



REPÚBLICA DE CUBA
ESCUELA NACIONAL DE SALUD PÚBLICA

MAESTRÍA EN EDUCACIÓN MÉDICA

Los exámenes teóricos y los criterios de calificación. Sus dificultades. Una experiencia en la Facultad de Ciencias Médicas “Dr. Enrique Cabrera”. Curso 2003-2004.

Autor: Nelson Tamargo Rodríguez

Tutora: Marta Pernas Gómez
Especialista de Segundo Grado en Histología.
Profesora Auxiliar.
Master en Ciencias de Educación Médica

Asesor: Dr C. Jorge Bacallao Gallestey
Master en Ciencias de Educación Médica

TRABAJO PARA OPTAR POR EL TÍTULO DE MASTER EN EDUCACIÓN MÉDICA

Ciudad de La Habana, 2005.

AGRADECIMIENTOS.

Agradezco a todos los que me estimularon y colaboraron en culminar la investigación, muy en particular a mi tutora Dra. Marta Pernas Gómez y a nuestro asesor Dr. Jorge Bacallao Gallestey.

RESUMEN

Con el propósito de caracterizar el proceso de evaluación teórica en el área clínica de la Facultad de Ciencias Médicas Dr. Enrique Cabrera se hizo un análisis de la correspondencia entre los objetivos finales de las asignaturas y los contenidos evaluados en los exámenes teóricos escritos aplicados en el curso 03-04 así como una valoración de la calidad metodológica de las preguntas utilizadas en dichos exámenes, la determinación de la concordancia en los criterios de calificación de los profesores y de la variabilidad de los criterios profesoriales acerca de las calificaciones finales de los exámenes a partir de las calificaciones de las preguntas componentes.

Los resultados ponen de manifiesto que existió una adecuada correspondencia de los aspectos evaluados en los exámenes finales y los objetivos previstos en los programas de las asignaturas. Se identificaron diversos problemas metodológicos en la elaboración de las preguntas empleadas en los exámenes, particularmente focalizados en las preguntas de los tests objetivos de asociación, en las preguntas de selección múltiple y de respuesta alternativa.

Se constató que hay deficiente concordancia en los criterios de los diferentes profesores para la calificación de los exámenes y una considerable variabilidad en la emisión de la calificación final de los exámenes a partir de los resultados por preguntas, en lo que, en alguna medida está influyendo el diseño de los exámenes utilizados. Estos resultados ponen de manifiesto la necesidad de lograr un perfeccionamiento en la preparación de los profesores para enfrentar la evaluación, tanto en la elaboración de los instrumentos como en la calificación de los exámenes aplicados.

INDICE

CAPÍTULO I INTRODUCCIÓN	1
CAPÍTULO II MARCO TEÓRICO	6
CAPÍTULO III OBJETIVOS	31
CAPÍTULO IV DISEÑO METODOLÓGICO	32
CAPÍTULO V RESULTADOS Y DISCUSIÓN	36
CONCLUSIONES	46
RECOMENDACIONES	47
REFERENCIAS BIBLIOGRÁFICAS	48
BIBLIOGRAFÍA CONSULTADA	52
ANEXOS	

CAPÍTULO I: INTRODUCCIÓN

La evaluación del aprendizaje conforma un marco de referencia para la reflexión y toma de decisiones inherentes a toda labor de mejora de la enseñanza ⁽¹⁾. La evaluación del aprendizaje es una parte esencial del proceso docente educativo que nos posibilita su propia dirección, así como el control y la valoración del desarrollo de los modos de actuación que los estudiantes van adquiriendo, y permite verificar el grado en que se van alcanzando los objetivos generales de estudio propuestos ⁽²⁾. La adecuada evaluación de los estudiantes tiene gran pertinencia ya que de ello depende la calidad de los futuros profesionales formados en el sistema para dar respuestas a las demandas de la sociedad.

Antecedentes

El problema del control y evaluación del aprendizaje es muy amplio y a la vez complejo. La literatura pedagógica consultada hasta el momento refleja que es motivo de preocupación no sólo para los estudiosos del problema, sino para profesores de diversas partes del mundo ⁽³⁻⁸⁾.

Un análisis de esta cuestión la hace el español José Luis Rodríguez Diéguez ⁽³⁾ basado en algunos trabajos escritos sobre el tema y que se refieren a tres líneas fundamentales en la evaluación del aprendizaje.

La primera relacionada con los que defienden la importancia del diseño científico de los instrumentos que se emplean para evaluar. La segunda relacionada con el juicio que se forme el profesor por la observación diaria del trabajo del estudiante y

mediante la comparación de los resultados obtenidos con aquello que se había previsto conseguir y la tercera línea, que supone el acercamiento pleno a una caracterización de la evaluación como proceso de control de la calidad en la educación.

Una investigación realizada en la Universidad de Girona en 1999 ⁽⁹⁾, considera la evaluación del aprendizaje como una variable especialmente relevante y expresa los siguientes motivos entre los que conducen a profundizar en el estudio de esta dimensión:

- 1- El primero de ellos sustenta que las conclusiones del informe del Plan Nacional de Evaluación de la calidad de las Universidades (1997), pone de manifiesto las carencias de los sistemas de evaluación del aprendizaje y la insatisfacción de los alumnos respecto a los mismos. Entre los argumentos que conducen a estas conclusiones destacaríamos la ausencia de criterios objetivos y generales de evaluación, la falta de coherencia entre los criterios de evaluación de diferentes asignaturas de un mismo estudio, la falta de publicidad de dichos criterios, la centralización de los procesos de control sobre los exámenes y el desajuste entre los contenidos impartidos y los niveles exigidos.
- 2- En segundo lugar, la existencia de investigaciones que demuestran cómo el sistema de evaluación que utilizan los profesores para valorar el aprendizaje de los alumnos afecta a la calidad de dicho aprendizaje, llegando a considerar que las intervenciones para mejorar los procedimientos de aprendizaje de los alumnos y su nivel de calidad, deberían empezar por cambiar los sistemas de evaluación ⁽¹⁰⁻¹¹⁾.

3- Un tercer motivo se basa en la constatación de la existencia de diferentes modelos de evaluación, centrados casi exclusivamente en los criterios que aplica cada profesor ⁽¹²⁻¹³⁾.

Si a este eclecticismo se suma la falta de reflexión y trabajo común sobre el tema, el resultado es la coexistencia de demasiados modelos evaluativos que conducen a la confusión y desorientación de los estudiantes respecto a cómo afrontar la demanda que se les formula.

La doctora N.G.Talízina ⁽⁴⁾ planteó en un artículo escrito hace unos años que la objetividad y fundamentación del control y la evaluación dependen de las siguientes condiciones:

- Tener en cuenta la solidez del control.
- Controlar los conocimientos y habilidades que se expresan en los objetivos de enseñanza. El control debe presuponer los indicadores diferenciados en cada una de sus direcciones.
- Las respuestas dadas deberán corresponderse con los objetivos de la enseñanza propuestos y en consecuencia otorgar notas.
- La evaluación total dependerá de las normas establecidas, considerando el significado de la evaluación y características de la asimilación del contenido.

En la práctica por lo general, dice Talízina, cada profesor tiene sus normas para evaluar. Esta es la causa de que al aplicar ejercicios de control iguales las calificaciones recibidas sean diferentes y esta es la razón por la cual el problema de las normas para calificar y evaluar reviste gran actualidad. Un error muy frecuente cuando se ha asumido un sistema de evaluación cualitativo es realizar la suma de las evaluaciones recibidas por el alumno durante una etapa, lo que es llevado a un promedio aritmético.

Al respecto, la autora expresa que se viola la objetividad al tener en cuenta resultados que responden a lagunas del conocimiento existentes en un momento dado, pero que después son superados por el estudiante, sin embargo, de todas formas tienen un peso considerable en su calificación, cuestión que no es justa.

El proceso de la evaluación ha sido y continúa siendo uno de los problemas más interesantes y preocupantes para los docentes de la universidad moderna, en la esfera de la enseñanza.

En el campo específico de la formación médica, reconocidas figuras en el campo de la medicina latinoamericana, le han brindado especial atención a esta tarea.

Salas Perea plantea que en los análisis que se efectúan en los colectivos de asignatura y disciplinas generalmente no se incluyen qué preguntas deben responder los estudiantes, qué problemas han de ser capaces de resolver, cuántos y cómo, qué criterios e índices nos pueden mostrar y en qué grado, ni tampoco el logro de los objetivos propuestos para cada actividad docente, tema impartido y para la asignatura en general. En la práctica todo esto se traduce en preguntas improvisadas que no están acordes con los objetivos a evaluar, ejercicios y problemas que no se corresponden con los niveles de asimilación propuestos, ni se relacionan con el tipo de control empleado, en la aceptación de respuestas incompletas y en diferencias de criterios entre los profesores al calificar los instrumentos evaluativos, todo lo cual crea desconfianza entre los estudiantes sobre los resultados ⁽¹⁴⁾.

La evaluación está determinada por los objetivos, pero sin lugar a dudas su efectividad va a depender en gran medida del número, frecuencia y calidad de los controles que se apliquen, así como de la correcta y uniforme calificación que se realice de los resultados. Su carácter de continuidad permite la constante comprobación de los

resultados y poder valorar su tendencia, que permite conformar un criterio dinámico sobre el aprovechamiento del estudiante.

En la educación médica superior cubana, los objetivos educacionales constituyen la categoría rectora del proceso docente educativo, pero no hay dudas de que la evaluación influye en lo que enseñamos y determina de manera directa lo que los educandos aprenden ⁽²⁾.

Como se puede observar, la problemática de la evaluación tiene gran actualidad y vigencia en el ámbito universal.

Justificación:

En los Consejos Ampliados de la FEU de la Facultad de Ciencias Médicas Dr. Enrique Cabrera se expresa la insatisfacción por los estudiantes con la evaluación teórica escrita en el área clínica. Aproximadamente un 10% de los estudiantes evaluados en el área clínica solicitan revisión de su examen escrito. La insatisfacción de los mismos se relaciona con la calidad en la elaboración de los instrumentos, con la heterogeneidad en los criterios de calificación de los exámenes teóricos finales entre asignaturas y con el ajuste de los temarios con los objetivos correspondientes.

El problema científico radica en la necesidad de caracterizar la evaluación teórica en el área clínica reflejando sus insuficiencias

CAPÍTULO II: MARCO TEÓRICO

En los discursos de Fidel Castro que tienen relacionados con la educación médica superior se plantea “Pero no debemos tener temor, cualquier cosa, menos graduar a un médico que no tenga calidad”. y continúa afirmando que “todos los cuadros deben analizar los problemas relacionados con la evaluación, la necesidad de unificar criterios”⁽¹⁵⁾.

La correcta aplicación del sistema de evaluación del aprendizaje que tiene un carácter cualitativo integrador, se basa en un conjunto de condiciones que constituyen las premisas básicas que debe satisfacer, las que se encuentran estrechamente relacionadas entre sí y cuya inobservancia impide que se cumplan efectivamente sus funciones. Con fines exclusivamente didácticos las estructuramos en principios y exigencias⁽¹⁴⁾.

Los principios de la evaluación son:

- La objetividad.
- La sistematización y continuidad.
- El carácter sistémico.

Para que una evaluación sea objetiva es imprescindible que, a partir de los objetivos educacionales definidos, determine el dominio del contenido de la enseñanza y las cualidades que lo caracterizan, de acuerdo con el nivel de profundidad precisado y su correspondencia con el nivel de asimilación exigido; y que, a su vez, esta valoración se realice mediante la forma y el tipo de evaluación que asegure la validez del control. La objetividad de un control se desarrolla, también, a la vez que se incrementa el grado de

concordancia entre los examinadores sobre lo que constituye una “buena respuesta”
(16).

La sistematización y continuidad de la evaluación está dada porque el número de muestras (controles) del universo de los conocimientos, las habilidades y actitudes, y la regularidad con que se efectúen, posibiliten la generalización del criterio sobre el nivel de aprendizaje alcanzado. En este sentido es necesario puntualizar que el aprovechamiento del educando siempre hay que caracterizarlo con un criterio dinámico, evolutivo y activo, de su tendencia.

Determinar su frecuencia es un problema bastante complejo, pues tan dañino es que el número de muestras sea insuficiente y no posibilite llegar a un correcto criterio de la efectividad y calidad del aprovechamiento docente, como que sea excesivo, atiborre y recargue al educando y al profesor, lo que por otra parte hace que los controles pierdan su importancia y fundamento en el proceso enseñanza aprendizaje.

Es por ello, que su frecuencia debe ser tal que permita retroalimentar oportunamente al educando y al profesor, y también debe estar en dependencia del éxito que se vaya logrando en el educando y de la necesidad que tenga de ser sometido a nuevos controles.

El carácter sistémico está dado por la necesidad de concebirlo como un componente del proceso docente educativo y en estrecha interrelación e interdependencia con los restantes componentes: objetivo, contenido, método, formas y medios de enseñanza, así como en el papel activo que tienen que desarrollar tanto el profesor como el estudiante para su efectividad.

Las exigencias que, en primer orden, tiene que cumplir todo procedimiento y técnica evaluativa son: validez y confiabilidad.

La validez es la correspondencia entre lo que se pretende verificar por el instrumento, procedimiento o método y lo que realmente se mide. Ello abarca tanto las exigencias propias del control como la determinación de los índices valorativos que permiten la calificación ^(14,17,18). Es el grado de precisión con que el instrumento utilizado mide realmente lo que está destinado a medir. Se relaciona específicamente con la interpretación que puede hacerse de los resultados obtenidos. Es un concepto siempre específico ^(16,19).

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

La validez de contenido está en correspondencia con el sistema de conocimientos y habilidades que mide un instrumento de control en cantidad y calidad, sean los suficientes para definir el grado de aprovechamiento del educando, sobre la base de los objetivos propuestos. En general mide a la pregunta; ¿mide el instrumento realmente la materia y comportamiento que se desea medir?

La validez funcional está en correspondencia con la selección del tipo de actividad (acciones, tareas o problemas) que debemos verificar en el educando para evaluar, de acuerdo con los objetivos propuestos, sus modos de actuación.

La validez predictiva está en correspondencia con la medida en que los resultados en una disciplina, semestre o ciclo nos permiten predecir o no su éxito durante otras disciplinas, semestres o ciclos. ¿En qué medida, por ejemplo, los resultados obtenidos

por un estudiante de Propedéutica Clínica y Semiología pueden ayudar a predecir los resultados de Medicina Interna?

La confiabilidad es la estabilidad de los resultados de un control, ya sea al repetirlo o al ser calificado por diferentes profesores. Significa que hay constancia en los resultados obtenidos y que su generalización es representativa del grado de aprovechamiento alcanzado por el educando en el tipo de control realizado. Se refiere a la constancia de los resultados en el tiempo, según los tipos de preguntas y los criterios de los examinadores. Es un concepto estrictamente estadístico y se expresa por un coeficiente de fiabilidad o mediante el error tipo de las mediciones efectuadas.

Con el fin de elevar la confiabilidad de un instrumento evaluativo debemos contemplar:

- Aumentar la evaluación de los modos de actuación en la vida real o lo más parecido posible.
- Crear la similitud en las condiciones en que se realiza la evaluación.
- Mantener el mismo nivel de exigencia.
- Estandarizar los criterios para emitir una calificación.

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. Por ejemplo, un control puede ser confiable, porque al ser calificado por diferentes profesores se obtienen los mismos resultados; sin embargo, puede no tener validez, ya que el instrumento no mide los objetivos previstos. Es por ello que, en primer orden hay que garantizar la validez de todo instrumento evaluativo, para completar en segundo orden, su confiabilidad.

En la educación médica se realizan diferentes tipos de controles en diferentes etapas del proceso docente: control preliminar, durante el proceso docente y el diferido.

El control preliminar es de dos tipos fundamentalmente: los exámenes de ingreso donde se evalúa que un aspirante sea aceptado y los exámenes de diagnóstico que se realizan antes de comenzar una asignatura o materia.

El control durante el desarrollo del proceso enseñanza aprendizaje es aquel que se realiza en el transcurso de la impartición o al concluir una asignatura, ciclo, carrera o especialidad. Puede ser frecuente, parcial, final y el examen estatal.

El control diferido es aquel que se realiza algún tiempo después de concluida la enseñanza de una asignatura, ciclo, carrera o especialidad y tiene como objetivo comprobar la retención de los conocimientos, las habilidades y los modos de actuación adquiridos en ella. Los exámenes de certificación y recertificación son ejemplos de este tipo de control.

La evaluación formativa se desarrolla durante el proceso docente y posee las características siguientes ⁽¹⁶⁾.

- Tiene una función de retroalimentación, diagnóstica o de control.
- Se diseña para informar al educando sobre sus éxitos e insuficiencias y deficiencias.
- Mide el progreso alcanzado por el educando en un período dado, su tendencia.
- Permite ajustar las actividades docentes en concordancia con los resultados.
- Es muy útil para guiar al estudiante y ayudarlo a pedir consejo.
- Será tan frecuente como sea necesario.
- Nunca será empleada por el profesor para emitir un juicio certificativo.

La evaluación acumulativa se desarrolla al concluir una asignatura, ciclo, carrera o especialidad y tiene las características siguientes:

- Tiene una función certificativa
- Se diseña para proteger a la sociedad y prevenir la incompetencia profesional.

- Se utiliza para gradar a los educandos por sus méritos, justificar las decisiones a adoptar y fundamentar la entrega de un título o diploma.

La aplicación de las diferentes formas de evaluación se puede realizar mediante múltiples métodos, procedimientos y técnicas ^(17, 20,21).

Los principales métodos empleados en la realización de los controles son: teóricos, prácticos y teórico-prácticos. Los métodos teóricos se emplean para evaluar los conocimientos y las habilidades intelectuales, mientras que los prácticos se emplean para evaluar las habilidades manuales o destrezas y los modos de actuación. Los métodos teóricos escritos son mundialmente los de mayor uso, los más económicos y los de mayor flexibilidad técnica.

Existen, básicamente, dos tipos de preguntas en los exámenes escritos: los de respuesta abierta y los de respuesta cerrada. En el primero, el candidato es instruido para que formule su respuesta; mientras que en el segundo tipo, al educando se le brindan varias posibles respuestas, de las que debe seleccionar una o varias, de acuerdo con su criterio específico y el instructivo que acompaña la pregunta.

Las preguntas de respuesta abierta se conocen comúnmente como de ensayo, y son más subjetivas que las otras. Este tipo incluye, también, las preguntas de respuestas cortas y las de ensayo modificado, y generalmente son de carácter reproductivo.

Las preguntas de respuesta cerrada se conocen como de tipo test, son más objetivas y posibilitan evaluar eficientemente –con amplio margen- el sistema de conocimientos, así como las habilidades intelectuales en la solución de los problemas de salud, incluso complejos ⁽²²⁾.

Todo instrumento evaluativo debe reunir las características generales siguientes:

- Su contenido debe estar directamente relacionado con los objetivos educacionales.

- Ser realista y práctico.
- Tener validez y confiabilidad.
- Ser objetivo y factible de realizar.
- Debe comprender los contenidos importantes y útiles (contenidos claves).
- Ser complejo, pero lo más breve posible.
- Ser preciso, pero claro en redacción.

EXAMEN TIPO ENSAYO

También es conocido como examen de ensayo largo o de desarrollo. Permite evaluar la información que posee el educando, así como sus habilidades para organizar las ideas, planear una respuesta y expresarla lúcidamente.

Presenta, dentro de sus problemas, que las preguntas sólo pueden abarcar un pequeño número de áreas, son relativamente generales y suponen respuestas amplias o de desarrollo. Los estudiantes consumen la mayor parte del tiempo pensando y escribiendo estas respuestas.

Son fáciles de confeccionar, pero su calificación consume una extraordinaria cantidad de tiempo.

En este tipo de examen no hay consistencia en los criterios de calificación empleados por los examinadores y la correlación de las calificaciones es baja entre ellos.

En general, la retroalimentación que proporcionan es pobre o nula, ya que cuando se les devuelven los exámenes a los estudiantes contienen pocos comentarios.

Sus resultados son muy difíciles de analizar, y la información obtenida por lo general no es medible.

La distribución de las calificaciones está fuertemente afectada por el subjetivismo de los examinadores, y sus rangos pueden variar en unos casos del 45 al 75 % y, en otros,

entre 0 y 100%. (2) Algunos califican estableciendo comparaciones entre las respuestas de diferentes estudiantes; mientras que otros los comparan con su nivel de conocimientos como especialistas. Las normas de calificación varían entre los grupos de examinadores ⁽²⁾.

En general la calificación de este tipo de examen es subjetiva y poco confiable.

VENTAJAS ^(16,23).

- Posibilita que el candidato exponga sus conocimientos, ordene sus ideas y demuestre su capacidad para expresarse en su idioma.

DESVENTAJAS ^(16,23).

- Limita sensiblemente el número de áreas a explorar.
- La objetividad es difícil de conseguir.
- Proporciona escasa retroalimentación al educando y al claustro.
- Su calificación consume demasiado tiempo, es subjetiva y sus resultados son poco confiables.

Este tipo de examen se debe emplear cuando: ⁽²⁴⁾

- El grupo de estudiantes sea pequeño.
- Los objetivos del control evalúen habilidades de autoexpresión escrita.
- Los examinadores se sientan satisfechos con evaluar, con pocas preguntas, una muestra de los contenidos que deben ser evaluados.
- Los examinadores tengan experiencia.
- Se considere que es más importante evaluar la libre expresión del candidato, que la objetividad y confiabilidad de una evaluación.
- Se formulen preguntas sobre una base problémica, que permita explorar su nivel de comprensión y/o su capacidad de solución.

- Para hacerlo más confiable es necesario incrementar el número de las preguntas/áreas, con una precisa definición de las posibles respuestas con sus correspondientes calificaciones.
- Tiene validez sólo para evaluar habilidades cognoscitivas. No son útiles para evaluar habilidades clínicas ni actitudes profesionales.

Entre las orientaciones para perfeccionar la confección y calificación de las preguntas tipo ensayo se encuentran: ⁽²⁾

- Escriba cada pregunta de modo que la tarea esté bien definida para cada estudiante.
- Utilice preguntas problémicas, donde le brinde al estudiante toda la información necesaria.
- Prepare, con antelación, el esquema de calificación con su clave bien detallada para cada tipo de respuesta.
- Muestree unas pocas respuestas de diferentes estudiantes, antes de comenzar a calificar cada pregunta.
- Califique una misma pregunta en todos los estudiantes, antes de pasar a otra.
- Ponga, como mínimo, a dos examinadores a calificar cada respuesta de un mismo estudiante y promedie sus calificaciones en cada pregunta.
- Brinde a cada estudiante, en cada pregunta, las razones de la calificación otorgada.
- Elimine el sesgo que supone una pobre presentación del examen.
- Oculte la identidad de los estudiantes.
- Elimine todo elemento que distraiga la atención del examinador.
- Garantice el tiempo necesario para la calificación del examen.
- Cada examinador calificará de modo independiente.

EXAMEN TIPO ENSAYO RESPUESTAS CORTAS

Las desventajas de las preguntas de tipo ensayo pueden reducirse, si se limitan las preguntas sólo a aquellas que tienen respuestas libres pero cortas ⁽²³⁾.

VENTAJAS

- Las preguntas son fáciles de confeccionar.
- Las áreas de materiales a examinar se incrementan, al aumentar el número de preguntas.
- La objetividad y confiabilidad de las calificaciones se incrementan considerablemente.
- Los examinadores pueden ponerse fácilmente de acuerdo en las posibles respuestas a aceptar como correctas y en cómo calificarlas.

DESVENTAJAS

- La amplitud de las áreas/materias a examinar aún es limitada.
- La calificación todavía consume mucho tiempo y persiste el juicio subjetivo alrededor del límite de lo correcto/incorrecto de las respuestas emitidas por los educandos.
- Las respuestas solicitadas son frecuentemente reproductivas más que de razonamiento y discusión; las preguntas pueden ser triviales o insignificantes, en relación con los objetivos propuestos.

Entre las recomendaciones para los exámenes de tipo ensayo de respuestas cortas se encuentran:

- Amplíe el número de preguntas a realizar, con el fin de incrementar el número de áreas materias a evaluar.

- Formule las preguntas de tipo problémicas, en correspondencia con los objetivos educativos a evaluar.
- Elabore las claves de respuesta bien detalladas con sus correspondientes calificaciones.
- Anote, en detalle, los errores e insuficiencias de cada respuesta sobre las cuales se basan las calificaciones emitidas, de forma que le sirva de retroalimentación al educando y al claustro.
- Cumpla con las orientaciones brindadas para las preguntas de tipo ensayo.

EXAMEN TIPO ENSAYO MODIFICADO

Es un tipo de examen cuyo formato se basa en un problema o situación de salud, donde los elementos son presentados en etapas sucesivas.

Se inicia con la descripción, por ejemplo, de la situación clínica, la cual es seguida de varias preguntas relacionadas con la interpretación de los síntomas y signos y qué conductas pudieran ser adoptadas.

Respondidas dichas preguntas, se le brinda al educando determinada información adicional que contiene datos significativos y otros de poca importancia, en relación con la enfermedad o con resultados complementarios sobre el paciente, seguida a continuación de otra batería de preguntas y así sucesivamente.

Por lo tanto, la técnica consiste en la presentación a partir de una situación problemática inicial de una serie sucesiva de etapas/problemas que constan de una batería de preguntas abiertas/cerradas cada una, acerca del desarrollo de la situación, donde las

respuestas en cada etapa son relativamente libres, ya que no son sugeridas al educando.

Las preguntas permiten explorar, entre otros aspectos la identificación de los factores claves que influyen en la toma de una decisión y sus actitudes correspondientes, en cuanto su conducta ante el trabajo profesional diario. Las preguntas no contienen aspectos explícitamente señalados. Se espera que los estudiantes enfrenten el problema aplicando lo que ellos hayan aprendido. También, es posible evaluar no sólo el resultado, en término de acciones a realizar, sino también la dirección del proceso de dichas acciones ^(25,26).

Estas preguntas llevan respuestas más completas de lo que aparentan a primera vista. Tales respuestas van a depender de la actitud, el nivel de entrenamiento y la experiencia del examinado.

Brindan la oportunidad de que los estudiantes demuestren su competencia y apliquen sus conocimientos en la solución de un problema de salud, cuando el uso de pacientes reales o simulados sean inapropiados. Un amplio rango de competencias puede ser muestreado en el período de 1 a 3 horas.

Aunque las preguntas, las claves de calificación y las respuestas de los educandos sean breves, su diseño y calificación consumen tiempo.

Este tipo de preguntas son fácilmente calificadas, en dependencia de las claves de las calificaciones elaboradas con anterioridad. Ello abarcará cada pregunta en particular y el examen en su totalidad.

VENTAJAS ⁽²⁾

- El empleo de descripciones clínicas incrementa la validez del examen, para medir competencias intelectuales.

- Las conductas a realizar pueden evaluarse ante un problema no diferenciado o general.
- Los profesores pueden identificar las áreas/materias débiles en cada estudiante y en cada grupo.
- La conducta a seguir puede ser evaluada por medio de las diferentes fases, en cuanto a diagnóstico, investigaciones, tratamiento y consejos al paciente.
- Con un adecuado diseño, se puede incrementar la validez del examen.

DESVENTAJAS

- Cada caso problema requiere tiempo para la preparación y realización del examen.
- Cada caso problema representa un examen en si mismo, y la información que se brinda al educando es limitada, lo que pudiera ser insuficiente y limitarle su interpretación y respuesta.
- Su calificación requiere tiempo y juicio profesional. Los examinadores deben sopesar cada pregunta y ponerse de acuerdo previamente, ante las posibles respuestas y valorar cuales son aceptadas como correctas.
- Los candidatos presentan dificultad en la distribución del tiempo entre las diferentes etapas de cada caso problema.

EXAMEN TIPO TEST OBJETIVO

Una forma alternativa de exámenes escritos está dada por el empleo de los test o exámenes objetivos, los cuales se han utilizado ampliamente en las últimas décadas (7,18,19,27,28) aun cuando algunos autores ⁽²⁹⁾ no son partidarios de su uso.

En un examen de este tipo, las respuestas de los estudiantes son calificadas objetivamente y son eliminados los problemas subjetivos que pueden encontrarse en los exámenes escritos de tipo ensayo.

Ciertas objeciones en torno a este formato de examen se relacionan con el hecho de que algunos consideran que las pruebas objetivas no miden la acción reflexiva del educando, sino meramente la memorización. Esto, en realidad no es una deficiencia del test, sino del que lo constituye, quien por falta de un planeamiento adecuado, no considera la diversidad de comportamientos susceptibles de una medición objetiva.

Un elemento esencial del planeamiento de un instrumento evaluativo se refiere a la precisión de los objetivos a medir, a partir de los cuales se definen las áreas a examinar, el énfasis que debe darse a cada una de ellas e, incluso, como consecuencia de esta planeación, pueden derivarse los tipos de ítems a construir ⁽³⁰⁾.

Un buen examen debe tener validez y confiabilidad. En este sentido, los test objetivos pueden presentar una mayor validez al permitir explorar una muestra mayor de los contenidos del programa. Por otra parte, pueden tener una mayor confiabilidad, ya que al establecerse una clave de respuesta unívoca se reduce considerablemente el efecto de "halo", la diversidad de criterios entre los que califican, así como su cansancio físico y mental.

VENTAJAS: ^(16, 24, 30)

- Asegura objetividad y, en general, propicia el aumento de la confiabilidad y la validez del examen.
- Incrementa la gama y variedad de los aspectos que se evalúan, en poco tiempo y con mínima supervisión.

- Posibilita la facilidad y rapidez de su calificación, lo que permite divulgar con prontitud las calificaciones obtenidas.
- Suministra retroalimentación, tanto al profesor como a los educandos, y evidencia los puntos débiles.
- Disminuye el costo de su aplicación, cuando el número de candidatos a evaluar es alto.
- Pueden almacenarse bancos de preguntas.
- El análisis estadístico es relativamente sencillo y suministra información confiable.

DESVENTAJAS

- Su preparación exige mucho tiempo, competencia y práctica, con el fin de evitar preguntas ambiguas y arbitrarias. Esta dificultad se incrementa cuando se desea comprobar aspectos complejos de interpretación o solución de problemas, más que los eminentemente factuales.
- Proporciona a los candidatos informaciones e indicaciones que no dispondrían en una situación real.
- Resulta costoso, cuando los educandos a examinar son pocos.
- Implica adoptar determinadas condiciones para tener en cuenta las respuestas correctas que pudieran ser adivinadas al azar.
- Exige el entrenamiento previo de los educandos.

Con el fin de elaborar un examen objetivo, debemos tener en cuenta los siguientes aspectos: ⁽²⁾

- Seleccionar una muestra de los aspectos que constituirán los ítems individuales.
- Definir cuáles son las técnicas o los tipos de test a emplear.

- Cómo escribir y presentar cada ítem.
- Cómo ordenar los ítem y cómo administrar el examen.
- Cómo evaluar su ejecución, cómo calificarlos y cómo interpretar los resultados.
- Cómo determinar su validez.

En épocas tan lejanas como los años 30 ⁽³¹⁾ se identificaban entre 40 y 50 variaciones en los formatos de los exámenes objetivos. Por las características de este trabajo, sólo enfatizaremos en los tres tipos básicos, que son, a su vez, los de mayor uso ^(16, 24, 30,32).

- 1- Ítem de respuesta alternativa (verdadero o falso).
- 2- Ítem de selección múltiple (Multiple Choice Questions) de la literatura anglosajona.
- 3- Ítem de asociación.

Cada uno de estos formatos tiene sus características particulares, pero podemos considerar algunas sugerencias generales para su confección:

- Desarrollar los ítems a partir de ideas relevantes y evitar trivialidades.
- Seleccionar temas o ideas que posibiliten la elaboración de ítems con el máximo poder de discriminación.
- Elaborar ítems que midan objetivos importantes.
- Presentar el ítem con la mayor claridad posible.
- Evitar la inclusión de elementos no funcionales.
- Evitar la construcción de ítems basados en elementos demasiados específicos.
- Adaptar el nivel de dificultad del ítem al nivel medio del grupo.
- Evitar la fraseología estereotipada.

Evitar la inclusión de elementos que pudieran sugerir respuesta, tales como:

- Construcción gramatical.
- Alternativa más larga y mejor elaborada como opción a seleccionar.
- Colocar la respuesta correcta en posición preferente o según un ritmo constante.
- Ítem interrelacionados entre sí.
- Utilización de determinadores específicos: todos, ninguno, nunca, siempre, etcétera.
- Presentación de una o más alternativas con el mismo significado.

ITEM DE RESPUESTA ALTERNATIVA (VERDADERO O FALSO)

Probablemente sea uno de los formatos más utilizados, debido a su preparación relativamente más fácil. Ello ha llevado a un real abuso su empleo y su uso debe privilegiarse en los casos en que sólo hay dos alternativas plausibles como respuesta, o en situaciones en las que no hay suficientes alternativas para conformar una pregunta de selección múltiple.

Este tipo de pregunta consiste en un conjunto de afirmaciones en las que el candidato debe seleccionar, para cada una de ellas, si el enunciado es verdadero o falso.

En su preparación debemos tener en cuenta los criterios siguientes:

- Las afirmaciones verdaderas deben ser aproximadamente iguales en número a las de las falsas, sin una alternativa regular.
- Los enunciados deben tener longitudes aproximadamente iguales para las afirmaciones verdaderas y las falsas.

- No debe contemplarse un elemento falso pequeño en una afirmación que es correcta, en relación con los elementos enfatizados.
- Redactar cada afirmación de modo que contenga una sola idea.
- Emplear lenguaje preciso, y cuando sea necesario utilizar términos numéricos más que elementos cualitativos.
- No insistir en cuestiones banales.
- Evitar frases entre paréntesis o incidencias que tengan poco que ver con la idea principal de la afirmación.
- Tratar de evitar las negaciones, en particular las dobles negaciones.

ITEM DE SELECCIÓN MÚLTIPLE

Consiste en la presentación de un enunciado (llamado también base, cuerpo, raíz o tallo) seguido de cuatro o cinco respuestas posibles (alternativas o complemento).

En este tipo de pregunta es importante mantener la utilización de cuatro o preferiblemente cinco alternativas de respuestas, ya que el uso de un número inferior afecta sensiblemente la confiabilidad de la pregunta.

El enunciado puede presentarse en forma de oración incompleta o de interrogación. Ambas variantes son válidas, pero la forma interrogativa es la más recomendable para los que recién comienzan a trabajar con este formato, ya que permite la elaboración de alternativas más específicas y homogéneas; su presentación es más fácil y es posible establecer las bases explícitamente.

No obstante, el enunciado en forma de oración incompleta economiza espacio en la base y en las alternativas y si está bien construido, facilita el paso de la lectura de la base a la búsqueda de la alternativa de respuesta correcta.

En su elaboración debemos considerar los criterios siguientes:

- Incluir en la base todos los elementos que se repitan en las alternativas.
- Elaborar alternativas breves y eliminar todo aquello que no sea esencial; se deberá incluir una idea completa.
- Tener cuidado con las bases negativas y evitar las doble negaciones, ya sea en la base como en las alternativas de respuestas.
- Elaborar alternativas que no provoquen controversias, en cuanto a cuál es la respuesta correcta.
- Todas las alternativas deben ser plausibles; se emplearán los errores comunes de los educandos.
- El lenguaje debe ser preciso, con una extensión y complejidad similar; se utilizarán expresiones frecuentes en los estudiantes.
- Evitar en lo posible el empleo de “todas las anteriores son ciertas” o “todas las anteriores son falsas”. En muchas ocasiones, esto se emplea erróneamente cuando no hay suficientes alternativas para seleccionar.
- Evitar situar la respuesta correcta en posición preferente y ubicarla siempre en una misma posición.
- No presentar nunca un conjunto de ítems de verdadero/falso en un ítem de selección múltiple: ¿Cuál de las siguientes alternativas es correcta?
- El contenido de una oración incompleta en la base debe ser equivalente a una pregunta.

ITEM DE ASOCIACIÓN

Se caracterizan, en general, por estar conformados por dos columnas. En una de ellas, generalmente la de la izquierda, se plantea un conjunto de enunciados; y en otra, se enumera un grupo de opciones o alternativas de respuestas.

Este tipo de formato puede considerarse, en lo fundamental, como un grupo de ítem de selección múltiple que comparte las mismas respuestas (opciones o alternativas).

Permiten-por sus características de construcción-la compactación de varios ítems; simplificar la estructura de las alternativas, al ser posible hacer todas las indicaciones en las instrucciones iniciales del ejercicio; y el uso de un mayor número de alternativas y, por ende, reducir el factor de respuesta al azar.

En su confección debemos tener en cuenta los criterios siguientes:

- Que cada columna recoja contenidos del mismo tipo.
- Los distractores deben estar ubicados en la columna de las opciones.
- Debe existir coincidencia gramatical entre ambas columnas.
- El número de ítems deberá oscilar entre 5 y 10.
- Deben incluirse más enunciados que respuestas, o más respuestas que enunciados. Si el número de enunciados y alternativas es el mismo, al menos una opción debe dar respuesta a más de un enunciado.
- Todas las opciones deben ser plausibles.
- Si la columna de opciones es una lista de palabras o números, es conveniente su ordenamiento.

DISCUSIÓN

Una vez confeccionado el cuestionario a emplear, es imprescindible su revisión por profesores que no hayan participado en su confección. Así mismo, se deben elaborar

las claves de calificación. No puede haber ambigüedades en cuanto a cuáles respuestas son las correctas. Causa bastante mala impresión y genera inseguridad en los estudiantes el hecho de que un examen que se supone sea objetivo no tenga las respuestas correctas bien definidas para todos.

ANÁLISIS DE LOS RESULTADOS

Una etapa importante en el trabajo con los exámenes objetivos consiste en el análisis de los resultados obtenidos. Ello se debe a que mediante éste es factible obtener información cuantitativa que permita analizar críticamente el examen empleado, así como evaluar las respuestas de los candidatos.

El principal objetivo de este análisis es validar la calidad de cada ítem y así garantizar la precisión y validez del instrumento evaluativo que estamos empleando.

Mediante este análisis de los ítems es posible:

- Identificar cuáles son los aspectos fáciles y los difíciles para los educandos.
- Diferenciar los ítems que, por deficiencias en su confección, no son capaces de discriminar a los educandos de buen rendimiento académico de aquellos que no lo tienen.

Como consecuencia de lo anterior, el análisis de los ítems ofrece informaciones que permiten focalizar la atención sobre posibles problemas en el proceso de enseñanza-aprendizaje y desarrolla en el profesor la capacidad de elaborar buenas preguntas, al modificar aquellos ítems que eran deficientes en su estructura.

La sola realización de exámenes objetivos no permite la evaluación del conjunto de conocimientos, habilidades y actitudes que conforman su competencia profesional; se requiere la complementación de otros métodos, procedimientos y técnicas.

EXAMEN TIPO “MANEJO DE PROBLEMAS DE PACIENTES”

Los estudiantes ganan experiencias con el enfrentamiento de los problemas de la vida real ⁽²⁶⁾. Cuando los pacientes no están disponibles o cuando es incorrecto probar diferentes soluciones alternativas en un paciente, los profesores han empleado la simulación en el proceso docente, durante el desarrollo de la enseñanza o al evaluar los resultados alcanzados.

Diferentes técnicas de simulación han sido desarrolladas y en cada una de ellas el candidato debe asumir el papel de estudiante, interno, residente o especialista y “manejar” el problema de salud en cuestión; en dependencia de esto y de los medios empleados, se han definido diferentes grupos de ejercicios clínicos de simulación. Otra clasificación se ha establecido atendiendo a las diferentes variedades de problemas (de diagnóstico, de emergencia médica, de alternativas de tratamiento, etc.) y se han descrito las fases para su elaboración (33, 34). Los estudios realizados acerca de esta metodología han permitido poner de manifiesto sus ventajas y desventajas ^(18,19,35) así como establecer en qué condiciones es recomendable su empleo.

CÓMO CALIFICAR UN EXAMEN

Para calificar un examen dirigido a evaluar la competencia de un educando, ^(36,37,38) en general, se emplean determinadas técnicas, que tienen que estar adecuadamente diseñadas con suficiente anterioridad y validadas, inicialmente, por consenso de expertos y, después, por el análisis matemático de sus resultados.

Entre las más empleadas tenemos:

- Listas de comprobación y escalas de calificación.

Las listas de comprobación recogen de forma bien detallada (y en secuencia requerida en el caso que se evalúe un proceso), las diferentes acciones de las tareas que deben ser objeto de evaluación. Pueden ser calificadas en unos casos de si se realizaron o no; en otros, verificar la secuencia o calidad con que se realizó la tarea.

Las escalas de calificación establecen, para un conjunto de acciones, diferentes gradaciones de calificaciones en relación con la calidad con que es ejecutada. Son muy útiles para evaluar actitudes, interacciones o el estilo con que se realiza una tarea clínica.

- Claves de calificación de respuestas.

Recogen las opciones o los patrones de respuestas que se aceptan como válidas para cada una de las preguntas realizadas en el examen. Pueden ser abiertas o cerradas. El empleo de dichas claves, elaboradas con anterioridad a la realización del examen, es un prerrequisito para garantizar su confiabilidad y eficiencia. Tiene que estar en correspondencia con la escala de calificación empleada, el nivel de complejidad de las preguntas, los patrones de respuesta aceptados y el entrenamiento previo que tenga el educando, en relación con el tipo de examen utilizado. Deben de validarse matemáticamente con posterioridad.

- Observación y grabación en video.

Es un método alternativo que se puede emplear. Se utiliza sobre todo con pacientes simulados, entrenados previamente. Proporciona una amplia retroalimentación. Su costo y los recursos que requiere limitan su empleo generalizado.

Al igual que preconizamos la necesidad de combinar los métodos, los procedimientos y las técnicas en un mismo examen, con el fin de incrementar su validez

y su confiabilidad, recomendamos a su vez, el empleo coordinado de diferentes técnicas que posibiliten incrementar la calidad de la calificación a otorgar.

Para emitir la calificación global de un examen se emplean por lo general tablas de decisiones en las cuales se ponderan las calificaciones de las diferentes preguntas, secciones o problemas ya calificados de forma independiente, ya sea con igual valor o no, con el fin de estandarizar la calificación global del examen.

El establecimiento de normas o patrones de calificación es un proceso complejo a desarrollar por el colectivo docente, con el objetivo de definir los elementos que determinan, en primer lugar, la calificación de aprobado de un examen y, en segundo orden, los elementos que caracterizan cada escala de calificación utilizada.

Es necesario comprender que cualquier norma o patrón de calificación es por su propia naturaleza arbitrario. Debemos partir del análisis de su contenido en correspondencia con los objetivos educacionales a evaluar. Se han desarrollado diferentes métodos a partir de la propuesta de Livingston y Ziesky en 1982, que podemos resumir en dos grandes grupos ^(37,39).

-Método normativo referencial.

Establece un patrón relativo que depende de la competencia mostrada por cada educando que ha sido examinado. Se basa en la distribución de frecuencias de los resultados de las calificaciones que se obtienen en un examen clínico dado. La calificación de aprobado se establece con posterioridad al examen.

Se aplica con el 15 o el 20% de los educandos examinados con las peores calificaciones, o aquellos candidatos cuyas calificaciones están por debajo de la media de los desaprobados. La escala puede ser numérica o alfabética. Tiene sus limitaciones para ser aplicada en la evaluación de la competencia profesional.

Método criterio referencial.

Parte de un patrón absoluto, que se establece con anterioridad a la realización del examen. Se basa en el análisis de los contenidos, en relación con los objetivos educativos a evaluar. La selección de la calificación de aprobado (por ejemplo, regular o 70 puntos), se arbitra independientemente de los resultados que obtengan los estudiantes.

Es el método más apropiado para calificar la competencia, y su escala también puede ser numérica o alfabética.

En los exámenes escritos de selección múltiple se tiene en cuenta el nivel de dificultad o complejidad de cada bloque de preguntas o problema evaluado (contenido esencial o no, obstáculos o distractores con diferentes niveles de complejidad, etcétera). Posibilita una mejor gradación de las calificaciones.

CAPÍTULO III: OBJETIVOS

OBJETIVO GENERAL: Caracterizar el proceso de evaluación teórica escrita en el área clínica de la Facultad de Ciencias Médicas “Dr. Enrique Cabrera” durante el curso 2003-2004.

OBJETIVOS ESPECÍFICOS:

1. Evaluar la correspondencia de los contenidos incluidos en los exámenes teóricos escritos con los establecidos por los objetivos de los programas de asignaturas.
2. Evaluar la calidad metodológica del diseño de las preguntas de los exámenes teóricos.
3. Evaluar la concordancia en los criterios de calificación de los exámenes teóricos escritos.
4. Evaluar la magnitud de la variabilidad entre profesores en relación con la evaluación global de los exámenes teóricos escritos.

CAPÍTULO IV: DISEÑO METODOLÓGICO.

Se trata de una investigación de carácter descriptivo.

Para caracterizar los exámenes teóricos finales escritos, tanto en relación con la correspondencia entre los objetivos de los programas ^(40,41,42,43) y el contenido evaluado en dichos exámenes como por la calidad metodológica de los cuestionarios utilizados, se constituyó un grupo de expertos integrado por 5 profesores con demostrada experiencia y dominio de estos aspectos, los que se encargaron de hacer el dictamen correspondiente, tomando como referencia los criterios que aparecen recogidos en el marco teórico. Para buscar la mayor uniformidad posible se efectuó un taller con dichos profesores, en el que se debatieron los puntos de vista del grupo, para llegar a un consenso que sirvió como marco de referencia para el trabajo.

Para las preguntas tipo ensayo, se constató si la tarea quedaba bien definida para el estudiante, y dentro de ellas, para las preguntas problémicas, si éstas ofrecían toda la información necesaria. Por último en las preguntas tipo test objetivos, se puso énfasis en los criterios establecidos ya descritos en el marco teórico para la confección de acuerdo a su tipo. ⁽²⁾

Se expresaron en porcentajes la frecuencia de preguntas con problemas de diseño en cada examen aplicado, la frecuencia de errores con relación a la totalidad de preguntas revisadas y al tipo de pregunta y los tipos de errores en relación con el total de errores identificados.

Para la evaluación de la concordancia en los criterios de evaluación, se sometieron a recalificación anónima por un trío de profesores cada uno de los exámenes teóricos escritos utilizados en el curso 2003-04 en la Facultad de Ciencias Médicas Dr. Enrique

Cabrera. Se tomó como patrón de referencia un examen escogido aleatoriamente entre aquellos aplicados y calificados en el momento en que se cursó cada una de las asignaturas. Los profesores seleccionados tenían no menos de 3 años de categorizados y de experiencia en la aplicación y calificación de instrumentos de evaluación, eran de la misma disciplina a la que correspondía el examen que debían recalificar y no habían tenido participación en la elaboración ni en la aplicación de dicho examen. Se utilizaron cuestionarios reales, respondidos por los estudiantes, a los que se le suprimieron las calificaciones otorgadas originalmente.

Según el procedimiento antes descrito, se obtuvieron para cada examen tres calificaciones para cada pregunta y tres calificaciones finales. Para lograr el análisis de concordancia se realizaron 6 comparaciones:

- Del patrón contra los resultados del primer profesor calificador.
- Del patrón contra los resultados del segundo profesor calificador.
- Del patrón contra los resultados del tercer profesor calificador.
- Los resultados del primer profesor calificador contra los del segundo.
- Los resultados del primer profesor calificador contra los del tercero.
- Los resultados del segundo profesor calificador contra los del tercero.

De acuerdo con el número k de preguntas de cada examen, habría $6(k+1)$ coincidencias posibles si incluimos la nota final en las comparaciones. A partir de ahí se obtuvieron las frecuencias de concordancias y discordancias en las calificaciones emitidas en cada examen. La medida final de concordancia elegida fue la frecuencia relativa (porcentaje) de concordancias.

En relación con las discordancias se calculó qué porcentaje de ellas fueron de un punto, qué porcentaje fueron de dos puntos y qué porcentaje fueron de tres puntos, puesto que todas las discordancias no tienen la misma gravedad. Para evaluar la magnitud de la variabilidad entre profesores en relación con la calificación final global de los exámenes teóricos escritos a partir de las calificaciones otorgadas a las diferentes preguntas, se constituyó un grupo de calificadores integrado intencionalmente por 50 profesores de más de 5 años de experiencia que manifestaron su disposición a colaborar con el estudio; a cada uno se le solicitó otorgar la calificación final a un total de 16 exámenes, tomados aleatoriamente entre los ya aplicados y que tenían todas sus preguntas calificadas. La información que recibieron los profesores se circunscribió a las secuencias de calificaciones de preguntas de los 16 exámenes utilizados como base para esta parte del estudio.

El número de preguntas de cada uno de los exámenes fue el siguiente:

Examen	Número de preguntas
1. Propedeútica	5
2. Med Interna	5
3. Ginecobst. 1 ^a	20
4. Ginecobst. 3 ^a	20
5. Ginecobst. 4 ^a	10
6. Pediatría I S	10
7. Pediatría II S	10
8. Salud Púb. I	10
9. Salud Púb. II	8
10. Med. Legal	5
11. MGI I	5
12. MGI II	5
13. Ginecobst. Int.	20
14. Medicina Int.	5
15. M G I Int	5
16. Pediatría Int	10

La variabilidad entre profesores (que resolvimos tomar como una medida de uniformidad en la calificación) se calculó como la varianza de una variable multinomial que toma valores entre 2 y 5. (En la práctica en ninguno de los exámenes se registraron más de 2 notas diferentes por lo cual fue posible trabajar con la varianza de una distribución binomial. Por tanto la variabilidad viene dada por el producto de las proporciones de profesores que respondieron cada una de las 2 alternativas que se presentaron, dividido este producto entre el total de profesores encuestados. Puesto que el número de profesores encuestados es constante ($n=50$) se trabajó simplemente con el producto de 2 proporciones $V= p*q$. Nótese además, que al presentarse sólo dos alternativas, ocurre que $q= 1 - p$.)

Una alternativa tal vez técnicamente preferible habría sido generar aleatoriamente diferentes patrones de respuestas para la asignación de una nota final por parte de los profesores. Esta variante se descartó por 2 razones:

- Para evitar que se generasen patrones que fueran poco típicos en la práctica.
- Por razones de tiempo y factibilidad para el desarrollo de las encuestas.

Previo a la realización de las tareas de investigación que han sido descritas, se explicó a los profesores participantes el propósito de la tarea que se les estaba solicitando, además de darles la garantía de la discreción en el manejo de los resultados, los que se utilizarían exclusivamente en función de una investigación y para ponerlos a disposición de cada colectivo docente y se decidió su participación una vez que expresaron su anuencia.

CAPÍTULO V- RESULTADOS Y DISCUSIÓN.

En el análisis realizado por los expertos se encontró un 100% de correspondencia de los contenidos evaluados con los objetivos de los programas. No es en este aspecto en el que se registran los problemas en la evaluación de los estudiantes. Se refleja que los profesores del área clínica fueron cuidadosos en el diseño de las preguntas, tratando de corroborar el dominio por los estudiantes de los contenidos correspondientes a los objetivos de los programas de estudio.

En la tabla I aparecen los resultados obtenidos en la revisión de los aspectos metodológicos en las preguntas de los cuestionarios de exámenes.

Tabla I

**Diseño metodológico de las preguntas en los exámenes teóricos escritos.
FCM Dr. E. Cabrera. Curso 2003-2004**

Examen	No. preguntas	Diseño adecuado	
		Número	%
Propedéutica	5	4	80
Med. Interna	5	4	80
G.O. 1ª Rot	20	15	75
G.O. 3ª Rot	20	14	70
G.O. 4ª Rot	10	7	70
MGI I	5	3	60
Pediatría I S	10	8	80
Pediatría 2 S	10	9	90
S Pública I S	10	9	90
S Pública2 S	8	7	87,5
MGI II	5	3	60
Med Legal	5	5	100
Internos GO	20	16	80
Internos M I	5	5	100
Internos MGI	5	5	100
Internos Ped	10	9	90
Total	153	123	80, 4

Se aprecia que de un total de 153 preguntas calificadas en la totalidad de los exámenes revisados, 123 resultaron correctas y 30 tuvieron problemas metodológicos de algún tipo, lo que representa casi un 20 % del total. En todos los casos se trató de preguntas tipo test objetivo, lo que refleja un mejor dominio por parte de los profesores de los aspectos metodológicos de las preguntas de ensayo, mientras que la metodología para la elaboración de preguntas de test no es de adecuado dominio de los profesores, lo que debe ser tenido en cuenta al diseñarse los planes de capacitación de los docentes. Este punto de vista se refuerza con el hecho de que las preguntas de test representan la mayor parte de las que suelen utilizar los profesores al elaborar los instrumentos evaluativos.

Aun cuando algunos de los criterios técnicos tomados como referencia pudieran ser objeto de discusión, por ahora son los que la práctica ha consolidado como válidos para asegurar la confiabilidad de las evaluaciones en las que se emplean. En última instancia, la realización de estudios futuros acerca del perfeccionamiento de este tipo de instrumento tiene que partir necesariamente de su adecuada aplicación, tomando como referencia los criterios de calidad existentes hasta el momento.

En todos los exámenes teóricos aplicados durante el curso 03-04 se confirmó que al menos en una de sus preguntas se presentaban problemas metodológicos de diseño. Los exámenes de MGI I y MGI II reflejaron mayores problemas metodológicos en su diseño lo que podría corresponderse con ser el colectivo de profesores más joven o de menor experiencia. Sin embargo, llama la atención el hecho de que esos mismos colectivos de profesores elaboraron correctamente los cuestionarios aplicados en la rotación de MGI del internado. Debido a que en el presente estudio no hemos abordado los aspectos causales de estos resultados, no es posible formular una explicación a

este hecho, aunque sí nos queda clara la conveniencia de profundizar en el tema de la calidad de los instrumentos de evaluación en futuras investigaciones. Un análisis similar se pudiera aplicar a la situación que muestran los instrumentos utilizados para evaluar a los estudiantes en la estancia de Ginecobstetricia, lo que refuerza el planteamiento de la necesidad de hacer estudios para la determinación de factores asociados causalmente con estos resultados, ya que casi seguramente el colectivo docente de Ginecobstetricia no se diferencie sustancialmente de otros del área clínica con mejores resultados, lo que requiere confirmación.

Tabla II
Tipo de preguntas con errores metodológicos en los exámenes aplicados en la FCM Dr. E. Cabrera. Curso 2003-2004

Tipo de pregunta	Número total según tipo	% del total de preguntas	No. de preguntas con deficiencias	% del total de su tipo.
Selección múltiple	63	41,17	14	22,2
Respuesta alternativa	44	28,76	9	20,45
Asociación	26	16,99	7	26,9
Ensayo	20	13,07	0	0
Total	153	100,0	30	19,6

De las 153 preguntas, 20 fueron de ensayo. De las 133 preguntas de tipo test, 63 fueron de selección múltiple, 44 de respuesta alternativa y 26 de asociación. Que se presente la mayor frecuencia de errores metodológicos en las preguntas de test en sus diferentes variantes es algo que merece la pena discutirse en los colectivos de asignatura como una deficiencia que es necesario erradicar, dadas las ventajas de este tipo de cuestionario en diversas circunstancias evaluativas, especialmente en

situaciones en que pudiera ser necesario evaluar cifras elevadas de estudiantes con el mayor nivel posible de objetividad.

Resulta de interés destacar que las preguntas de asociación son las de tipo test menos usadas por los profesores y al mismo tiempo, las más deficientes desde el punto de vista de la calidad técnica de su elaboración. Por otra parte, aunque el % de preguntas de selección múltiple con deficiencias metodológicas no es el más elevado, en nuestra opinión sí es el de mayor impacto en la calidad de los instrumentos evaluativos, dada la elevada proporción de ese tipo de pregunta en los cuestionarios.

Por último, es notable la poca preferencia de los profesores por las preguntas de ensayo. Tomando en cuenta que son las que más oportunidad brindan para que el alumno demuestre la capacidad para expresar sus ideas coherentemente, profundizar en los factores que están incidiendo en su relativamente baja utilización es otro aspecto a tener en cuenta para estudios ulteriores. En opinión del autor, el hecho de ser mucho más subjetivas y requerir mayor tiempo para su calificación pudieran ser factores que estén incidiendo en la pobre utilización que se les da en los exámenes finales, lo que requiere comprobación.

Tabla III

**Tipo de error metodológico detectado en los exámenes escritos de la FCM
Dr. E. Cabrera Curso 2003-2004**

Errores	Número de veces que se detectó	% del total de preguntas con errores
Preguntas de respuesta alternativa con diferencia considerable* en número de proposiciones verdaderas y falsas.	9	30
Preguntas de selección múltiple en que se utiliza "todas las anteriores son verdaderas" o "todas las anteriores son falsas".	8	26,7

Preguntas de selección múltiple en que se utiliza conjunto de ítems de verdadero o falso. (Ej. ¿Cuáles alternativas son correctas?)	6	20,0
Preguntas de asociación con igual número de enunciados que de respuestas.	5	16,6
Preguntas de asociación que en las columnas recogen contenidos de diferente tipo.	2	6,7
Total	30	100

* Se tomó cuando la diferencia era mayor o igual de 70 %

En la tabla III mostramos el tipo de error metodológico detectado en la confección de las preguntas según su frecuencia. Como se aprecia, son deficiencias cualitativas de fácil solución, por lo que en nuestra opinión, esta información nos proporciona una mayor precisión en relación con los aspectos a enfatizar en las actividades de capacitación que se realicen con los profesores para mejorar su dominio de la tecnología que se analiza y lograr un impacto cualitativo importante en los instrumentos evaluativos.

Tabla IV

**Concordancias en los criterios de calificación. Exámenes teóricos escritos.
FCM Dr. E. Cabrera Curso 2003-2004**

Examen	Número de coincidencias posibles $6(k+1)^*$	Concordancias	
		Número	%
Propedeútica	36	23	63,9
Med. Interna	36	22	61,1
G.O. 1ª Rot	126	83	65,9
G.O. 3ª Rot	126	110	87,3
G.O. 4ª Rot	66	49	74,24
MGI I	36	32	88,9
Pediatría I S	66	36	54,5
Pediatría 2 S	66	48	72,7
S Pública I S	66	51	77,3
S Pública2 S	54	33	61,1
MGI II	36	24	66,7
Med Legal	36	14	38,9
Internos GO	126	103	81,7
Internos M I	36	24	66,7
Internos MGI	36	36	100,00
Internos Ped	66	39	59,1

Las diferencias en los criterios de calificación empleados se evidencian al evaluar la concordancia. En opinión del autor, estos resultados ponen en evidencia uno de los problemas que mayor insatisfacción genera en los estudiantes en relación con el área clínica. Se puede afirmar que en algunas disciplinas la situación identificada requiere ser atendida a la mayor brevedad posible. Estos resultados nos llevan a plantear algunas preguntas: ¿por qué en algunas asignaturas se presentan por cientos de concordancia tan bajos?; ¿por qué son tan diferentes los resultados en las rotaciones de Ginecobstetricia entre sí y con la rotación del internado?; ¿por qué también ocurre lo mismo en Pediatría?; ¿por qué resulta tan diferente la concordancia en la calificación entre la MGI-I, la MGI-II y la MGI del internado? Encontrar las respuestas a estas preguntas requiere de un estudio de causalidad mucho más complejo, pero ineludible si se tiene en cuenta que resulta imprescindible para poder fundamentar un trabajo metodológico en los colectivos docentes de las diferentes disciplinas, orientado a perfeccionar cualitativamente la evaluación en el área clínica de la Facultad Dr. Enrique Cabrera, no sólo en lo referente a la calidad técnica de los instrumentos que se utilicen sino para el establecimiento de criterios de calificación con una adecuada fundamentación. Si bien el establecimiento de patrones rígidos ha resultado inoperante, la falta de criterios metodológicos, que es el otro extremo, resulta peligroso, como se ratifica en los resultados que se muestran a continuación.

Tabla V

**Discordancias en los criterios de calificación.
Exámenes teóricos escritos. FCM Dr. Enrique Cabrera. Curso 2003-2004**

Exámenes por asignaturas	No. de Coincidencias Posibles 6(k+1)	Discordancias		Discordancias		
		No.	%	De 1 punto	De 2 puntos	De 3 puntos
Propedéutica	36	13	36,1	13 (100%)		
Med. Interna	36	14	38,8	14 (100%)		
G.O. 1ª Rot	126	43	34,1	35 (81,4%)	8 (18,6%)	
G.O. 3ª Rot	126	16	12,7	14 (87,5%)	2(12,5%)	
G.O. 4ª Rot	66	17	25,8	17 (100%)		
MGI I	36	4	11,1	4(100%)		
Pediatría I S	66	30	45,5	17(56,7%)	10 (33,3%)	3 (10%)
Pediatría 2 S	66	18	27,3	14(77,8%)	1 (6.3%)	3 (16,7%)
S Pública I S	66	15	22,7	15(100%)		
S Pública2 S	54	21	38,8	19(90,4%)	2 (9,6%)	
MGI II	36	12	33,3	8 (66,7%)	4 (33,3%)	
Med Legal	36	22	61,1	19(86,4%)	3(13,6%)	
Internos GO	126	23	18,3	20(86,9%)	3(13,1%)	
Internos M I	36	12	33,3	12(100%)		
Internos MGI	36	--	--	--	--	--
Internos Ped	66	27	40,9	21(77,8%)	5(18,5%)	1(3,7%)

Con excepción de la disciplina Medicina Interna, en la que la discordancia identificada en las calificaciones de todas sus asignaturas fue de un punto, en las restantes la situación caracterizada es muy seria y en casos como Pediatría, la necesidad de profundizar en los factores que están gravitando sobre esta discordancia tan marcada es evidente y apremiante. Estamos ante una situación que no solamente tiene que estar resultando generadora de una justificada frustración para los estudiantes, sino que además tiene implicaciones éticas, económicas, gerenciales y humanas, por la repercusión que estos resultados tienen en particular para los estudiantes afectados y en general, para la institución, que apoya el análisis de su gestión sobre la base de resultados que en realidad no pueden ser considerados confiables. Téngase en cuenta que una discordancia de dos puntos significa que lo que unos profesores valoran como 3 (regular), otros lo valoran como 5 (excelente); lo que unos valoran como 2 (desaprobado), otros lo valoran como 4 (bien) y mucho más crítica es la discordancia de

3 puntos, presentada particularmente en la disciplina Pediatría, en la que lo que para algunos profesores es excelente, para otros es desaprobado.

Tabla VI

Magnitud de la variabilidad entre profesores en la calificación final de los exámenes. FCM Dr. E Cabrera. Curso 2003-2004

Examen por asignatura	Notas finales								X
	5		4		3		2		
	No.	%	No.	%	No.	%	No.	%	
Propedeútica	-	-	50	100	-	-	-	-	0
Med Interna	29	58%	21	42%	-	-	-	-	0,24
Ginecobst . 1 ^a	-	-	23	46%	27	54%	-	-	0,25
Ginecobst. 3 ^a	20	40%	30	60%	-	-	-	-	0,24
Ginecobst. 4 ^a	-	-	50	100%	-	-	-	-	0
Pediatría I S	-	-	50	100%	-	-	-	-	0
Pediatría II S	-	-	25	50%	5	50%	-	-	0,25
Salud Púb. I	-	-	9	18%	41	82%	-	-	0,14
Salud Púb. II	-	-	7	14%	43	86%	-	-	0,12
Med. Legal	-	-	50	100%	-	-	-	-	0
MGI II	-	-	-	-	-	-	50	100%	0
Ginecobst. Int.	50	100%	-	-	-	-	-	-	0
Medicina Int.	-	-	50	100%	-	-	-	-	0
M G I Int	50	100%	-	-	-	-	-	-	0
Pediatría Int.	-	-	45	90%	5	10%	-	-	0,09

Como complemento de los resultados mostrados hasta aquí, aparecen los obtenidos al someter al criterio de 50 profesores de experiencia las 16 secuencias de preguntas calificadas para que emitieran la calificación global que a su juicio debía tener cada examen, los que aparecen recogidos en la tabla VI, en la que se muestra la varianza de una distribución binomial, que en este caso se ha multiplicado por el total de encuestas. La varianza de una distribución binomial es el producto de $p*(1-p)$ en donde p denota la probabilidad de un tipo de respuesta. Las calificaciones siguen, en rigor, una distribución multinomial pero, como ya se señaló al describir el diseño metodológico, en la práctica las calificaciones fueron solamente 2, y fueron además, contiguas, es decir

no se diferenciaron en más de 1 punto, lo que ocurrió en 7 de los 16 exámenes recalificados.

En opinión del autor, este es otro de los problemas más directamente relacionados con la insatisfacción mostrada por los estudiantes con la evaluación y, conjuntamente con las discordancias caracterizadas anteriormente, le resta el carácter objetivo que la evaluación pudiera haber tenido a partir de la utilización preferencial de cuestionarios tipo test objetivo.

El análisis conjunto de los resultados de las tablas V y VI pone de manifiesto que principalmente los colectivos docentes de Ginecobstetricia, Salud Pública y sobre todo Pediatría, tienen los peores resultados. Independientemente de que se trata de un problema generalizado, en opinión del autor, estos colectivos deben ser objeto de la atención inmediata de investigadores, metodólogos y autoridades académicas, para buscar las causas y las soluciones a tan compleja problemática.

Al analizar los resultados en los exámenes agrupados según el número de preguntas encontramos lo siguiente:

Tabla VII

Variabilidad en la calificación de exámenes clasificados según número de preguntas. FCM Dr. E Cabrera. Curso 2003-2004

Número de preguntas	Exámenes según número de preguntas	Exámenes con más de una calificación	Por ciento
Con 5 preguntas	7	1	14,3
Con 8 preguntas	1	1	100
Con 10 preguntas	5	3	60
Con 20 preguntas	3	2	67
TOTAL	16	7	43,8

Estos últimos resultados ponen de manifiesto que aquellos exámenes diseñados con 5 preguntas, según se recomienda en los documentos metodológicos emitidos por el Viceministerio de Docencia e Investigaciones, tuvieron una variabilidad considerablemente menor que los confeccionados con un mayor número de preguntas. Aun cuando es importante dejar a los colectivos docentes las libertades que necesiten para tomar decisiones directamente relacionadas con el proceso que supuestamente dirigen, en nuestra opinión resulta evidente que la reducción de esta variabilidad es necesaria y requiere de la existencia de alguna guía o tabla de decisiones que, sin utilizarse como una camisa de fuerza, de alguna manera contribuya a fomentar el análisis colectivo de las evaluaciones y la búsqueda de criterios cada vez más homogéneos en cada colectivo docente.

CONCLUSIONES

El estudio llevado a cabo permitió realizar la caracterización de la evaluación final teórica escrita en el área clínica de la Facultad Dr. Enrique Cabrera en relación con los aspectos propuestos en los objetivos de la investigación, con lo que se pusieron de manifiesto algunos hechos que presumiblemente son importantes fuentes generadoras de insatisfacciones por los estudiantes.

Si bien se pudo constatar que ha existido una buena correspondencia entre los objetivos planteados por las asignaturas y los contenidos que han sido objeto de evaluación, éste es un tema no agotado, en vista de que las puntualizaciones que hacen los objetivos no se limitan a la selección de los contenidos, sino que precisan la habilidad a desarrollar con esos contenidos, su nivel de asimilación y de profundidad, lo que también es de gran importancia para la verificación de la correspondencia entre el currículo diseñado y el ejecutado.

La constatación de deficiencias metodológicas en la elaboración de los cuestionarios de examen, particularmente en los de tipo test objetivo y el hallazgo de una considerable discordancia en los criterios de calificación de las preguntas de examen, con elevada variabilidad en los criterios para la asignación de una calificación global a los exámenes a partir de la valoración de las calificaciones de las preguntas y la focalización de los problemas de la calidad de la evaluación de forma más marcada en determinados colectivos docentes, han puesto en evidencia la necesidad de ampliar y profundizar las investigaciones relacionadas con la calidad de la evaluación docente en la Facultad Dr. Enrique Cabrera, así como la adopción de medidas orientadas a dar solución a los problemas que pudieran ser identificados y caracterizados.

RECOMENDACIONES

- Proponer la realización de reuniones en los colectivos de asignaturas de la Facultad, donde se expongan con mayor detalle los resultados de este trabajo en lo que concierne a cada uno y se promuevan la discusión y el análisis valorativo, para la búsqueda de posibles causas y soluciones y el fomento del trabajo metodológico en dichos colectivos.
- Proponer y desarrollar cursos o sesiones de capacitación entre los docentes en el tema evaluación del aprendizaje.
- Recomendar a los colectivos docentes la aplicación de la Instrucción VAD 5 de 1991, relacionada con la elaboración y calificación de exámenes finales en el área clínica de la carrera de Medicina.
- Dar continuidad al presenta estudio mediante la realización de investigaciones centradas en el tema de la evaluación docente en la Facultad.

REFERENCIAS BIBLIOGRÁFICAS

1. González Pérez M: La evaluación del aprendizaje: Tendencias y reflexión crítica. Rev. Cubana de Educ. Superior. Vol. XX: 47-62. 2000
2. Salas Perea RS: Educación en salud. Competencia y desempeño profesionales. La Habana: Editorial Ciencias Médicas, 1999
3. Rodríguez Diéguez JL: Didáctica general y objetivos de evaluación. España; Editorial Ciencias, 1984.
4. Talízina NF. El control y sus funciones en el proceso docente. Revista Pedagogía Soviética. 1989.
5. González Pérez M: La evaluación del aprendizaje: Tendencias y reflexión crítica Revista Cubana de Educación Superior. XX(1): 47-62. 2000.
6. Friedman Ben, David M: The role of assessment in expanding professional horizons. Medical Teacher 22(5): 472-77. 2000.
7. Guntsche E, Montbrun María de la C: La evaluación del aprendizaje en Medicina. Universidad Nacional de Cruz Argentina, 2001. Disponible en: www.fac.org.ar/scvc/llave/edu/guntsche.htm
8. 11ª Conferencia Internacional de Ottawa sobre Educación Médica. 2002. Disponible en: www.bcmedic.com/ottawa/ottawa_esp.htm
9. Pérez Cavan M L; Carretero Torres M R; Palma Muñoz M; Rafel Cufí E: La Evaluación de la calidad del aprendizaje en la Universidad. Departamento de Psicología. Universidad de Girona. 1999
10. Elton L y Laurillard D: Trends in student learning. Studies in Higher Education, 4, 87-102. 1997
11. Biggs J B: Assessing learning Quality : reconciling institutional, staff and educational demands, Assessment and evaluation in higher education vol 21, 1, 5-15 .1996
12. Gros B y Romana T: Ser profesor palabras sobre la docencia Universitaria. Barcelona. Publicaciones de la Universidad de Barcelona. 1995
13. Sayer M, Bowman D, Evans D, Wessler A, Word D. Use of patients in professional medical examinations: current UK practice and the ethical implications for medical education. BMJ 2002, 324:404-407.

14. Salas Perea RS. El papel de la evaluación en el proceso enseñanza aprendizaje. Educ. Med. Sup. 5 (1): 3-17. 1991.
15. Castro Ruz F. Discurso de clausura pronunciado en el Encuentro Nacional de Estudiantes de Ciencias Médicas. La Habana. Palacio de las Convenciones, 1989.
16. Guilbert JJ. Educational Handbook for Health Personnel. Revised ed. Geneva, 1981. WHO Offset Publication No 35.
17. Newble DI. Assesing clinical competence at the undergraduate level. Dundee: Association for the Study of Medical Education, 1992, Booklet No 25.
18. Handfield-Jones R, Holling Worth G. CEPC's Certification Examination. An Overview, Canadian Family Physician 1990; 36: 2069-2074.
19. Kane MT. The assesment of Profesional Competence. Evaluation & The Health Professions 1992, 15(2): 163-182.
20. Harden RM. Assesment of Clinical Competence Examiners' tool-kit. En: Hart IR, Harden RM and Walton HS, eds. Newer Development in Assesing Clinical Competence. Montreal: Can-Heal Publications, 1986: 11-19
21. Salas Perea RS. El proceso de enseñanza aprendizaje: Su esencia. Educ Med Super 1992 6(1): 3-16.
22. Mc Guirre C. Written methods for Assessing Clinical competence. En: Hart IR, Harden RM and Walton HS, eds. Newer Development in Assessing Clinical Competence, Montreal: Can Heal Publications, 1987; 46-50.
23. Cox KR, Ewan ChE. The Medical Teacher. Edinburgh: Churchill Livengstone, Longman Group Limited, 1982.
24. Fabb WE, Marshall JR. The Assesment of Clinical Competence in General Family Practice. Lancaster: MTP Press Limited, 1983.
25. Knox JDE. The modified Essay Question. Dundee: Association for the Study of Medical Education, 1975, Booklet No 5.
26. Ilizástigui F, Douglas R, Calvo I. Los problemas docentes y la evaluación de los estudiantes en el nuevo plan de estudio de Medicina (folleto). La Habana: Vicerrectoría de Desarrollo del Instituto Superior de Ciencias Médicas de la Habana, 1988.

27. Sutnick AI, Ross LP, Wilson MP. Assessment of Clinical Competences of Graduates of Foreign Medical Schools. JAMA 1993; 270 (9): 1041-1045.
28. Sutnick AI, Ross LP, Wilson MP. Assessment of Clinical Competencies by the Foreign Medical Graduate Examination in the Medical Sciences. Teaching and Learning in Medicine 1992; 4(3): 150-150-155.
29. Pappworth MH. Passing Medical Examinations 2nd ed. London: Butterworths, 1976: 37-41.
30. Marelín Vianna H. Los test en la educación. Pamplona: Ediciones Universidad de Navarra SA, 1983.
31. Hawkes HE, Lindquist EF, Mann CR, eds. The construction and use of achievement examinations Cambridge: Houghton Mifflin Company. The Riverside Press, 1936.
32. Thorndike RL, Hayen EP. Measurement and Evaluation in Psychology and Education. 4th ed. New York: John Wiley & Sons, 1977
33. Harden RM. Preparation and presentation of patient- Management Problemas (PMPs). Medical Education. 1983, 17:256. ASME Booklet No 12.
34. Marshall JR, Fabb WE. The Construction of Patient-Management Problems. Medical Education 1981; 15:126 ASME Booklet No 12.
35. Dawson-Saunders B. Overview of Assessment Methods: What's New & What's "Best". Paper from the Second Cambridge Conference Workshop. Cambridge: University School of Medicine, 1993.
36. Ilizástigui F, Calvo I, Douglas R. El Programa Director de la Medicina General Integral para el Médico General Básico. (folleto) LA Habana. Ministerio de Salud Pública. 1985.
37. Newble DI, Dauphinee D, Dawson- Saunders B, Macdonald M, Mulholland H, Page G, et al. Guidelines for the Development of Effective and Efficient Procedures for the Assessment of Clinical Competence. Paper from the Fourth Cambridge: University School of Medicine, 1989.
38. Ilizástigui F. De la incoordinación a la integración de los estudios médicos. Revista sobre Educación Superior Universidad de la Habana, Enero-Junio 1971:31-70.

39. Livingston SA, Ziesky MJ. Passing Scores. Pricetown: NJ, Educational Testing Service, 1982.
40. Programas de las asignaturas de tercer año de Medicina. La Habana: Ed. Pueblo y Educación. 1991.
41. Programas de las asignaturas de cuarto año. La Habana: Ed. Pueblo y Educación. 1991.
42. Programas de las asignaturas de quinto año de Medicina. La Habana: Ed. Pueblo y Educación. 1991
43. Programas del internado. La Habana: Ed. Pueblo y Educación. 1987.

BIBLIOGRAFÍA CONSULTADA:

- Argyris, C: Theories of action that individual learning. *American Psychologist* , 31, 638-654. 1976.
- Area Moreira, M. Que aporta internet al cambio pedagógico en la educación superior? Universidad de la Laguna. Publicado en R. Pérez (Coord): *Redes multimedia y diseños virtuales. Actas del III Congreso Internacional de Comunicación, Tecnología y Educación. Universidad de Oviedo, septiembre 2000*, pgs 128-135.
- Avila Cruz VR, *El correo electrónico y su uso óptimo en la búsqueda de información. 2000*
- Barbier, J M. *La evaluación en los procesos de formación. Barcelona: Paidos / MEC. 1993.*
- Barnett,R: *Improving Higher Education: total quality care. Buckingham. Society for Research into Higher Education and Open University Press, 185-228. 1992.*
- Biggs, J B: *What do inventories of students learning processes really measure? A theoretical review and clarification. British Journal of Educational Psychology . 63, 3-19. 1993.*
- Biggs,J B: *Assessing for Learning : Some Dimensions Underlyng New Approaches to Educational Assessment . The Alberta Journal of Educational Research, vol XLI, 1, 1-17. 1995.*
- Biggs, J B: *Assessing learning Quality: reconciling institucional, staff and educational demands. Assessment and Evaluation in Higher Education, vol 21, 1, 5-15, 1996.*
- Biggs, J B: *Enhancing teaching through constructive alignment. Higher Education, 32, 347-364, 1996.*
- Boulton- Lewis G. *Tertiary students knowledge of their own learning. The SOLO taxonomy (Structure of the Observed Learning Outcome). New York Academic Press. 1982*
- Boulton- Lewis G. *The SOLO Taxonomy as a Means of shaping and Assessing Learning in Higher Education. Higher Education Research and Development. Vol 14. 2. 143-154. 1995.*

- Chadwick, C.B. y Rivera, N.I.: Evaluación formativa para el docente. Barcelona. Paidós. 1991.
- Cirigliano FJ: Reflexiones del viejo profesor. Enviado por el autor de la lista de discusión cuedistanca@liserv.uned.es 2001.
- Clariana. M, Monereo C y Pérez Cabani, M.L. : Percepción del modelo instruccional del profesor y rendimiento académico. En C. Marcelo y P. Mingorance (eds.) Pensamiento de profesores y Desarrollo Profesional. Formación inicial y permanente. Sevilla. 453-462. 1992.
- Consejo de Universidades: Programa experimental de evaluación de la calidad del sistema universitario: Informe final. Madrid. MEC 1994.
- Consejo de Universidades: Programa de evaluación de la calidad de las universidades. Madrid. MEC 1995.
- Consejo de Universidades: Plan Nacional de Evaluación Nacional de la calidad de las Universidades. Guía de evaluación. Madrid. MEC.1996
- Consejo de Universidades: Informe de los resultados de la primera convocatoria del Plan Nacional de Evaluación de la calidad de las Universidades. Madrid. Secretaría Nacional del Consejo de Universidades. 1997.
- De Miguel, M: La calidad de la educación y las variables del proceso y el producto. Educación. 6, 29-51. 1995
- Díaz Barriga, A. El problema de la teoría en la evaluación y de la cuantificación del aprendizaje. En Díaz B: (Comp) El examen: textos para su historia y debate. UNAM . México 1993.
- Entwistle N: The impacto of Teaching on Learning Outcomes in Higher Education. Edimburgh: USDU/Commissioned by the Employment Department Training, Enterprise and Education Directorate. 1992.
- Entwistle A y Entwistle N: Experiences of understanding in revising for degree examinations. Learning and instruction, 2: 1-22. 1992.
- Fermín M: La evaluación, los exámenes y las calificaciones. Buenos Aires. Kapelusz. 1972.
- Greenjalgh T. Computer assisted learning in undergraduate medical education. BMJ 2001; 322:40-44

- Lobo LC, Marielba, JM Galpecin J. Evaluación, enseñanza y aprendizaje. Rev Educ. Med. Salud 1986; 20 (4)
- Mateo J Presentation: En J Espín y M Rodríguez L'avaluació dels aprenentatges a la univesitat Barcelona Universidad de Barcelona. 11-13. 1993 .
- Mavis BE, Cole BL, Hoppe RB: A survey of information sources used for progress decisions about medical students: [http:// www.med.ed.online.org](http://www.med.ed.online.org).2000
- Monereo C, et al: Estrategias de enseñanza aprendizaje. Formación del profesorado y aplicación en la escuela Barcelona, 1994.
- Pérez Cabani y Terradellas Piferrrer M.R.: Enseñar, aprender y evaluar lo aprendido en la Universidad. Comunicación presentada en las Jornadas de formación del profesorado universitario y la calidad de la enseñanza. Universidad del País Vasco. 21 al 23 de Nov 1999
- Pérez González OL, Portuondo Padrón R: La evaluación del aprendizaje. Revista Cubana de Educación Superior. (2) 67-77. 1998.
- Pérez Rabré M, Gonzalez Cardeli B: Sistema de experto aplicado al rendimiento académico en la enseñanza superior. Revista Cubana de Educación Superior (2) 67-77. 1998.
- Pérez Serrano, G: Investigación cualitativa. Retos e interrogantes. Técnica y análisis de datos. Madrid. 1994.
- Salas Perea, RS: La evaluación de la educación superior contemporánea. Universidad Mayor de San Andrés. Biblioteca Médica Volumen: 24. La Paz Bolivia. 1998.
- Salas Perea, RS: Evaluación curricular. Carpeta No 5 Centro Nacional de Perfeccionamiento Médico. C. Habana. 1997.
- Salas Perea, RS: La calidad del desarrollo profesional: Avances y desafíos: Trabajo presentado en el Congreso Nacional de Educación Médica: reto del Siglo XXI. C. Habana 1998
- Salas Perea, RS: La comunicación educacional. Material de estudio de la maestría de Educación Médica. La Habana, 2002.
- Tang, K C : Effects of different assessment methods on tertiary students approaches to studyng. Unpublished doctoral dissertation. University of Hong Kong 1991.

- Taylor, C: Assessment for measurement or standards: the peril and promise of large scale assessment reform. American Educational Research Journal, 31, 231-262. 1994.
- Van der Vleuten C: Validity of final examinations in undergraduate medical training BMJ: 321:1217-19. 2000
- Van der Vleuten C, et al: Clerkship assessment assessed Medical Teacher. 22(6) 392-600. 2000.
- Venturelli J: La evaluación en el proceso de formación de los profesionales médicos. Principios de evaluación formativa. En: Educación Médica: Nuevos enfoques, metas y métodos capítulo 6. serie PALTEX, Salud y Sociedad. 2000 No 5. OPS. Washington. 1997.
- Wood. LP, O Donnell E: Assessment of competence and performance at interview BMJ: 320: S2-7231. 2000

ANEXOS

Anexo 1

Patrón de repuestas para el estudio de variabilidad

Estimado profesor:

Como parte de un estudio orientado al perfeccionamiento de la evaluación docente en nuestra Facultad, es necesario evaluar la variabilidad por parte de los profesores en la asignación de la calificación global de los exámenes finales teóricos a partir de las calificaciones de las preguntas que lo integran. Con ese propósito solicitamos su participación en el llenado de esta encuesta, en la que le ofrecemos 16 series de calificaciones de preguntas para que ud. decida la calificación final de los correspondientes exámenes.

Se trata de una encuesta anónima, sin repercusión individual, cuyos resultados se emplearán solamente con fines de investigación y para la retroalimentación de los colectivos docentes en aquellos aspectos que particularmente le conciernen. La utilidad de la investigación para nuestro centro depende de la calidad de la contribución que realicen los profesores individualmente en su condición de expertos, por lo que agradecemos de antemano la cooperación que en este sentido pueda dar.

Examen 1:

P1	P2	P3	P4	P5
3	5	4	5	3

Nota final: _____

Examen 2:

P1	P2	P3	P4	P5
5	5	5	5	3

Nota final: _____

Examen 3:

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
4	3	3	4	4	3	5	4	2	5	3

P12	P13	P14	P15	P16	P17	P18	P19	P20
3	2	4	2	4	5	3	4	3

Nota Final _____

Examen 4:

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
5	3	5	5	5	5	2	5	5	5	5

P12	P13	P14	P15	P16	P17	P18	P19	P20
4	4	5	5	5	5	5	5	5

Nota Final _____

Examen 5

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
5	2	3	5	5	5	2	5	5	5

Nota Final _____

Examen 6

P1	P2	P3	P4	P5
5	3	5	5	4

Nota Final_____

Examen 7

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
5	5	4	4	3	5	3	4	3	5

Nota Final_____

Examen 8

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
5	2	5	4	5	5	5	4	2	5

Nota final:_____

Examen 9

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
3	2	5	5	2	4	2	3	4	5

Nota Final_____

Examen 10

P1	P2	P3	P4	P5	P6	P7	P8
3	5	2	2	4	4	5	5

Nota Final_____

Examen 11

P1	P2	P3	P4	P5
5	4	5	3	5

Nota final: _____

Examen 12

P1	P2	P3	P4	P5
2	5	2	2	5

Nota final: _____

Examen 13

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
5	5	5	5	5	4	5	5	5	5	5

P12	P13	P14	P15	P16	P17	P18	P19	P20
5	5	5	5	5	5	5	3	4

Nota final: _____

Examen 14

P1	P2	P3	P4	P5
3	5	4	5	4

Nota final: _____

Examen 15

P1	P2	P3	P4	P5
5	5	5	5	5

Nota final: _____

Examen 16

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
5	5	5	2	4	5	5	5	5	3

Nota final: _____