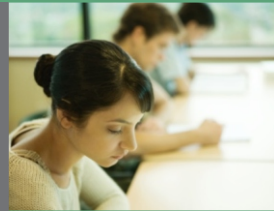


VOL
1
NÚMERO 1

Suplemento Especial de la Editorial Physiological Mini Reviews sobre Educación



EVALUACIÓN DE LOS APRENDIZAJES

Licenciada Amanda Galli

SOCIEDAD ARGENTINA DE INVESTIGACIÓN Y DESARROLLO EN EDUCACIÓN MÉDICA
(SAIDEM).

Contacto: E-mail: amanda.galli@saidem.org.ar

EVALUACIÓN DE LOS APRENDIZAJES

“Las Facultades de Medicina han de asumir su responsabilidad profesional respecto a la sociedad en general y respecto a los estudiantes en particular, asegurando que éstos posean el mínimo de competencias que les permita iniciar su vida profesional con garantías para los enfermos y para ellos mismos. Cara a todo ello, el primer paso debe ser definir estas Competencias”. Jordi Palés. [1]

“Assessment plays a major role in the process of medical education, in the lives of medical students, and in society by certifying competent physicians who can take care of the public”. Shumway J M and Harden R M. [2]

“Si bien se evalúa por muchas razones - para decidir que niño va a entrar a una escuela, para controlar a los profesores, para reconocer escuelas- el sentido fundamental es que uno pueda tomar decisiones Inteligentes”. Miguel Zabalza. [3]

“La evaluación educacional es el proceso de delinear, obtener y proveer información útil para juzgar alternativas de decisión”. Daniel Stufflebeam. [4]

“... el sistema de exámenes influyó mucho más en el aprendizaje que los objetivos de la enseñanza, la organización del programa de estudio o los métodos didácticos... Por desgracia, los exámenes en la escuela de medicina parecían estar más a menudo en conflicto que en armonía con los objetivos explícitos e implícitos de la enseñanza y con la organización del programa. ... la verdadera reforma del programa docente debía empezar por el sistema de exámenes y no por el plan de estudios”. George Miller. [5]

Se pueden mencionar varios propósitos de la Evaluación Educacional:

- Identificar necesidades de aprendizaje, nivel de conocimientos previos.
- Orientar el aprendizaje de los alumnos, brindar feedback, motivar.
- Proteger al público de médicos incompetentes.
- Certificar la idoneidad de los especialistas, recertificación.
- Conocer calidad del programa, fundamentar cambios curriculares, rendir cuentas (responsabilidad social), decidir la continuación o no de determinados programas de posgrado.

Se identifica como **evaluación diagnóstica** a la que tiene como propósito identificar necesidades de capacitación; a la que explora los conocimientos previos de los estudiantes para ajustar la programación de la enseñanza.

La **evaluación formativa** es la que utiliza los resultados de pruebas y exámenes para orientar al alumno en su proceso de aprendizaje. La evaluación formativa no adjudica notas ni calificaciones, identifica los puntos fuertes y los débiles y brinda planes de mejora.

La **evaluación sumativa** es la que verifica si el sujeto reúne los requisitos necesarios y suficientes para aprobar un curso, para recibir el título, para ganar una beca. De alguna manera se toma la decisión sumando o resumiendo varios datos. A través de la evaluación sumativa (nota final) los docentes certifican, responsablemente, la competencia de los egresados.

En todos los casos la información que brinda la evaluación es un insumo para la toma de decisiones.

Tradicionalmente, en nuestras facultades de medicina, en el ciclo biomédico se utilizan exámenes escritos para evaluar el nivel de conocimientos de los alumnos y decidir su aprobación y promoción o su aplazo y/o recursado de la materia. Todos los profesores reconocen que la evaluación es un proceso necesario y que es una actividad docente ineludible; pero la importancia que tiene la evaluación no siempre se ve reflejada en la práctica docente cotidiana. En algunas cátedras, la evaluación es vivida como un trámite administrativo, más o menos burocrático, en el que cada docente pregunta sobre

aquellas cosas que le parecen importantes y cada uno lo hace del modo que considera más apropiado. También es cierto que algunos profesores han reflexionado sobre la evaluación, advierten que es una práctica compleja y sienten que no poseen el tiempo ni las herramientas necesarias para evaluar de manera satisfactoria. También hay que destacar que en muchas cátedras la evaluación es sistemáticamente planificada en procesos participativos en los que intervienen todos los docentes que tuvieron a su cargo la enseñanza.

A la hora de elaborar los exámenes y de tomar estas decisiones los docentes deberían tener presente que en nuestro país no existe ningún examen final de carrera o prueba de habilitación profesional de modo que los exámenes que se administran en cada curso o asignatura de la carrera van “entregando” un pedacito del título, como si fuera un rompecabezas que se va armando.

En todos los casos, los exámenes deben evaluar lo que realmente se quiere evaluar y los resultados deben ser confiables. Necesariamente se requiere el más alto nivel de validez y de objetividad posible [6].

Los exámenes que utilizan los docentes son la forma de controlar la calidad del egresado que la universidad entrega a la sociedad.

El propósito de esta comunicación es presentar algunos criterios y recomendaciones para construir exámenes escritos objetivos de modo que los resultados sean lo más confiable posible.

¿Qué evaluar?

En el contexto de las carreras universitarias, en cada una de los cursos se trata de verificar o comprobar el logro de los objetivos de aprendizaje. Es decir, si se han alcanzado los resultados esperados.

¿Qué objetivos educativos figuran en su programa de enseñanza?

Los docentes son los que explicitan lo que se espera como resultado del aprendizaje.

Además de explicitar los objetivos, los docentes deciden las formas, acciones o estrategias de enseñanza que conducirán al logro de las metas propuestas. El principal papel de la evaluación consiste en proporcionar evidencias apropiadas que ayuden a estudiantes y docentes a ver en qué medida se van logrando los objetivos de aprendizaje.

Los objetivos específicos de un curso orientan la selección de contenidos y actividades de enseñanza y constituyen la variable a evaluar [7]. Ante todo cabe señalar que no hay un tipo de examen que permita obtener información sobre los distintos objetivos de aprendizaje por eso la primera recomendación es que deben usarse distintos instrumentos.

Generalmente se formulan objetivos que corresponden a tres categorías o dominios.

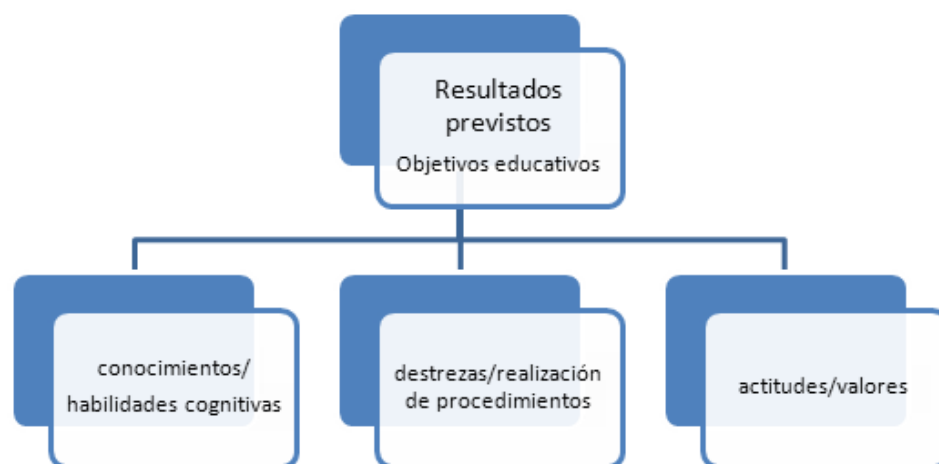


Figura 1.

Los objetivos educativos son los resultados del aprendizaje previstos o esperados.



Figura 2.

Resultados del aprendizaje e instrumentos para recoger información.

Cada tipo de examen tiene sus limitaciones por eso se recomienda utilizar distintos tipos o formatos [8].

Últimamente los exámenes orales se han revalorizado como instancias para evaluar las habilidades para expresarse oralmente y/o para hacer presentaciones en ateneos, congresos y en otros contextos profesionales. Los exámenes orales tienen serias limitaciones para evaluar conocimientos ya que, generalmente, exploran una muestra poco representativa del total de contenidos de una asignatura. Por otra parte, todas las instancias de evaluación que se basan en la observación directa requieren la elaboración de guías para sistematizar la observación (listas de cotejo, escalas, rúbricas) y controlar la subjetividad del observador. Es aconsejable que estas listas o guías sean producto de un trabajo colaborativo y que sean el resultado de un consenso sobre cómo es la mejor manera de realizar un procedimiento y/o cuáles serán los indicadores de una conducta aceptable. Existen muchos instrumentos para la evaluación del aprendizaje, algunos ya estandarizados como el mini-cex [9].

La evaluación de la competencia profesional requiere distintas instancias de evaluación y se utilizan distintos formatos para obtener información sobre los conocimientos y habilidades del sujeto.

En la siguiente tabla se presentan algunos componentes de la competencia profesional y los formatos de preguntas y pruebas a utilizar.

Competencia profesional	Formato evaluación
Conocimientos biomédicos	PEM - PA
Anamnesis	OBS - PE
Exploración física	OBS - PE
Interpretación datos clínicos	IgC- EO- PEM - PA
Decisiones diagnósticas	EO- PEM - PA
Decisiones terapéuticas	EO- PEM - PA
Realización de procedimientos	OBS - PE - MA
Habilidades comunicacionales	OBS - PE
Relaciones interprofesionales	OBS - PE- SINF
Asuntos legales	EO - PE- PEM- PA- SINF
Ética	OBS- EO - PE- PEM- PA
Prácticas preventivas	EO - PE- PEM- PA
Prácticas comunitarias	EO- PEM- PA - SINF - PY
Investigación	EO- PEM- PA- PY
Docencia	OBS- EO- PE- PEM- PA

Referencias:

PEM: pregunta elección múltiple;
PA: pregunta de respuesta abierta;
OBS: observación directa;
PE: paciente estandarizado: persona entrenada para representar el papel de usuario del servicio de salud;
IgC: interpretación de imágenes en la computadora;
EO: examen oral;
SINF: simulaciones informáticas, en la computadora;
PY: presentación de proyectos.

Qué se evalúa y cómo se evalúa tiene impacto sobre el aprendizaje porque el alumno orienta su esfuerzo para responder satisfactoriamente lo que los docentes valoran.

La mayoría de las preguntas de los exámenes escritos son preguntas informativas cuya respuesta requiere recordar un dato específico. Es responsabilidad de los docentes elaborar exámenes que requieran razonamiento y/o aplicación.

Otra forma de presentar la diversidad de instrumentos a utilizar en la evaluación se presenta en la Figura N° 3. La Pirámide de Miller resume los distintos niveles taxonómicos en el aprendizaje de la competencia profesional y el tipo de instrumento que puede ser utilizado para su evaluación. Los exámenes son instrumentos adecuados para evaluar los primeros niveles de la pirámide de Miller. En los otros niveles aparece la observación directa del desempeño y los distintos modos de sistematizar la observación para ganar en objetividad.

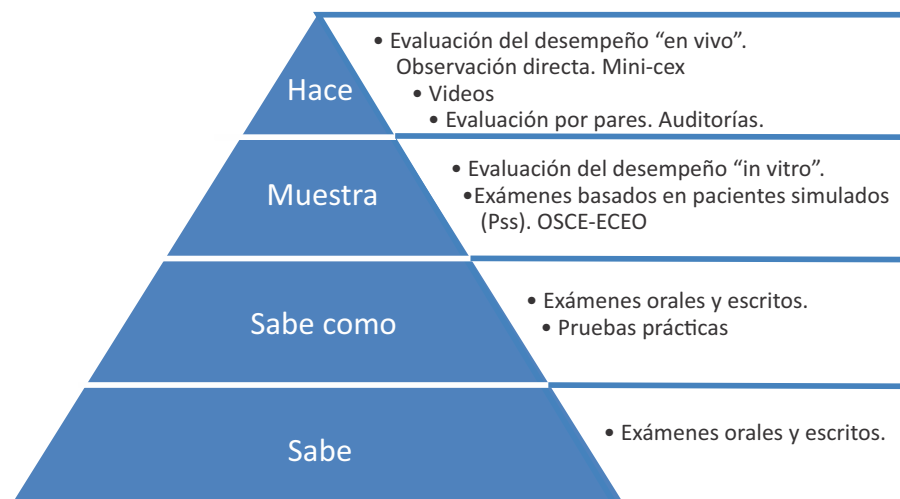


Figura 3.

Pirámide de Miller y tipo de instrumentos a utilizar para la evaluación en cada uno de los niveles.



Figura 4.

Exámenes orales y escritos, distinto formato.

Exámenes orales

Estos exámenes se pueden estructurar de dos maneras distintas:

- Disertación
- Interrogatorio oral

La disertación es una exposición oral en la que el disertante desarrolla un tema.

En el contexto escolar, el docente indica el tema y el alumno debe realizar su exposición según una secuencia, extensión y profundidad determinada por él, sin que nadie intervenga durante su desarrollo. El Interrogatorio oral consiste en una serie de preguntas sobre tópicos específicos formuladas por el docente y que el alumno responde puntualmente.

Clásicamente las preguntas se clasifican como preguntas informativas y preguntas reflexivas.

- **Ejemplos de preguntas informativas:** requieren el recuerdo de datos específicos

¿Cuál es el porcentaje de sobrevivencia...?

Establezca las diferencias y similitudes entre.....

¿Cuál es la clasificación de.....

- **Ejemplos de preguntas reflexivas:** requieren cierto nivel de análisis, sacar conclusiones, elaborar juicios de valor.

¿Está de acuerdo con...? ¿Por qué?

¿Qué actitud asumiría frente a ...? ¿Por qué?

La pregunta es un medio adecuado para estimular el pensamiento. La forma de la pregunta es un estímulo que promueve ciertos procesos mentales:

- recordar datos específicos
- establecer relaciones entre conceptos
- establecer conclusiones
- elaborar juicios de valor

El siguiente es un ejemplo de dos preguntas sobre un mismo contenido que exploran diferente nivel de conocimiento.

Preguntas [10]	Nivel de conocimiento
<i>¿ Qué región recibe el suministro sanguíneo a través de la arteria cerebelosa inferior posterior?</i>	Retención – memoria de un hecho aislado
<i>Un hombre de 62 años de edad presenta ataxia de las extremidades del lado izquierdo, síndrome de Horner, nistagmo y pérdida de las sensaciones de dolor y temperatura faciales. ¿Qué arteria es más probable que esté obstruida?</i>	Aplicación de conocimientos de ciencias básicas

¿Qué resultados del aprendizaje se pueden evaluar con un examen oral?

- Utilización de terminología, conocimiento de hechos específicos
- Habilidad para buscar y seleccionar información significativa
- Capacidad para expresar ideas: organización del discurso, coherencia, secuencia lógica
- Confianza en sí mismo, manejo del estrés

¿Cuáles son las principales limitaciones del interrogatorio oral como instrumento para recoger información sobre el nivel de conocimientos alcanzado por los alumnos?

- Escasa validez porque se formulan pocas preguntas que no representan la totalidad de la materia.
- Escasa objetividad porque los docentes para juzgar el desempeño del estudiante utilizan criterios personales, implícitos. Además en la apreciación del rendimiento del alumno influyen otros factores subjetivos.

Para mejorar la confiabilidad de los resultados de los exámenes orales es recomendable que los docentes “preparen” el examen.

Preparación	Tareas
Antes del día del examen: planificar el examen	<ul style="list-style-type: none"> - establecer los objetivos que serán evaluados (“temas que entran”) - confeccionar una tabla de especificaciones: cantidad de preguntas sobre cada tema consistente con el nivel taxonómico de los objetivos. - redactar preguntas con respuesta esperada (clave) - establecer el nivel de exigencia para la aprobación: cuántas preguntas de cada tema deben ser respondidas correctamente para aprobar el examen.
Durante el examen: no improvisar preguntas en el momento	<ul style="list-style-type: none"> - llevar una lista de preguntas agrupadas por tema. - hacer la pregunta y esperar la respuesta - reformular la pregunta si es necesario - registrar las respuestas del alumno (+ -) en relación a la clave
Después del examen: brindar feed back	<ul style="list-style-type: none"> - informar los resultados a los estudiantes: identificar puntos fuertes y débiles.

La tarea de brindar feed back a los estudiantes es fundamental en las evaluaciones parciales, durante el proceso de enseñanza-aprendizaje, para orientar el esfuerzo del alumno.

Existen otras instancias donde el docente puede observar las competencias comunicacionales de un estudiante. La capacidad para expresar y confrontar ideas se pone en juego en actividades tales como:

- Debate: controversia entre dos alumnos puntos de vista distintos
- Mesa redonda: cada participante realiza una breve presentación inicial y responde preguntas del auditorio; el coordinador de la mesa hace una síntesis del tema tratado.
- Panel: el debate se realiza entre los panelistas que deben presentar enfoque diferentes.
- Foro: el alumno debe responder las preguntas del auditorio.

Coloquio

Esta modalidad de examen oral tiene como propósito comprobar si los alumnos han adquirido, además de los conocimientos específicos de la materia, la habilidad de trabajar cooperativamente en un grupo [11]. Es decir, si el alumno es capaz de comunicar, discutir y compartir la información que posee. El desarrollo de esta habilidad es un requisito necesario para poder trabajar en un equipo de salud con actitud científica.

La evidencia que se toma en cuenta para evaluar si se han logrado los objetivos propuestos es el desempeño de cada alumno en una situación de trabajo grupal.

¿En qué consiste un coloquio? Se trata de un grupo de 5 a 7 alumnos que discuten algunos problemas planteados por un docente que cumple las funciones de coordinador de la discusión. Es recomendable que los alumnos hayan participado en algunas experiencias de aprendizaje en común, por ejemplo: constituyendo previamente una comisión de Trabajos Prácticos.

Se pueden identificar 5 momentos en el desarrollo de un coloquio:

- El docente plantea los temas/problemas. Se recomienda 2 ó 3 temas o problemas como mínimo, el primero puede ser propuesto por los alumnos para facilitar el comienzo del examen, aflojando la tensión inicial.
- Discusión grupal, se observa la participación y calidad de los aportes que realiza cada uno de los alumnos. El docente que coordina un coloquio debe tener un especial entrenamiento en dinámica de grupo.
- El docente formula preguntas, cada alumno debe responder por lo menos dos o tres preguntas. Esta segunda parte del coloquio le permite al docente ratificar la apreciación del grado de conocimiento demostrado por el alumno durante la discusión de los temas.
- El docente plantea una instancia de autoevaluación / reflexión. Cada uno de los alumnos juzga su propio desempeño en el grupo.
- El docente asigna a cada uno de los alumnos la calificación, fundamentando en todos los casos cuáles son sus elementos de juicio. Esta circunstancia convierte el examen en una instancia de feed-back muy valiosa para el estudiante que puede, con ayuda del profesor, percibir más nítidamente cuáles son sus falencias y cuáles son sus puntos fuertes.

Se estima que un coloquio se desarrolla en un tiempo mínimo de 45 minutos. Es recomendable la presencia de un observador que registre las interacciones y que colabore en la evaluación de la dinámica del grupo.

Exámenes escritos

"Mientras que casi todos los profesores estaban de acuerdo en que es importante ayudar a los alumnos a adquirir la capacidad necesaria para identificar, analizar y resolver problemas, en los exámenes se ponía a prueba, sobre todo, la capacidad para recordar fragmentos aislados de información" ...
George Miller [12]



Preguntas de respuesta abierta: desarrollo o de ensayo.

Ejemplos:

- Explique las interacciones entre drogas antihipertensivas.
- ¿Qué relaciones se pueden establecer entre fecundidad y grado de escolaridad de las mujeres en los países desarrollados?

Los exámenes que tienen el formato de ensayo son difíciles de corregir [13] Una forma de aumentar la objetividad de las preguntas de respuesta abierta es restringir la respuesta y establecer por consenso la respuesta que será considerada correcta.

Ejemplos:

- Mencione por lo menos tres mecanismos por los cuales una droga puede cerrar un canal iónico
Clave: la respuesta correcta debe incluir los tres siguientes:
 - mediante proteína G
 - por unión directa al canal
 - bloqueando un receptor con canal intrínseco
- Paciente de 65 años que consulta por un episodio transitorio de pérdida de fuerza en miembro superior derecho y trastornos del habla que duró 30 minutos.
 - 1.- ¿Qué antecedentes investigaría en el interrogatorio? (No menos de cinco y no más de 10)
 - 2.- Tiene una TAC cerebral sin anomalías a las 24 hs, ¿Qué estudios solicitaría para determinar posibles causas del accidente isquémico transitorio? (excluyendo laboratorio) (no más de cuatro)**Clave:**
 - 1.- La respuesta debe incluir por lo menos 5 de los siguientes: Episodio similar, Hipertensión arterial, Hiperlipidemia, Enfermedad coronaria, Obesidad, Diabetes mellitus, Arritmia, Tabaquismo
 - 2.- La respuesta debe incluir los tres siguientes: Ecocardiograma, electrocardiograma, ecodoppler carotídeo.

Preguntas que requieren una respuesta breve.

Ejemplos:

- ¿Cuál es la tasa de mortalidad neonatal en la República Argentina?
- Un hombre de 65 años de edad tiene dificultad para levantarse y para enderezar su tronco pero no tiene inconvenientes para flexionar sus piernas. ¿Qué músculo es más probable que haya sufrido una lesión?

Completar espacios en blanco.

Se presenta un párrafo con algunos espacios en blanco que el alumno debe completar con una palabra o cifra de modo que el enunciado sea correcto.

Ejemplo:

- Las formas histopatológicas de cáncer mamario en las embarazadas no difieren de las halladas en las pacientes no gestantes. Predominan los carcinomas.

Clave: debe responder: ductales y lobulillares

Los espacios o líneas para contestar deben ser uniformes para que su tamaño no sugiera la respuesta. El párrafo no debe tener tantos espacios en blanco que dificulten la comprensión.

Se considera respuesta correcta si completa correctamente todo el párrafo independientemente de la cantidad de palabras-conceptos a recordar y escribir.

Establecer la correspondencia

Ejemplo:

- Establezca la correspondencia entre cada fase del potencial de acción que figura en la columna de la izquierda, con el fenómeno que la genera, que figura en la columna de la derecha. Escriba en el espacio en blanco, la letra que corresponda. Cada letra puede ser usada una, varias o ninguna vez.

Fases potencial de acción	Fenómeno que la genera
Fase 1g.....	a) Despolarización rápida
Fase 2e.....	b) Despolarización de los músculos papilares
Fase 3d.....	c) Repolarización auricular
Fase 4f.....	d) Repolarización Ventricular
Fase 0a.....	e) Plateau
	f) Potencial de reposo
	g) Repolarización rápida

Alternativas constantes (verdadero-falso)

Ejemplos:

Marque V si considera que el enunciado es verdadero /correcto y F si falso o incorrecto.

- La válvula aórtica es la que más frecuentemente se afecta con la fiebre reumática V F

<i>Las siguientes son enfermedades recesivas ligadas al cromosoma X.</i>	V	F
1. Hemofilia A (hemofilia clásica)	x	
2. Fibrosis quística		x
3. Distrofia muscular de Duchenne	x	
4. Enfermedad de Tay-Sach		x

Las opciones deben ser totalmente correctas y/o falsas sin ninguna duda. Se recomienda redactar más enunciados falsos ya que se ha demostrado que tienen mejor discriminación que los enunciados verdaderos.

Preguntas de selección múltiple

Es un tipo de ítem muy utilizado, generalmente se identifica como "choice". En las publicaciones internacionales se encuentran con la sigla MCQ (multiple choice question) y la sigla en castellano es PEM (pregunta de elección múltiple).

Estas preguntas se caracterizan por tener un enunciado o base y 3-4 opciones breves [14] y concisas. Una y solamente una de las opciones es la clave (respuesta correcta).

La evaluación de los conocimientos adquiridos no se limita a explorar el recuerdo de datos específicos sino que incluye otros niveles taxonómicos como comprensión, razonamiento, análisis y síntesis, aplicación.

Con frecuencia se escucha decir que las pruebas de selección múltiple sólo evalúan memoria. Esta observación es correcta. La mayoría de los exámenes estructurados elaborados por los docentes demandan pura memoria. Pero esto se debe a la falta de habilidad de los docentes para construir preguntas más elaboradas, no se trata de una limitación intrínseca al tipo de ítem. Los docentes que construyen pruebas estructuradas que exploran sólo memoria frecuentemente también formulan preguntas sólo informativas en los exámenes orales.

Ejemplos:

- ¿Cuál es, en la población general, la incidencia de las cardiopatías congénitas? (enunciado o base)
 - a) 1% (clave)
 - b) 3%
 - c) 5%
 - d) 10%
- ¿Cuál de los siguientes es el factor de riesgo más importante para enfermedad coronaria?
 - a) **Dislipidemia**
 - b) Tabaquismo
 - c) Diabetes

Nivel taxonómico: memoria de dato específico

En el enunciado se pueden incluir imágenes (gráficos, placas RX, trazados ECG, etc), los resultados de un estudio, los datos de un paciente (viñeta). El enunciado debe incluir toda la información necesaria y una pregunta bien concisa; las opciones deben ser más bien breves.

Es frecuente encontrar preguntas de selección múltiple que carecen de pregunta.

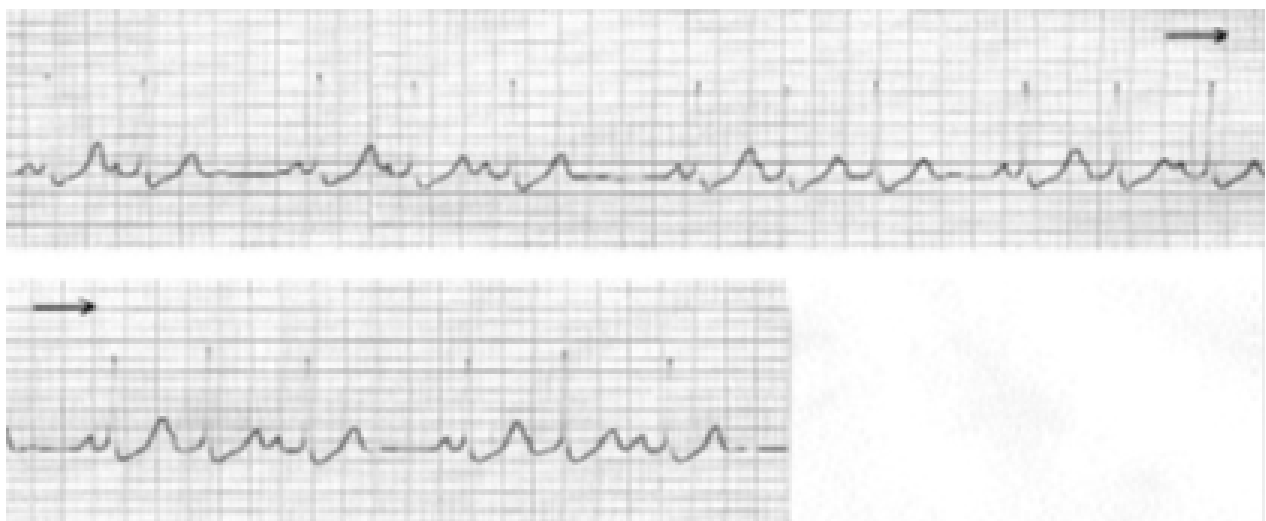
Por ejemplo: enunciado incompleto, falta pregunta.

- Número de hijos es una variable:
 - a) Cualitativa dicotómica
 - b) Cuantitativa discreta
 - c) Cuantitativa continua

El estudiante al terminar de leer el enunciado debe poder pensar la respuesta sin necesidad de leer las opciones. En el ejemplo anterior la redacción de la pregunta debería ser: ¿Qué tipo de variable es el número de hijos?

Ejemplo: enunciado con imagen.

- ¿A qué arritmia corresponde el siguiente trazado?



- a) taquicardia supraventricular con bloqueo AV de 2° grado.
- b) bloqueo sinoauricular de 2° grado.
- c) taquicardia sinusal con bloqueo AV de 2° grado (Wenckebach).
- d) extrasistolia supraventricular

Nivel taxonómico: comprensión, traducción, interpretación.

Ejemplo con enunciado con viñeta clínica

• Un niño de 2 años de edad presenta edema desde hace 1 semana. Su presión arterial es de 100/60mm Hg y presenta edema generalizado y ascitis. Las concentraciones de suero son: creatinina 0,4 mg/dL y colesterol 569mg/dL. El análisis de orina presenta proteína 4+ y no presenta sangre. ¿Cuál es el diagnóstico más probable?

- a) Glomerulonefritis postestreptocócica aguda
- b) Síndrome urémico hemolítico
- c) Síndrome nefrótico con cambios mínimos
- d) Púrpura de Schönlein-Henoch con nefritis

Nivel taxonómico: aplicación, análisis y síntesis.

GUÍA PARA LA ELABORACIÓN DE PREGUNTAS DE MÚLTIPLE ELECCIÓN [15]

Adaptado de:

- Developing and Validating Multiple-Choice Test Items, Thomas M. Haladyna, Third Edition, Lawrence Erlbaum Associates Inc Publishers, 2004, New Jersey. Chapter 5, Guidelines for developing MC items, pages 99-100.
- Pauta de comprobación para pruebas escritas, Ana Cecilia Wright. Diplomado en Educación Médica. Escuela de Medicina PUC, 2008.

Contenido de la pregunta

1. Está directamente relacionado con los objetivos del curso.
2. Evalúa un contenido importante que debe ser aprendido.
3. Se basa en evidencia científica (no opiniones).
4. En la redacción se usan expresiones diferentes a las que aparecen en los textos o apuntes.
5. Las preguntas son independientes entre sí.
6. Está formulada en forma correcta y transparente (no capciosa).

Estilo y formato de la pregunta

7. El vocabulario y las abreviaturas son de conocimiento de los estudiantes.
8. La puntuación está usada correctamente. Las mayúsculas están usadas correctamente. La ortografía está usada correctamente (incluye tipeo).
9. Hay concordancia gramatical entre el enunciado y cada una de las opciones.
10. Su lectura consume poco tiempo. El tamaño de la letra es el adecuado.

Enunciado

11. La información es relevante e indispensable para la respuesta.
12. La formulación es clara y precisa.
13. El enunciado es tan breve como sea posible.
14. La idea principal está incorporada en el enunciado (no en las opciones).
15. Los términos “Más correcto”, “Más adecuada”, “Más probable” o “Mejor” están escritos en forma destacada (mayúscula o negrita o subrayado).
16. Las palabras No, Excepto, Falso, Incorrecto o Menos correcto están ausentes o debidamente justificados.

Opciones de respuesta

17. Tiene tres o cuatro opciones (no cinco).
18. Una de las opciones es inequívocamente la correcta (clave) .

19. Los distractores son creíbles o convincentes.
20. Son independientes entre si.
21. Son homogéneas en contenido y presentan aproximadamente la misma extensión.
22. Presentan la misma precisión y la misma complejidad.
23. Presentan la misma estructura gramatical.
24. “Ninguna de las anteriores” está ausente o debidamente justificado.
25. “Todas las anteriores” está ausente o debidamente justificado.
26. Palabras como No o Excepto están ausentes.
27. Al inicio de las opciones no se repiten palabras.
28. Están ordenadas (orden lógico o numérico).
29. La ubicación de la respuesta correcta (clave) cambia al azar.

Últimamente se ha desarrollado otro formato de pregunta para evaluar, específicamente, el razonamiento clínico. El Script Concordance Test surge como un modelo novedoso de valoración del razonamiento clínico a partir del modelo clásico hipotético - deductivo y de la denominada “teoría de Script”. Esta teoría afirma que los conocimientos están archivados en la memoria como “redes” - “estructuras cognitivas”- que van siendo cada vez más complejas a medida que avanza la experiencia clínica. Estas estructuras cognitivas se activan ante cada caso clínico concreto y orientan la selección e interpretación de las nuevas informaciones. Estas matrices de experiencia interiorizada (guiones o scripts) facilitan la captación e interpretación de los datos que se produce en forma automática, ahorran el análisis paso por paso y el profesional con experiencia, frente a una situación determinada, puede llegar a una respuesta intuitivamente y a gran velocidad ("conocimiento experto condensado"). El objetivo del Script es evaluar la organización del conocimiento en redes conceptuales y la capacidad de activación de estas redes en la toma de decisiones. La redacción de script requiere un entrenamiento especial y la colaboración de 10 expertos para elaborar la clave de respuestas aceptables [16].

Construcción y aplicación de un examen escrito.

La construcción de un examen escrito requiere un importante esfuerzo. En la práctica se observa que suelen distribuirse los temas entre los docentes (según preferencias o especialidades), cada uno redacta preguntas, luego se juntan todas en un mismo cuestionario y se hacen las fotocopias. Esta modalidad de trabajo suele producir exámenes con sesgos por temas con mayor número de preguntas – según interés y dedicación del docente- o ausencia de otros temas porque el docente no entregó a tiempo sus preguntas. También suelen aparecer, en un mismo examen, preguntas con distinta cantidad de opciones y otros defectos.

La elaboración de un examen debe ser un trabajo colaborativo; la cantidad y calidad de las preguntas debe ser analizada por un grupo de docentes/colegas.

Esta tarea implica una serie de pasos:

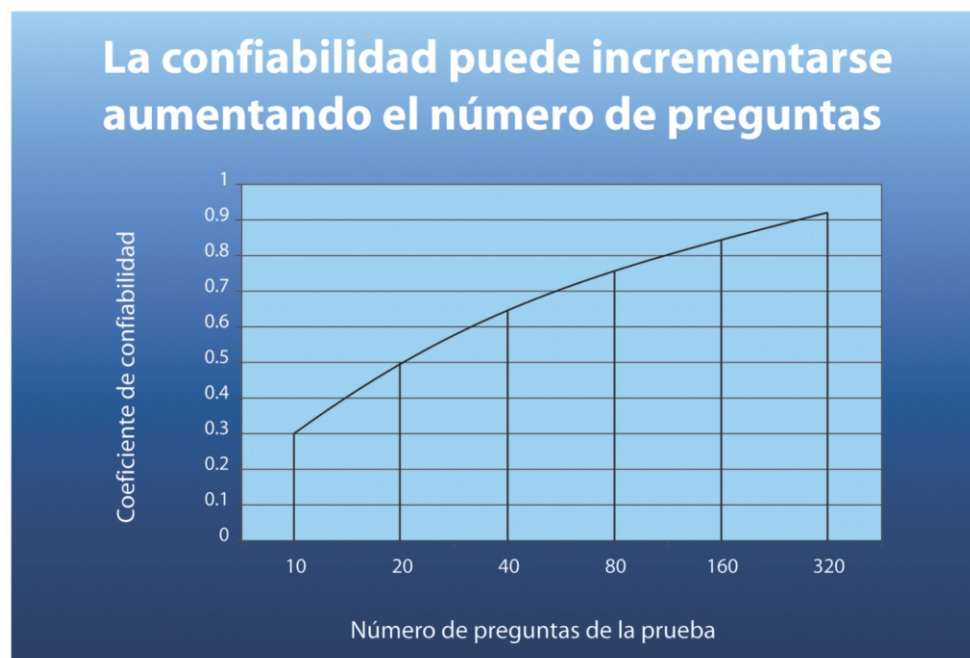
- **Construcción de la Tabla de especificaciones.** Objetivos, temas que se van a evaluar y cantidad de preguntas en cada uno de los temas, formato de las preguntas, cantidad de preguntas según nivel taxonómico.

Niveles taxonómicos	Conductas
Memoria	nombrar, identificar, enumerar, definir, describir
Comprensión	Interpretar datos, ejemplificar, clasificar, comparar, resumir
Aplicación	analizar datos y sacar conclusiones, tomar decisiones (indicar estudios, prescribir un tratamiento), justificar.

Ejemplo: Tabla de especificaciones.

Temas	Memoria	Comprensión	Aplicación	Totales
A	0	10	30	40
B	5	5	25	35
C	10	10	5	25
Totales	15	25	60	100

Es importante establecer el total de preguntas que tendrá el examen ya que la confiabilidad de los resultados depende de la cantidad de preguntas. Para una confiabilidad de 0,7 la prueba debe tener por lo menos 80-100 preguntas.



Fuente: Juárez Díaz González N, García Carrasco F. Taller examen escrito. <http://www.conacem.org.mx/assets/11.pdf>

- **Revisión de cada una de las preguntas** (enunciado-opciones y clave) por un grupo de colegas. Aquellas preguntas en las que no hay consenso no deben ser incluidas en el examen. Existe un instrumento - escala Galofré [17] - para evaluar la calidad de las preguntas antes de ser incluidas en una prueba.
- **Imprimir las copias, una para cada alumno.** Se recomienda registrar las respuestas en formularios que permitan la lectura con scanner lo que facilita la tabulación inmediata de los resultados. Puede optarse por el soporte electrónico, requiere que cada alumno tenga acceso a una PC el día y hora del examen (si se hace en forma sincrónica). Se necesita el apoyo de un programador.
- **Analizar los resultados.** Una vez corregidos los exámenes: se analizan cuidadosamente los resultados en términos de rendimiento de los estudiantes y se extraen conclusiones sobre la calidad del instrumento aplicado. Se debe revisar la distribución de las respuestas en las distintas opciones lo que permite identificar preguntas que “funcionaron mal”, que son defectuosas o que la clave de respuesta que se dio como correcta es dudosa. Es recomendable anular las preguntas que presentan estos defectos. El análisis de los resultados de un examen de rendimiento preparado por el profesor brinda información sobre la calidad del instrumento de evaluación [18]. Estos análisis se pueden realizar a partir de un instrumento administrado a por lo menos 30 alumnos.

• **Informar los resultados a los interesados.** No se trata de informar qué puntaje obtuvo y si aprobó o no. Discutir con los alumnos los resultados del examen revisando cada una de las preguntas y las opciones que fueron más elegidas – a veces lo es una opción incorrecta- es una excelente actividad de aprendizaje. Es una forma de hacer feed back. Toda persona que rinde un examen tiene derecho a revisar su prueba para identificar sus aciertos y errores.

Criterios de aprobación. Nivel de exigencia.

La decisión que los profesores toman más frecuentemente en base a los resultados obtenidos en los exámenes es la promoción o no de los alumnos.

¿Cuántas preguntas deben ser respondidas correctamente para aprobar el examen? En la literatura este tema se conoce como “standard setting” y se mencionan métodos tales como Ebel, Angoff y Hofstee [19].

Existen diferentes criterios para decidir el nivel de exigencia para la aprobación de un examen [20].

Criterio absoluto

El rendimiento de cada uno de los alumnos se compara con un estándar deseable preestablecido. Teóricamente el estándar lo constituyen los objetivos de aprendizaje establecidos como "resultados esperados del aprendizaje" y tomados como base para la planificación del examen. ¿Son todos los objetivos esenciales? ¿Qué porcentaje de "desconocimiento" o equivocación podría admitirse? ¿Cuál es el mínimo indispensable?

En respuesta a estas preguntas es que los profesores fijan el nivel de aprobación: por ejemplo 60% del puntaje máximo posible.

Por ejemplo: si administramos una prueba de 80 preguntas con un nivel de exigencia de 60%, y cada pregunta vale un punto, el puntaje máximo posible es de 80 puntos y el puntaje mínimo para aprobar es 48 puntos (60% de 80). Todos los alumnos que obtengan 48 puntos y más estarán aprobados y todos aquellos que obtengan menos de 48 estarán aplazados en el examen.

Criterio relativo

El puntaje necesario para aprobar se establece después de corregir todas las pruebas y conociendo el puntaje más alto obtenido. El nivel de exigencia para aprobar se establece de acuerdo al mejor puntaje obtenido. Por ejemplo: en el mismo caso anterior- examen de 80 preguntas- el puntaje mas alto obtenido es 69 correctas, el nivel de exigencia del 60% significa que el examen se aprueba con 41 correctas (60% de 69). cantidad de alumnos que aprueban varía según se utilice un criterio u otro.

Lógicamente la cantidad de alumnos que resultan aprobados/aplazados varía según se utilice un criterio u otro. En ambos casos es importante determinar, antes de establecer la línea de corte, cuál es el puntaje máximo posible: después de corregida la prueba se debe hacer un análisis cuali-cuantitativo de cada una de las preguntas y decidir si hay que anular alguna pregunta lo que necesariamente modifica el puntaje máximo posible. Por ejemplo: si un examen tiene 80 preguntas y tres de ellas resultan deficientes y son anuladas el puntaje máximo posible pasará a ser 77 y el nivel de aprobación, con criterio absoluto, será 46 correctas.

Adjudicación de notas. Sistemas de calificaciones.

Generalmente para informar al estudiante, a los padres y a la sociedad sobre el rendimiento de los estudiantes se utilizan "las notas". Y las notas se toman en cuenta para decisiones importantes: el promedio de la carrera suele tomarse como uno de los criterios para la admisión al sistema de residencias, para otorgar una beca, para postular a un empleo, etc.

Existen distintos Sistemas de Calificaciones que varían en la cantidad de "categorías o clases" en las que se pueden ubicar a los alumnos según su rendimiento.

- Sólo dos categorías: insuficiente - suficiente; aprobado - desaprobado
- Tres categorías: insuficiente - suficiente - distinguido
- Cinco categorías: A-B-C-D-E
- Diez categorías: escala decimal, de 0 a 10 .

A	Excelente
B	Muy bueno
C	Bueno- termino medio
D	Regular
E	Deficiente-malo

Con cierta frecuencia se combina la escala decimal con categorías cualitativas y resulta la siguiente combinación.

Sobresaliente	10
Distinguido	9
Muy bueno	8-7
Bueno	6-5
Regular	4
Aplazado	3-2-1-0

El sistema de calificaciones escolares es una convención, un código que está muy lejos de expresar con exactitud "una medida del conocimiento".

Chile califica con una escala del 0 al 7; Argentina con una escala del 0 al 10 y Venezuela utiliza del 0 al 21. Y la nota que significa "aprobado" varía en cada caso: Venezuela aprueba con el 7 que es el valor más alto de la escala chilena. Argentina aprueba con 4 que en Venezuela significa aplazado. Los sistemas de calificaciones son convenciones y representan ciertas tradiciones.

"Como cualquier otro aspecto profundamente implicado en la cultura, los procedimientos de calificación y las calificaciones a menudo se han considerado cosas ya hechas y sabidas, con un mínimo de análisis racional acerca de su naturaleza y sus funciones. A medida que las examinamos, nuestro enfoque debería ser, por lo menos en parte, el del antropólogo cultural que fija su interés en una serie de patrones de conducta que le parecen raras y presumiblemente significativas, y que trata de entender las funciones para las que sirven y la manera en la que se relacionan con la cultura de la que son parte. Debemos hacer a un lado nuestro compromiso personal y mirar estos fenómenos con los fríos ojos del científico social, con el fin de tratar de identificar las fuerzas que moldean y sustentan las prácticas de calificación actuales y las presiones dentro de la cultura educativa que hace de estas prácticas algo tan irracional y resistente al cambio" Thorndike R. Hagen E [21].

Múltiples factores intervienen en la adjudicación de las notas por parte de los docentes. Algunas investigaciones han demostrado que:

- A igualdad de rendimiento y capacidad las niñas reciben mejor notas que los varones.
- Los alumnos más aceptados por sus maestros, a igualdad de rendimiento y capacidad, obtienen calificaciones más altas que los alumnos menos aceptados.

- Las maestras tienden a asignar mejores notas a las niñas que a los varones y a ser menos generosas que los maestros.

De modo que una nota alta o baja puede estar relacionada hasta con el sexo de quien califica. Cuando los maestros y profesores adjudican notas, el valor de las mismas depende de lo que ellos *creen* debe ser aprendido en su curso.

Una de las críticas más serias al sistema de calificaciones en vigencia es que en un sólo símbolo se incluye una gran variedad de objetivos cumplidos. Una única nota jamás podrá ser un buen indicador de logros tan diversos. A pesar de todas estas limitaciones un sistema de calificaciones es necesario para comunicar los resultados.

Cualquiera sea el sistema que se utilice (cualitativo, alfabético, decimal, etc.) las "notas" informan:

- Al alumno sobre su propio rendimiento.
- A los padres sobre el progreso de sus hijos.
- A los maestros sobre la calidad de su enseñanza.
- Al futuro empleador sobre la capacidad de los postulantes.

Conclusiones

- La evaluación es un componente esencial de la enseñanza.
- Los exámenes son un poderoso motivador: los alumnos aprenderán lo que creen que el docente considera valioso y que preguntará en las evaluaciones
- La mayoría de los instrumentos para recoger evidencias sobre los conocimientos y actitudes de los alumnos son elaborados por los propios docentes.
- La evaluación diagnóstica brinda información para ajustar la enseñanza al nivel de los alumnos y promover un aprendizaje significativo.
- La evaluación formativa orienta el proceso de aprendizaje. Su objetivo fundamental es proporcionar a los estudiantes feedback sobre su aprendizaje y facilitar la reflexión sobre su progreso.
- La evaluación sumativa verifica si el sujeto reúne los requisitos necesarios y suficientes para aprobar un curso, para recibir el título, para ganar una beca. Los docentes certifican la competencia de los egresados.
- Los resultados de la evaluación brindan información sobre los aprendizajes logrados por los alumnos y la calidad de la enseñanza impartida.

Referencias.

[1] **Palés. J.** *Educación Médica* 2001, 4 (1).

[2] AMEE Medical Education Guide No 25: The assessment of learning outcomes for the competent and reflective physician. *Medical Teacher* 2000, 25 (6): 569-584.

[3] **Zabalza M.** ¿Qué puede aportar la evaluación para mejorar el proceso educativo? *Revista Novedades Educativas* 2002; Año 14. N° 135.

[4] **Stufflebeam D, Shinkfield A.** *Evaluación Sistemática, Paidós. Buenos Aires 1987.*

- [5] **Miller G.** Los principios en la práctica en Estrategias Educativas para las profesiones de la salud. *Cuadernos de Salud Pública N° 61. OMS. Ginebra. 1975.*
- [6] **Durante Montiel, Lozano Sánchez, Martínez Gonzáles, Morales López, Sánchez Mendiola.** Evaluación de Competencias en ciencias de la Salud. *Panamericana. México. 2011 . Ver capítulo 1.*
- [7] **García Barbero M, Alfonso Roca MT, Cancillo Salas J, Castejón Ortega.** Planificación educativa en ciencias de la salud. *Massón. Madrid 1995. Ver capítulo 3.*
- [8] **Durante Montiel, Lozano Sánchez, Martínez Gonzáles, Morales López, Sánchez Mendiola.** Evaluación de Competencias en ciencias de la Salud. *Panamericana. México. 2011 . Ver capítulo 2.*
- [9] **Fornells-Vallés JM.** El ABC del Mini-CEX. *EDUC MED 2009; 12 (2): 83-89.*
- [10] **Ejemplo tomado de Case SM y Swanson DB.** Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. *Disponible en <http://www.nbme.org/pdf/iwg-sp/iwg-spanish2006.pdf>.*
- [11] **Galli A, de Hernández NC, Galperín J.** Una técnica no convencional para la evaluación del aprendizaje: el coloquio. *Revista de Planificación Universitaria, N° 5, junio 1978, Barquisimeto, Venezuela. Disponible en www.saidem.org.ar*
- [12] **Miller G.** Los principios en la práctica. *Cuadernos de Salud Pública N° 61 Estrategias Educativas para las profesiones de la Salud. OMS. Ginebra. 1975.*
- [13] **Torrubia R y Pérez J.** La difícil objetividad de las pruebas de ensayo en la evaluación del rendimiento académico. *Educación Médica 2005; 8(1): 17-21.*
- [14] Cantidad de opciones. **Ver Vyas R, Supe A.** Multiple choice questions: a literature review on the optimal number of options. *Ntl Med J India. 2008; 21(3):130-3. Disponible en <http://www.ncbi.nlm.nih.gov/pubmed/19004145>*
- [15] Documento elaborado por Ana Cecilia Wright. Universidad Católica de Chile.
- [16] **Galli A, Gimeno G.** Evaluación de competencias clínicas. *SCRIPT. Concordance test. Disponible en www.saidem.org.ar*
Resumen y adaptación de: **Brailovsky C.** *SCRIPT. Concordance test. Monografía original Carlos Brailovsky: <http://www.script.md/site/es/monografia.html>.*
- [17] **Galofré A.** Preguntas de selección múltiple. Índice de calidad. *Disponible en www.saidem.org.ar*
- [18] **Case SM y Swanson DB.** Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Sección IV. *Disponible en <http://www.nbme.org/pdf/iwg-sp/iwg-spanish2006.pdf>.*
- [19] **Case SM, y Swanson DB.** Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Capítulo 9. *Disponible en <http://www.nbme.org/pdf/iwg-sp/iwg-spanish2006.pdf>.*
- [20] **Norcini JJ.** Setting standards on educational tests. *Med Educ 2003; 37:464-469.*
Cusimano MD Standard setting in medical education. *Acad Med 1996;71(Suppl.10):S112-S120.*

[21] **Thorndike R y Hagen E.** Medición y evaluación en Psicología y educación. *Trillas. México. 1996.*

Lecturas recomendadas.

Castro C, Galli A. (2002) Evaluación Educacional. Módulo 2 : Construcción y aplicación de exámenes. AFACIMERA. Buenos Aires. Disponible en www.saidem.org.ar

Epstein RM. Assessment in medical education. *New Engl J Med* 2007; 356:387–396.

Hays R. Assessment in medical education: roles for clinical teachers. *The Clinical Teacher* 2008; 5: 23-27.

Lambert WT, Schuwirth, Van der Vleuten CPM. Written assessment. ABC of learning and teaching in medicine. *BMJ* 2003; 326 (22): 643-645.

Ruiz Moral R. Educación Médica. Manual práctico para clínicos. *Panamericana. Madrid. 2010.*

Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education* 1996;1:41-67.