GA_{\text{aciertos}} - GB_{\text{aciertos}}}{N_{\text{grupo mayor}}}$$

Donde:

$D_i$ : Índice de discriminación del reactivo  $i$

$GA_{\text{aciertos}}$ : Número de aciertos en el reactivo  $i$  del 27 % de personas con las puntuaciones más altas en el test.

$GB_{\text{aciertos}}$ : Número de aciertos en el reactivo  $i$  del 27 % de personas con las puntuaciones más bajas en el test.

$N_{\text{grupomayor}}$ : Número de personas en el grupo más numeroso (GA o GB).

Mientras más alto es el índice de discriminación, el reactivo diferenciará 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).

Algunos autores han definido la siguiente regla para determinar la calidad de los reactivos, en términos del índice de discriminación. A continuación se muestran los valores  $D$  y su correspondiente interpretación. Asimismo, se señalan las recomendaciones para cada uno de estos valores (tabla).

**Tabla.** Calidad de los reactivos y recomendaciones

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 - 0,19	Pobre	Descartar o revisar a profundidad
< -0,01	Pésima	Descartar definitivamente

Debe tomarse en consideración que el máximo valor del índice de discriminación de un ítem tiene estrecha relación con su índice de dificultad, ya que un reactivo con valor muy alto o muy bajo de índice de dificultad en realidad no discrimina, pues es solucionado por la mayoría de los estudiantes o por la minoría de ellos respectivamente.

A partir de este análisis, se propone utilizar la norma discriminativa, que tiene como objetivo determinar el valor óptimo que debe presentar un ítem en su índice de discriminación, de acuerdo al índice de dificultad del mismo.

El valor de la norma discriminativa es la siguiente fórmula (fórmula 6):

$$ND = \begin{cases} 0,3 * ID & \text{si } 27 \% \leq ID \leq 73 \% \\ 100 - ID & \text{si } 73 \% \leq ID \leq 100 \% \end{cases}$$

Con la norma discriminativa podemos elaborar otro indicador que indique si un reactivo es aceptable o no según su valor de índice de discriminación, este índice se conoce como relación discriminativa (RD), y está dado por el cociente entre el índice de discriminación (D) del ítem y su norma discriminativa (ND) (fórmula 7):

$$RD_i = \frac{D_i}{ND_i}$$

Se considera que si la relación discriminativa es mayor que 1, significa que el índice de discriminación es mayor que la norma discriminativa y por tanto el ítem se puede considerar aceptable, si el valor es menor que 1, se recomienda analizar el ítem, si el valor es inferior a 0,6 se sugiere desechar el ítem.

- Coeficiente de discriminación del punto biserial ( $r_{pbis}$ ):

Este indicador permite calcular la efectividad discriminativa de un reactivo, entiéndase por ello un ítem o una pregunta.

La correlación del punto biserial ( $r_{pbis}$ ) se emplea para saber si los estudiantes aplicados son los que obtienen las respuestas correctas, qué tanto poder predictivo tiene el reactivo, y cómo puede contribuir a las predicciones.<sup>20</sup>

Se utiliza la siguiente fórmula (fórmula 8):

$$r_{pbis} = \frac{\overline{X_1} - \overline{X_0}}{S_x} * \sqrt{\frac{n_1 \cdot n_0}{n(n-1)}}$$

Donde:

$X_1$ = Media de las puntuaciones totales de aquellos que respondieron correctamente el ítem.

$X_0$ = Media de las puntuaciones totales de aquellos que respondieron incorrectamente el ítem.

$S_x$  = Desviación estándar de las puntuaciones totales.

$n_1$ = Número de estudiantes que respondieron correctamente el ítem.

$n_0$ = Número de estudiantes que respondieron incorrectamente el ítem.

$n = n_1 + n_0$

Según estándares internacionales la distribución de los resultados del coeficiente discriminación punto de correlación biserial ( $r_{pbis}$ ) es:

- < 0 discriminan negativamente.
- 0 0,14 discriminan pobremente.
- 0,15 0,25 discriminan regular.
- 0,26 0,35 buen poder discriminativo.
- > 0,35 excelente poder de discriminación.

El  $r_{pbis}$  puede ser calculado para cada ítem y para cada pregunta.

#### VI. Análisis de la confiabilidad de las preguntas y el examen.

Existe una variedad de indicadores que se utilizan para evaluar la confiabilidad de un instrumento evaluativo. En nuestro caso proponemos usar el coeficiente de alfa de Cronbach que permite determinar la consistencia interna de un instrumento y brinda información relacionada con el aporte que hace cada pregunta a la consistencia general del examen.

- Coeficiente  $\alpha$  de Cronbach

La fórmula que se emplea es la siguiente (fórmula 9):

$$\alpha = \frac{nr}{1 + r(n - 1)}$$

Donde n es el número de elementos del examen y r su correlación media.

Este coeficiente se puede obtener utilizando las aplicaciones estadísticas profesionales como el SPSS y el SyStat.

El alfa de Cronbach toma un valor positivo. Para muchos autores un valor aceptable para un examen es de 0,60 a 0,70.<sup>22,23</sup>

Los resultados deben tabularse atendiendo al alfa del examen y de cada pregunta, que se calcula excluyendo los datos de esa pregunta.

Su interpretación se realiza tomando en cuenta cómo aumenta o disminuye el alfa sin la pregunta, si el valor es igual o inferior al alfa del examen la pregunta aporta consistencia interna al examen, pues al eliminarla el coeficiente desciende, en caso contrario, si el valor del alfa del examen se eleva al eliminar la pregunta, es un indicador de que la pregunta afecta la consistencia interna y disocia los resultados del examen.

### **Aplicación de la metodología al examen final de la asignatura Morfofisiología Humana III del curso 2010 2011 de la Facultad de Ciencias Médicas de Holguín**

#### **I. Criterio de expertos.**

Los resultados de la entrevista a expertos nos permitió por consenso clasificar el índice de dificultad que debe caracterizar las preguntas del examen que quedó de la siguiente forma:

Más de 90 se consideran fáciles.

Entre 80 y 89 se consideran medianamente fáciles.

Entre 70 y 79 se consideran de dificultad media.

Entre 50 y 69 se consideran medianamente difíciles.

Menos de 50 se consideran difíciles.

Además se clasificaron los tipos de preguntas predominando las de tipo verdadero y falso (tres preguntas), ensayo corto (2 preguntas), y de complemento simple (2 preguntas).

#### **II. Análisis del instrumento evaluativo.**

De la recolección de los datos de los documentos oficiales de la asignatura y de los 261 exámenes se procedió a la tabulación y obtención de los diversos indicadores de la metodología explicada.

Fondo de tiempo de los contenidos y su presencia en el examen.

En el cuadro 1 se aprecia que el contenido al que se le asignó mayor tiempo para el tratamiento didáctico fue sistema sensoriales con 24 h que representa 20,69 % del total de horas de la asignatura, sin embargo este contenido cuenta en el examen 7 incisos que representa el 17,07 %. La temática sistema nervioso suprasegmentario se evaluó con 10 incisos para un 24,39 %.

La relación entre fondo de tiempo e ítem del examen no fue la más adecuada, con un coeficiente de correlación de Spearman de 0,55 (cuadro 1).

**Cuadro 1.** Distribución del fondo de tiempo según programa y presencia de los incisos en las temáticas

Contenidos	Fondo de tiempo		Debían planearse		Presencia en el examen	
	Horas	%	Incisos	%	Incisos	%
Tema 1. Generalidades del sistema nervioso.	12	10,34	4	10,34	7	17,07
Tema 2. Sistema nervioso segmentario.	20	17,24	7	17,24	5	12,20
Tema 3. Sistema nervioso suprasegmentario.	20	17,24	7	17,24	10	24,39
Tema 4. Meninges.	10	8,62	3	8,62	5	12,20
Tema 5. Sistemas sensoriales.	24	20,69	10	20,69	7	17,07
Tema 6. Sistemas motores somático y visceral.	20	17,24	7	17,24	5	12,20
Tema 7. Actividad nerviosa superior.	10	8,62	3	8,62	2	4,88

Correlación de Spearman  $n=0,55$

En el cuadro 2 observamos que la correspondencia entre la cantidad de incisos según su índice de dificultad esperada con los reales del examen.

**Cuadro 2.** Incisos del examen según índice de dificultad esperado y real

Incisos según su índice de dificultad	Incisos esperados	% Incisos esperados	Incisos reales	% Incisos reales
Fáciles	2	5,00	5	12,20
Medianamente fáciles	8	20,00	12	29,27
De dificultad media	21	50,00	17	41,46
Medianamente difíciles	9	20,00	6	14,63
Difíciles	2	5,00	1	2,44
Totales	41	100	41	100

Correlación de Spearman 0,872

Solo llama la atención el incremento de íncisos fáciles (5) sobre lo esperado (2). El resto se comporta de manera bastante cercana a lo esperado. Al aplicar el coeficiente de Correlación de Spearman el valor obtenido es de 0,87, mostrando adecuada correlación.

- Índice de dificultad por preguntas, temáticas y examen:

El índice de dificultad para el examen de Morfofisiología Humana III fue de 0,97. Las preguntas de mayor dificultad fueron 6, 7 y 3, con valores por debajo de 0,9. Meninges, actividad nerviosa superior y sistemas motores somático y visceral representan los contenidos evaluados en dichas preguntas respectivamente (cuadro 3).

**Cuadro 3.** Índice de dificultad de las preguntas, temáticas y el examen

Preguntas	Temática	Tipo de pregunta	Índice de dificultad
1	I	Verdadero y falso	0,91
2	II	Complemento simple	0,95
3	VI	Complemento simple	0,89
4	III	Verdadero y falso	0,92
5	V	Verdadero y falso	0,97
6	IV	Ensayo corto	0,80
7	VII	Ensayo corto	0,88
		Examen	0,97

- Índice de dificultad por áreas de las ciencias básicas

El índice de dificultad por áreas de las ciencias que integran la Morfofisiología Humana III presentó valores superiores a 0,80 en la Anatomía, Fisiología y Embriología, mientras que la Histología y la Bioquímica fueron las de mayor dificultad con índices por debajo de 0,75.

En todos los casos se evalúan en el rango de preguntas medianamente fáciles y fáciles.

- Índice de discriminación de los ítems de las preguntas

Empleamos la escala aceptada internacionalmente para este índice, y los incisos se distribuyen entre 0,00 a 0,19 y 0,20 hasta 0,29. En el primer caso se sugiere descartar o revisar a profundidad y en la segunda la necesidad de revisar.

De forma general se aprecia que 29 ítems requieren una revisión profunda y 12 una revisión, lo que demuestra que la discriminación no es la más adecuada. Las preguntas más afectadas son la 4 y la 5 (cuadro 4).

**Cuadro 4.** Índice de discriminación de los incisos según preguntas

Preguntas	Temática	0,00 a 0,19	0,20 a 0,29
1	I	3	4
2	II	3	2
3	VI	2	3
4	III	6	1
5	V	7	0
6	IV	4	1
7	VII	4	1
	Examen	29	12

- Análisis del índice de dificultad, del índice de discriminación y la relación discriminativa de las preguntas

De igual manera que en el caso de los ítems, las preguntas 5, 6 y 7 deben ser revisadas en profundidad y las restantes deben ser revisadas, como se observa no hay correspondencia entre los tres indicadores y el tipo de pregunta, ello debe estar en relación con la clave de calificación aplicada en el examen. La relación discriminativa habla a favor de la pregunta 3 (cuadro 5).

**Cuadro 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	Verdadero y falso	0,91	0,21	1,77
2	Complemento simple	0,95	0,23	1,15
3	Complemento simple	0,89	0,23	7,68
4	Verdadero y falso	0,92	0,20	2,50
5	Verdadero y falso	0,97	0,16	1,49
6	Ensayo corto	0,80	0,16	3,15
7	Ensayo corto	0,88	0,19	2,15

- Confiabilidad del examen por preguntas

La confiabilidad del examen y por preguntas es aceptable, oscilando con valores del Alfa de Cronbach de 0,75 a 0,80 lo que demuestra consistencia interna en la

estructura del instrumento, no obstante los aspectos señalados en relación con la dificultad y la discriminación.

La correcta aplicación de un sistema de evaluación del aprendizaje, que tiene un carácter cualitativo integrador, se basa en un conjunto de condiciones que constituyen las premisas básicas que se deben satisfacer, las cuales se encuentran estrechamente relacionadas entre sí y cuya inobservancia impide que se cumplan efectivamente sus funciones.

En este instrumento empleado como ejemplo se aprecia dificultades en su planeación, en la elaboración de las claves de calificación y en la decisión de éxito fracaso que hace que se considere fácil y con baja discriminación.

## **CONCLUSIONES**

La calidad de un instrumento depende en gran medida de la planeación que se realice con suficiente antelación y constante revisión y actualización.

La aplicación de la metodología propuesta tiene como objetivo ayudar a la realización de un análisis más científico y normalizado, que permita determinar los contenidos con resultados deficientes y establecer su tratamiento posterior.

El resultado final del análisis de calidad de un instrumento debe incluir la valoración que realicen los profesores de los colectivos de las asignaturas en su planeación y los resultados de la aplicación de la metodología propuesta.

## **REFERENCIAS BIBLIOGRÁFICAS**

1. Álvarez de Zayas C. La dirección del proceso docente educativo. En: Hacia una escuela de Excelencia. La Habana: Editorial Academia; 1996.
2. Ministerio de Educación Superior. Reglamento de Trabajo Docente y Metodológico. Resolución 210; 2007. La Habana: Gaceta Oficial de la República de Cuba; 2007.
3. Salas Perea RS. La evaluación en la educación superior contemporánea. Biblioteca de Medicina. Volumen XXIV. UMSA, La Paz, Bolivia, 1998. CD-ROM Maestría en Educación Médica, ENSAP; 2005.
4. González Pérez M. La evaluación del aprendizaje.- tendencias y reflexión crítica. Educ Med Super. 2000;20(1):47-62.
5. Salas Perea R. Planeamiento de un examen. Biblioteca de Medicina. Volumen XXIV. UMSA, La Paz, Bolivia, 1998. En CD-ROM Maestría en Educación Médica, ENSAP; 2005.
6. Vicedo Tomey A. La integración de conocimientos en la educación médica. Educ Med Super. 2009;23(4).



7. Segura Castillo M. La perspectiva ética de la evaluación de los aprendizajes desde un enfoque constructivista. Rev Actualidades Investigativas en Educación [Internet]. 2007 ene-abr [citado 27 Sep 2011]; 7(1) [Aprox. 22 p]. Disponible en: <http://www.dialnet.unirioja.es/servlet/articulo?codigo=2273870>
8. Pérez Tapia JH, Acuña Aguila N, Arratia Cuela E. Nivel de dificultad y poder de discriminación del tercer y quinto examen parcial de la cátedra de cito-histología 2007 de la carrera de medicina de la UMSA. Rev Cuadernos [Internet]. 2008 [citado 22 Abr 2011];53(2): [aprox. 14 p]. Disponible en: <http://www.revistasbolivianas.org.bo/pdf/chc/v53n2/v53n2a03.pdf>
9. Backhoff E, Larrazolo N, Rosas M. Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). Revista Electrónica de Investigación Educativa [Internet]. 2000. [citado 2 Oct 2011];2(1): [aprox. 14 p]. Disponible en: <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>
10. Marín Hernández D. Estructura y valoración de la calidad del examen final de Morfofisiología Humana I en el Estado Táchira. Curso 2005-2006 [tesis] La Habana: ENSAP; 2007.
11. Carrazana Lee A. Nivel de dificultad y poder de Discriminación del examen diagnóstico de la asignatura Morfofisiología Humana I. Educ Med Super [Internet]. 2011 ene-mar [citado 22 abril 2011];25(1): [aprox. 12 p]. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864=es)
12. Argudín Samonte E, Díaz Rojas P, Leyva Sánchez E. Índice de Dificultad del examen de Morfofisiología Humana I. Educ Med Super [Internet]. 2011 abr-Jun [citado 2 Oct 2011];25(2): [aprox. 16 p]. Disponible en: [http://bvs.sld.cu/revistas/ems/vol25\\_2\\_11/ems07211.htm](http://bvs.sld.cu/revistas/ems/vol25_2_11/ems07211.htm)
13. Moreno Montañez M. Caracterización del examen estatal escrito de la especialidad en Medicina General Integral. Primera convocatoria. Venezuela. Mayo 2006 [tesis]. La Habana: ENSAP; 2007.
14. Santiesteban Leyva E. Propuesta de metodología para el análisis de los resultados de los exámenes finales de las asignaturas de la disciplina Morfofisiología Humana del Programa Nacional de Formación de Medicina Integral Comunitaria en la República Bolivariana de Venezuela [tesis]. La Habana: ENSAP; 2007.
15. Abreu Reyes D. Caracterización del examen teórico final de Morfofisiología Humana del perfil Laboratorio Clínico. Educ Med Super [Internet]. 2011 ene-mar [consultado 6 Feb 2012];25(1): [aprox. 9 p]. Disponible en: [http://www.scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21412011000100003&lng=es&nrm=iso&tlng=es](http://www.scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000100003&lng=es&nrm=iso&tlng=es)
16. Díaz Rojas P. Introducción a la Investigación en Ciencias de la Salud. La Habana: Escuela Nacional de Salud Pública; 2010.
17. Domínguez Mendoza TN, Godoy Calderón S. Recomendaciones metodológicas para el análisis de reactivos de opción múltiple en exámenes que integran diversas áreas del conocimiento. Ciudad México: Instituto Politécnico Nacional; 2008.
18. Pérez Tapia JH, Acuña Aguilar N, Arratia Cuela ER. Nivel de dificultad y poder de discriminación del tercer y quinto examen parcial de la cátedra de cito-histología 2007 de la carrera de medicina de la UMSA. Revista Cuadernos. 2008;53(4).

19. Sánchez González MaG. Descripción del nivel de facilidad y poder de discriminación del examen de inferencia estadística en métodos estadísticos en Medicina veterinaria y Zootecnia. Facultad de Medicina Veterinaria y Zootecnia. México. Jornada de Educación Médica [Internet]. 2010 [citado 2 Ene 2012]: [aprox. 9 p]. Disponible en:  
[http://www.facmed.unam.mx/sem/jem2010/Abstracts/trabajos\\_orales.pdf](http://www.facmed.unam.mx/sem/jem2010/Abstracts/trabajos_orales.pdf)
20. Caraballo Carballo RM, Rico Romero L, Lupiáñez Gómez JL. Análisis de los ítems de las evaluaciones autonómicas de diagnóstico en España: 2008-2009. Rev Iberoamericana de Educación Matemática [Internet]. 2011 Junio [citado 2 Oct 2011]; (26): [aprox. 16 p]. Disponible en: [http://www.fisem.org/web-/union/index.php?option=com\\_content&view=article&id=32&Itemid=3](http://www.fisem.org/web-/union/index.php?option=com_content&view=article&id=32&Itemid=3)
21. Molina Soldán EM. Instrumentos de evaluación en el proceso de enseñanza aprendizaje. Revista digital Investigación y Educación [Internet]. 2006 Agosto [citado 25 Feb 2012]; 26(3): [aprox. 7 p]. Disponible desde:  
[http://www.google.com.cu/#sclient=psy-ab&hl=es\\_19&q=Molina+E.=558](http://www.google.com.cu/#sclient=psy-ab&hl=es_19&q=Molina+E.=558)
22. Ministerio de Salud Pública. Proyecto de Indicaciones Metodológicas para el planeamiento, diseño y control de la calidad de los exámenes escritos para evaluación de la competencia y el desempeño profesionales. La Habana: Escuela Nacional de Salud Pública; 2001.
23. Córlica JL, Hernández Aguilar ML, Holloway Creed J, Dimos C. Propuesta de indicadores de calidad para la evaluación de la calidad de las preguntas en los exámenes de opción múltiple. I Congreso virtual de educación a distancia [Internet]. 2004 [consultado 6 Feb 2012]. Disponible en:  
[http://www.ateneonline.net/datos/91\\_03%20PONENCIA%20CORICA-ERNANDEZ-HOLLOWAY-DIMOU.pdf](http://www.ateneonline.net/datos/91_03%20PONENCIA%20CORICA-ERNANDEZ-HOLLOWAY-DIMOU.pdf)

Recibido: 30 de noviembre de 2012.  
Aprobado: 10 de diciembre de 2012.

*Pedro Augusto Díaz Rojas*. Universidad de Ciencias Médicas de Holguín. Ave. Lenin No. 4. Holguín, Cuba. Teléfono: 462650. Correo electrónico: [pdiaz@infomed.sld.cu](mailto:pdiaz@infomed.sld.cu)