

## DISEÑO DE UN INSTRUMENTO PARA EVALUAR LA CALIDAD DE LOS EXAMENES ESCRITOS (TEORIA) DE LOS ESTUDIANTES DE MEDICINA

### Designing an instrument to assess the quality of written examinations (theory) of medical students

\*Eduardo Luis Suárez Barrientos,

\*\*Ciro Larrazabal Córdova,

\*\*\* Julio Hinojoza Rodriguez

Recibido: 16-09-10; Aceptado: 1-10-10

#### RESUMEN

Presentamos el diseño de un instrumento para analizar cómo se realiza la evaluación de los conocimientos en las pruebas escritas (exámenes teóricos) que se aplican a los estudiantes de medicina. Si bien se trata de una evaluación del diseño de la evaluación sumativa propiamente dicha, cuyo objetivo es medir el aprovechamiento de los alumnos de forma cuantitativa, entonces deberemos preocuparnos de la calidad de este diseño.

Se procedió inicialmente a una revisión bibliográfica, deteniéndonos fundamentalmente en tres trabajos: la taxonomía de los objetivos educacionales de Bloom, la versión revisada de Anderson, y el trabajo de Case y Swanson sobre como elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Luego de un análisis y reflexión de estos y otros trabajos, se diseñó un instrumento que exige tres ejercicios: primero identificar los contenidos de determinada asignatura o materia, que se consideren imprescindibles, importantes y complementarios; segundo identificar la distribución de las preguntas en el examen; y tercero analizar las habilidades de pensamiento que implícitamente se exige utilizar a los estudiantes en la resolución del examen. El análisis global de estos criterios permitirá evaluar la coherencia en la construcción global del examen.

Se discute como el diseño de una evaluación sumativa desafiante, productiva y enriquecedora como resultado de un ejercicio cuidadoso de planificación, exigiría al estudiante una forma más calificada de aprender, por lo tanto este ejercicio no dejaría de reconocer simultáneamente también su valor formativo en los estudiantes

**Palabras Clave:** Metaevaluación, exámenes escritos, evaluación sumativa, evaluación formativa

#### ABSTRACT

We present the design of an instrument that allows us to know how to perform the evaluation of knowledge in the written (theoretical) examinations of medical students. While this is an assessment of the summative evaluation design itself, which aims to measure student achievement in a quantitative way, then we worry about the quality of this design. It initially undertook a literature review, stopping primarily on three works: the taxonomy of Bloom's educational objectives, the revised version of Anderson, and the work of Case and Swanson on how to prepare questions for written evaluations in the area of basic science and clinics. After an analysis and discussion of these and other studies, we designed an instrument that requires three exercises: first to identify that particular course content, course or subject are considered essential, important and others which are complementary, and second to identify the distribution of questions on the test, and third analyzing the thinking skills that implicitly require students to use in solving the test. The overall analysis of these criteria will assess the consistency or inconsistency in the overall construction of the test.

We discuss how the design of a summative evaluation challenging, productive and rewarding as a result of careful planning, the student would require a more qualified mode to learn, so this exercise would also recognize the educational value simultaneously in students

**Keywords:** Meta-evaluation, written examinations, summative evaluation, formative evaluation

\*Jefe Departamento de Educación Médica y planificación (DEMYP)

\*\* Director Académico Facultad de Medicina UMSS,

\*\*\* Responsable de la Unidad de Educación Continua

## INTRODUCCIÓN

Evaluar la evaluación (metaevaluación) tiene por objeto, mejorar la calidad de la misma, generando un proceso de reflexión (autoevaluación), este ejercicio permite diagnosticar fortalezas y debilidades y como resultado de esta valoración, tomar decisiones para asegurar la calidad de la evaluación.

La evaluación como tal es un componente amplio, por lo que pretendemos focalizar la atención, en el diseño de un instrumento que permita conocer como se realiza la evaluación de los conocimientos en los exámenes escritos que se aplican a los estudiantes de medicina.

Es importante evaluar este componente, porque este es uno de los núcleos centrales del proceso de enseñanza y aprendizaje. La forma de la evaluación es uno de los factores principales que determina la forma en lo que los estudiantes estudian y por lo tanto la forma en la que aprenden o las razones por las que no lo hacen; por lo que se considera importante conocer cuáles son las características de los exámenes escritos que se realizan para evaluar los conocimientos en los estudiantes de lo que se ha venido a llamar "la teoría" en la Facultad de Medicina. Los exámenes de tipo test que se aplican actualmente, han sido la respuesta a la necesidad de evaluar un elevado número de estudiantes, (masificación) especialmente en los primeros cursos de la carrera.

Si bien se trata de una evaluación sumativa cuyo objetivo es medir el aprovechamiento de los alumnos de forma cuantitativa como el sistema actual lo hace necesario, entonces deberemos preocuparnos de la calidad de esta evaluación, es decir que y como se está evaluando y para qué. Si el test se ha diseñado solo para evaluar la capacidad de memorizar, que se conoce como conocimientos declarativos, entonces los estudiantes solo se preocuparán de memorizar. Si el objetivo es promover el pensamiento crítico en los estudiantes, entonces se debería diseñar un examen que estimule en los estudiantes las habilidades de pensamiento superior, de modo que el estudiante tenga la necesidad de aprender de un modo más cualificado que le haga necesario analizar, evaluar, discriminar, crear; dicho de otro modo pensar en profundidad<sup>1</sup>. En este caso si bien seguiríamos refiriéndonos a una evaluación esencialmente

sumativa, por su necesidad de poner una nota al estudiante; su diseño desafiante, productivo y enriquecedor, como resultado de un ejercicio cuidadoso de planificación, permitiría reconocer simultáneamente también su valor formativo en los estudiantes.

Es importante analizar que tan coherentes somos en los procesos de enseñanza y aprendizaje. Lo cierto es que no deberíamos enseñar de una forma y evaluar de otra totalmente diferente<sup>2</sup>, reflexionar sobre estos aspectos probablemente también nos permita comprender una de las razones por las cuales los estudiantes asisten o no a las actividades teóricas o solo algunas de ellas.

## MATERIAL Y MÉTODOS

Para diseñar esta herramienta a la que denominaremos el Instrumento Metaevaluator de la Teoría (IMT), se procedió inicialmente a una revisión bibliográfica deteniéndonos fundamentalmente en tres trabajos: el primero es el trabajo de Benjamín Bloom realizado en 1956, que examina el dominio cognitivo<sup>3</sup>. La idea central de este trabajo es identificar aquello que los educadores quieren que los estudiantes sepan, categorizando y ordenando habilidades de pensamiento y objetivos. La taxonomía en general sigue el proceso del pensamiento empleando progresivamente habilidades de pensamiento que avanzan desde las habilidades de pensamiento inferior hacia las de orden superior. El segundo documento revisado es el de Lorin Anderson -un antiguo estudiante de Bloom que junto a sus colegas publicaron en 1999 una versión actualizada de la Taxonomía de Bloom<sup>4</sup>, en la cual incluyen un amplio rango de factores que ejercen impacto en la enseñanza y el aprendizaje. Esta taxonomía revisada intenta corregir algunos de los problemas detectados en la taxonomía original. A diferencia de la versión de 1956, la taxonomía revisada distingue entre el saber qué (el contenido del pensamiento) y el saber cómo (los procedimientos utilizados en la resolución de problemas). La tercera revisión realizada fue el trabajo elaborado por la National Board of Medical Examiners, titulado "Como elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas"<sup>5</sup>.

Estas tres fuentes proporcionaron varias ideas sobre las cuales se diseñó el IMT, que se espera pueda servir para efectuar el diagnóstico de nuestros exámenes reforzando sus fortalezas y corrigiendo los problemas existentes en

ellos.

Para esta tarea se seleccionaron algunos criterios que debían ser incluidos en el mencionado instrumento.

El primero se relaciona con el hecho de que pretender que los estudiantes aprendan determinados contenidos no es un ejercicio del todo o nada, se trata más bien de una cuestión de grado, de tal modo de que es necesario identificar que contenidos de una asignatura, materia o curso se consideran imprescindibles, cuales importantes y que otros son complementarios. El IMT deberá presentar una muestra sobre cómo se tendrían que construir los criterios para rotular con propiedad estos contenidos, y cuantificarlos.

En segundo lugar si este ejercicio se ha realizado con propiedad entonces tendrá que verse reflejado en la distribución de las preguntas en el examen, por lo tanto el IMT tendría que tener la capacidad de identificar esta distribución y evaluar en qué medida el test es adecuado o no a este propósito.

El tercer criterio se relaciona con el análisis de las habilidades de pensamiento que implícitamente se exige utilizar a los estudiantes en la resolución del examen, test o prueba. El IMT deberá presentar una sugerencia sobre el empleo de determinados verbos en la construcción de las preguntas que respondan a las necesidades de su diseño, con algunos ejemplos en su redacción.

El análisis global de estos criterios permitirá evaluar la coherencia o no en la construcción global de la prueba.

## RESULTADOS

A continuación se presenta como está diseñado el IMT.

### INSTRUMENTO META-EVALUADOR DE LOS EXAMENES TEÓRICOS (IMT)

I. Clasifique y cuantifique los temas de su materia o asignatura que está evaluando en su test.

En la tabla No 1 se especifican los criterios y definiciones de los contenidos de estas definiciones, las cuales pueden ser reformulados o complementados de acuerdo a sus necesidades. Lo importante es que usted especifique y procedimentalice.

Tabla No 1 Definiendo los Contenidos

CONTENIDOS	CRITERIOS DE LA DEFINICIÓN
<b>Imprescindible</b>	Se refiere a contenidos relacionados con un problema de salud importante en la región o en el país. Generalmente se encuentra entre las diez principales causas de consulta externa (Ej. Diarrea aguda), de hospitalización (Ej. Abdomen agudo), o representan los principales problemas de salud de la comunidad (Ej. Anemia ferropénica). Si se trata de una materia, son contenidos centrales (Ej. En Anatomía el sistema cardiovascular, en Farmacología antibióticos de primera línea). Se recomienda que los contenidos deben ser visto en su integridad, es decir de la prevención a la rehabilitación. Generalmente no se necesita un conocimiento especializado para su abordaje. Esto son los temas que un medico general debe dominar
<b>Importante</b>	El impacto de este problema de salud sobre la sociedad la hace importante por: su alta mortalidad (Ej. VIH), secuelas invalidantes (Ej. Quemaduras), alteraciones de la fertilidad (Ej. CA de cérvix), de las funciones mentales (Ej. Esquizofrenia), etc. Si se trata de una materia sus contenidos son relevantes (Ej. En farmacología antibióticos de segunda línea). Es posible que para la atención de este tipo de problemas se necesite el concurso de un especialista; sin embargo el medico general debe tener un conocimiento sólido, es decir debe comprender claramente la problemática para poder actuar con oportunidad
<b>Complementario</b>	Estos problemas de salud requieren habitualmente el manejo de un profesional de alta especialidad para su diagnóstico, intervención y resolución. Si se trata de una materia sus contenidos son de valor informativo (Ej. En Farmacología nuevos antibióticos). Estos son los temas que un médico debe reconocer y definir.

#### I.1. Cuantifique los temas empleando la Tabla No 2

Tabla No 2. Cuantifique los contenidos identificados

Tema	Número de temas	Porcentaje
Imprescindible		
Importante		
Complementario		
Total de temas		100 %

#### I.2 Evalúe la distribución de los temas en su test.

En función a los resultados obtenidos evalúe su test, para ello se propone emplear los criterios usados en la Tabla No 3 en lo que se refiere a la distribución de temas marcando con una "X", la casilla que lo identifique.

Tabla No 3 Resultado de la distribución

Resultado de la distribución	Criterios
Adecuada -----	Los temas imprescindibles representan el 50 % del examen, el otro 50 % está conformado por los temas importantes y los complementarios, sin embargo se observa un predominio de los temas importantes en esta otra mitad
Regular -----	Los temas imprescindibles representan el 50 % del examen, el otro 50 % está conformado por los temas importantes y los complementarios, sin embargo se observa un predominio de los temas complementarios en esta otra mitad.
Inadecuada -----	Los temas imprescindibles representan menos del 50 % independientemente de cómo estén distribuidos los otros temas.

#### I. Identifique las habilidades de pensamiento evaluados

1. Preguntas para evaluar como se recoge información (recordar información relevante) son aquellas que empleen los siguiente verbos como indicadores: reconocer (Ej. a que mecanismo de resistencia corresponde el enunciado mencionado, cualquier pregunta de falso / verdadero, de selección, o las preguntas que emplean "excepto"), definir (Ej. Defina la neumonía atípica), listar, nombrar, identificar, reproducir y citar

2. Preguntas para evaluar la comprensión (convertir la información en conocimiento habilidad, habilidad de construir significado) emplean los siguiente verbos como indicadores: diferenciar (Ej. Cuál es la diferencia en la fisiopatología entre el desnutrido emaciado y el edematoso), clasificar (Ej. Cómo se clasifica la hiponatremia) interpretar (Ej. Qué significa que el resultado es estadísticamente significativo pero clínicamente no), distinguir, asociar, estima, resumir, contrastar, comparar, ejemplificar, explicar.

3. Preguntas para evaluar la aplicación (hacer uso de la información) utilizan los siguiente verbos como indicadores: aplicar (Ej. Aplique un protocolo de diagnóstico y / o tratamiento), completar (A la tabla de dos columnas donde se muestran las diferencias agregue una tercera columna con las palabras claves donde se expliquen las diferencias), modificar (corregir las incoherencias del texto), cambiar (Ej. Cambie las conductas siguiendo las guías internacionales basadas en la evidencia), elaborar, implementar, descubrir.

4. Preguntas para evaluar la capacidad de analizar (descomponer y relacionar) son aquellas que empleen los siguientes verbos como indicadores: separar (Por Ej. Lo positivo de lo negativo), ordenar (Ej. organice la información en orden de prioridades), conectar (Ej. Elabore los conectores en este mapa conceptual sobre la diarrea adicionando la información relevante), comparar, seleccionar, inferir, clasificar, categorizar, separar, atribuir.

5. Preguntas para evaluar la capacidad de evaluar propiamente dicha (comprobar, juzgar y criticar el resultado) utilizan los siguientes verbos como indicadores: establecer gradación (Ej. Explique el tratamiento gradual que esta-

blecería en las diferentes formas de presentación del Asma), comprobar (Ej. Analice el artículo y señale las contradicciones), medir (Ej. Cuál es la sensibilidad y especificidad del test), recomendar, juzgar, explicar, valorar, criticar, justificar, argumentar, predecir, hipotetizar.

6. Preguntas para evaluar la capacidad de crear (incorporar cosas para hacer algo nuevo) emplean los siguientes verbos como indicadores: integrar (Ej. Integre lo que se conoce hasta el momento sobre la patología X y construya una nueva teoría), reordenar (Ej. Si usted tendría que ordenar las conclusiones presentadas como las replantearía), substituir (Ej. Después de analizar las metaevaluaciones efectuadas sobre el tratamiento "X" que tratamiento substituiría y con qué), diseñar (Ej. Diseñe un protocolo de estudio), generar, componer, planear, reescribir.

7. Preguntas para evaluar actitudes (saber ser). Son aquellas que empleen los siguientes verbos como indicadores: plantear posturas (Ej. Cuál es su posición en el problema "X"), opinar (Ej. Qué opina sobre la posición "X"), conciliar (Ej. Cómo se podría conciliar entre las posturas encontradas), analizar dilemas (Ej. De los tres pacientes que llegaron al mismo tiempo a UTI cuál paciente usará el único ventilador disponible).

II.1 Determine el número y porcentaje del tipo de preguntas en los temas imprescindibles.

<b>Las habilidades de pensamiento inferior</b> (saber) se analizan con las tres primeras preguntas es decir recordar, comprender y aplicar.
<b>Las habilidades de pensamiento superior</b> (saber pensar, pensar en profundidad) se evalúan con las preguntas 4, 5, 6; Es decir: analizar, sintetizar y crear. La evaluación de las actitudes (Saber ser) se evalúa con la pregunta No 7 y la consideramos una habilidad superior.

Consideramos que esta actividad le servirá para iniciar la evaluación de la coherencia global del test, con este propósito por favor le solicitamos llenar las tablas No 4, No 5 y No 6.

Tabla No 4 Número y porcentaje de preguntas en los temas imprescindibles

	Recordar	Comprender	Aplicar	Analizar	Sintetizar	Crear	Saber ser
Número							
Porcentaje							
Total							

Tabla No 5 Número y porcentaje de preguntas en los temas importantes.

	Recordar	Comprender	Aplicar	Anali- zar	Sinteti- zar	Crear	Saber ser
Número							
Porcen- taje							
Total							

Tabla No 6 Número y porcentaje de preguntas en los temas complementarios.

	Recordar	Comprender	Aplicar	Anali- zar	Sinteti- zar	Crear	Saber ser
Número							
Porcen- taje							
Total							

### II.2 Analice la coherencia global del texto.

La coherencia global se refiere a que el test es percibido como una unidad de sentido y no como un conjunto de preguntas inconexas, la formulación de las preguntas son el resultado de una planificación cuidadosa. Para facilitar esta evaluación por favor analice y responda la Tabla No 7.

Tabla No 7 Evaluación de la Coherencia global del test.

Criterios de coherencia	Indicadores
1. En los temas imprescindibles, las preguntas para evaluar habilidades de pensamiento superior deberían representar más del 50%, y las preguntas para evaluar las habilidades de pensamiento inferior no deberían sobrepasar el 30%.	Sí al menos dos criterios están presentes y uno de ellos es el que se refiere a los temas imprescindibles, marque coherente "X"  -----
2. En los temas importantes las preguntas para evaluar habilidades de pensamiento superior e inferior están distribuidas en un porcentaje cercano al 50%, aunque es deseable que predominen las habilidades de pensamiento superior	Cualquier otra posibilidad que no sea la referida arriba, marque incoherente "X"  -----
3. En los temas complementarios Las habilidades de pensamiento inferior representan la mayoría. Las preguntas para evaluar el pensamiento superior no deberían sobrepasar el 30%	-----

## DISCUSIÓN

El IMT es una propuesta para analizar la calidad de los exámenes que se elaboran para evaluar los conocimientos de los estudiantes en exámenes escritos; sin embargo para

su implementación es recomendable llegar a acuerdos internos en el departamento, curso, asignatura, etc. Lo ideal es que los criterios para seleccionar los contenidos se definan por consenso, considerando las particularidades de cada situación, de igual manera se deberían definir los criterios para rotular cada contenido, y la forma cómo se van a distribuir las preguntas analizando la coherencia global de la prueba. Este ejercicio nos proporcionará el lenguaje común para planificar cualquier actividad posterior sobre el respaldo de una actividad participativa que permita empoderarse del planteamiento.

El IMT puede ser empleado con diferentes propósitos evaluativos, sin embargo en el fondo la razón de ser de la evaluación es servir a la acción educativa, que debe entenderse desde una visión integral; situación que al docente le debería preocupar antes de cualquier otra consideración. Como dice Stenhouse (1984), "para evaluar hay que comprender. La evaluación no es cuestión de éxito o fracaso. El profesor debería ser un crítico, un estratega y no un simple calificador"<sup>4</sup>.

Es probable que el IMT le proporcione un procedimiento de planificación de una parte del examen que indudablemente mejorará la calidad del mismo; sin embargo no pretende ser suficiente, será necesario revisar también otros documentos que concentren su interés en la redacción de preguntas de opción múltiple de alta calidad como el documento del National Board of Medical Examiners<sup>5</sup>. Finalmente para hacer más completo el ejercicio se debería analizar en qué medida el diseño del test evaluado está respondiendo al Perfil Profesional de la Facultad de Medicina en que se implementa esta prueba.

Diseñar un test para evaluar los conocimientos adquiridos por los estudiantes en las actividades teóricas, es un proceso complejo, en tanto que no solo pretende mostrar lo que los estudiantes "saben", sino también refleja a su diseñador como estratega y como educador y su responsabilidad procurando que la evaluación sirva también para aprender y formar, el principal desafío.

El IMT no pretende ser una guía de que es lo que se "debe hacer" es más bien un camino entre varios posibles de lo que se podría hacer para mejorar las acciones de evaluación. Sea cual fuere el que decidamos, la evaluación es un ejercicio que requiere una planificación cuidadosa por su

diseñador, reflejando que detrás del diseño, existe la intención de estimular en los estudiantes las habilidades de pensamiento superior, condiciones centrales de un pensador crítico<sup>1</sup>.

Se espera que este instrumento propicie un espacio de diálogo entre los actores involucrados, reconociendo la importancia de la evaluación en los procesos de enseñanza – aprendizaje, situación que nos permitirá iniciar procesos de mejora de la calidad de la misma.

## **AGRADECIMIENTOS**

---

Al Lic. David Aranibar y Lic. Franklin Mayan Echalar, miembros del departamento de Educación Médica y Planificación quienes colaboraron en la realización del trabajo.

## **BIBLIOGRAFÍA**

---

1. Cottrell S. Critical Thinkings Skills. Developing Effective Analysis and Argument. New York: Palgrave, 2005
2. Díaz F, Hernández G. Estrategias docentes para un aprendizaje significativo. Pensamiento crítico. 2 ed. Mexico: Mc Graw Hill, 2001
3. Bloom B, ed. Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. New York ; Toronto: Longmans, Green; 1956
4. Anderson L, Krathwohl D, ed. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Abridged Edition; 2001
5. Case S, Swanson D. Como elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas 3 ed. Philadelphia: National Board of Medical Examiners, 1998
6. Molnar G. Reflexiones en evaluación educativa. Revista Electrónica de Investigación y Evaluación Educativa 1997; 3 (1) - ISSN 1134-4032 - D.L. SE-1138-94. Disponible en: <http://www.chasque.net/gamolnar/evaluacion%20educativa/evaluacion.04.html>