La evaluación de los conocimientos: Lo que parece ser, ¿es realmente lo que es?

Eduardo Durante

Como se desprende dramáticamente del artículo del diario Clarín (fig. 1), la evaluación del conocimiento parece ser un poco más compleja que lo que habitualmente consideramos, aunque no tanto como nos podríamos imaginar. En este artículo, se pretende trasmitir las bases metodológicas para planear una adecuada evaluación de los conocimientos así como desmitificar lo complejo del tema.
Actualmente, se le atribuye una importancia muy alta en la educación, ya que numerosas investigaciones indican que la forma de evaluar el conocimiento ejerce una influencia sustancial en el cambio de los estilos de aprendizaje de los alumnos[1].
Se escucha decir en ocasiones, que los alumnos deberían estudiar para "saber" y no para aprobar los exámenes. Dado el estado actual de la investigación, parecería que esto no es así. Los alumnos en general estudian para "pasar los exámenes" y por otra parte, lo que no está en los exámenes es percibido por ellos como de no mucha importancia. Por lo tanto, dentro de la planificación curricular la evaluación del conocimiento debe ser concebida como un poderoso instrumento para orientar el estilo de aprendizaje de los alumnos y ser incluida desde el inicio del diseño curricular y no como algo que ocurre después que se decidieron los contenidos y las actividades de enseñanza-aprendizaje. No se puede pretender que los alumnos "estudien" para aprender habilidades de razonamiento clínico, si la evaluación de esta compleja habilidad se realiza a través de una prueba de elección de opciones múltiples. La prioridad de los alumnos es "pasar" el examen y en función de esta necesidad es que orientarán sus esfuerzos para adquirir los conocimientos[1,2].

BASES CONCEPTUALES

Los médicos y otros profesionales de la salud, dado su entrenamiento en estadística, metodología de la investigación y epidemiología clínica, manejan conocimientos relacionados con la evaluación clínica. Desde este punto de vista, hacer un diagnóstico es una forma de evaluación, ya que a través de la recolección de información por variados medios y de su interpretación, se puede llegar a realizar un juicio clínico con diferentes grados de incertidumbre para diagnosticar una enfermedad o condición.
Afortunadamente, una parte importante de esta base de conocimientos puede ser "transferida" al área de la evaluación de los conocimientos. Para lograr ese fin, utilizaremos una aproximación donde se facilite la transferencia a través del uso de analogías entre lo "médico" y lo "educativo".
Esto es posible porque el paradigma subyacente es el mismo: la psicometría. En este paradigma, las pruebas intentan conocer la verdad (rasgos, *traits* o *constructs*) que está de alguna manera oculta en el fenómeno en estudio (la enfermedad en el paciente, el conocimiento en el alumno).

¿CÓMO DIAGNOSTICAR LA ENFERMEDAD "INCOMPETENCIA PROFESIONAL"?

Una manera de mirar a la evaluación es verla como una medida de la competencia profesional[3]. Se podría, entonces, considerar a los exámenes o pruebas de conocimiento (de hecho se usa la misma palabra "examen" o "prueba" en ambos campos) como una herramienta diagnóstica para la enfermedad "incompetencia médica". Uno podría pensar que los alumnos que aprueban un examen son competentes mientras que los que no lo hacen, son incompetentes. Como sabemos, esta es una verdad a medias. Como en el caso de los estudios clínicos, cada examen puede producir resultados falsos positivos (desaprueban el examen pero son competentes) y falsos negativos (aprueban el examen pero son incompetentes). En la evaluación, sobre todo la sumatoria en la que se promueve a un curso superior o se otorga un certificado profesional de maestría en un área, ambos errores tienen serias consecuencias:

1. Si pasa un estudiante que tiene la enfermedad, podría existir un serio riesgo para la sociedad, dando a su vez una mala reputación a la Universidad que otorga el título o permite el ingreso a una institución (como el caso del panadero de San Pablo, Brasil, de la figura 1).
2. Si no aprueba un estudiante sin la enfermedad, entonces se puede condenar a alguien a perder un tiempo de

Recibido: 3 jun. 2005 Aceptado: 13 jun. 2005

Comité de Investigación en Educación Médica. Hospital Italiano de Buenos Aires. Correspondencia: eduardo.durante@hospitalitaliano.org.ar

su vida, a veces un año, con la consiguiente desmotivación.

Se trata, por lo tanto, como en la medicina, de minimizar los resultados falsos positivos y negativos. Una manera de mirar a la "capacidad diagnóstica" de los exámenes de conocimiento, es calculando y conociendo su confiabilidad, validez e impacto educacional.

LA ECUACIÓN DE LA UTILIDAD DE LOS EXÁMENES DE CONOCIMIENTOS

En la práctica, las decisiones sobre qué tipo de exámenes utilizar no sólo se toman en función de las variables mencionadas más arriba sino también de las opiniones, sentimientos y tradición educativa de los profesores, así como el costo del proceso.

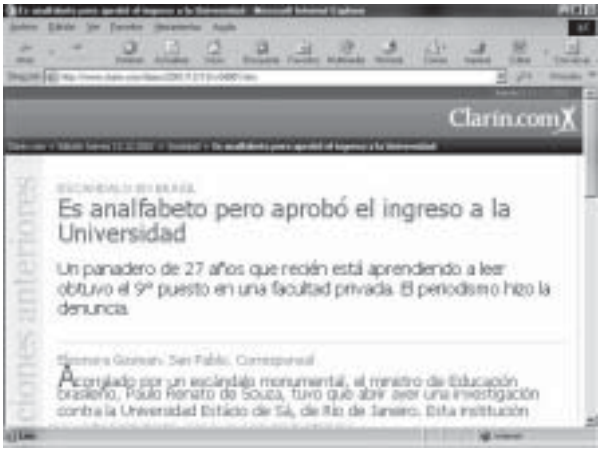
Para decidir cuál o cuáles son los métodos de evaluación más adecuados, usaremos la noción de utilidad de un examen en donde se vinculan estas variables y se les da diferente peso[2]. Así:

UTILIDAD = CONFIABILIDAD X VALIDEZ X IMPACTO EDUCACIONAL X ACEPTABILIDAD X 1/COSTO

Para poder "calcular" la utilidad de un examen con esta ecuación, es necesario definir cada una de las variables.

1. *Confiabilidad*: La confiabilidad se refiere a la precisión de la medición o la reproducibilidad del puntaje obtenido con el instrumento[3-5]. Haciendo una analogía con la medicina, se pretende que cuando se mida la glucemia en una muestra de sangre, el resultado sea consistente a lo largo del tiempo. Es decir, si en una primera medición es 140 mg% y luego 110 mg% en la misma muestra, la confiabilidad es baja y no se pueden sacar conclusiones diagnósticas sobre si el paciente tiene diabetes o no. En educación es similar: los puntajes de la misma prueba a lo largo del tiempo deberían medir la misma "cantidad" de conocimiento. En este caso, no podremos hacer juicios sobre si el alumno tiene la enfermedad "incompetencia profesional" o no. O sea, los resultados de los exámenes deben ser *reproducibles*.

Figura 1. URL: http://www.clarin.com./diario/2001/12/13/s-04001.htm



En la teoría clásica de la evaluación (teoría psicométrica), el coeficiente de confiabilidad se expresa en una escala de 0 (nada de confiabilidad) a 1 (confiabilidad total) como indicador de precisión. Este coeficiente puede interpretarse como un coeficiente de correlación entre esta medición y una re-medición hipotética tomada bajo condiciones similares en un tiempo no muy alejado[2-3]. Si bien la interpretación de los índices de confiabilidad requiere ciertos conocimientos psicométricos, se considera que un nivel arbitrario generalmente aceptado del índice de confiabilidad para evaluaciones sumativas o terminales ("de certificación de competencia", como el final de un ciclo de materias o profesional) es de 0,80. El problema con estos métodos es que es necesario "tomar" dos veces el examen. Para solucionar esto, se han diseñado métodos estadísticos que permiten desdoblar el examen en dos mitades y compararlos como si hubieran sido administrados en dos momentos diferentes. Esos métodos son: el de formas equivalentes (con o sin intervalos de tiempo) y el método de *split-half* (se divide por dos el mismo examen y se correlacionan ambas partes con el método de r de *Spearman*).

La confiabilidad también puede entenderse como la medida en que los ítems individuales de un examen (por ejemplo, cada pregunta en un *multiple-choice*) se comportan de manera similar (co-varían) al interior de un examen en relación al desempeño global del sujeto en ese examen. Esta determinación se realiza con métodos estadísticos, de los cuales uno de los más conocidos son el Kuder-Richardson o KR-20. Este método se aplica a las pruebas que tienen ítems con respuestas dicotómicas, tipo correcto-incorrecto. Cuando las respuestas tienen más de una respuesta (escalas nominales tipo Likert, o un puntaje) se utiliza el conocido método de alfa de Chronbach. En este caso, el coeficiente también varía entre 0 y 1[2-3].

Cuando la medición está a cargo de dos observadores diferentes, se puede determinar la confiabilidad inter observador. Cuando la medición está a cargo del mismo observador en dos ocasiones diferentes, se considera la confiabilidad intra observador. En ambos casos, se realiza una correlación y varía entre 0 y 1[2-3].

La confiabilidad intercasos constituye un caso especial de confiabilidad[6]. Existe gran cantidad de evidencia que indica que la confiabilidad de las mediciones

sobre las competencias clínicas se ve obstaculizada por el hecho de que las competencias son específicas según el contenido o área[5-6]. Esto significa que el hecho de lograr una competencia óptima en un área no es un buen predictor de competencia en otra, aún en el caso de que dichas áreas se encuentren muy relacionadas. Esto tiene que ver con que la adquisición de competencias es específica para contenidos diferentes aunque parezcan similares (la competencia del examen físico en un paciente con insuficiencia cardíaca no predice el desempeño en el examen físico en un paciente con fibrosis pulmonar). La consecuencia de este fenómeno es que para alcanzar resultados confiables en las pruebas de evaluación se requieren muestras relativamente numerosas de las diferentes competencias clínicas (entrevistas, examen físico, interpretación de pruebas diagnósticas etc.), lo que aumenta considerablemente el tiempo de evaluación (por ejemplo, para que los resultados de un ECOE (Examen Clínico Objetivo y Estructurado)[14] sean confiables se requieren alrededor de 13 estaciones de entre 10 y 15 minutos de duración lo que lleva el tiempo de examen a 4 ó 5 horas)[6]. Como consecuencia de esto, hay que tener en cuenta algunas recomendaciones al seleccionar o diseñar exámenes de evaluación del desempeño:

a. En general, a mayor número de ítems en un examen, mayor será la confiabilidad. Esto es debido a que en los exámenes más largos, la probabilidad de error aleatorio disminuye como consecuencia de que existe un muestreo más amplio de las conductas o conocimientos a evaluar. O sea, que las diferencias en los puntajes de los exámenes de los alumnos se deben a diferencias reales en el desempeño de los alumnos y no a diferencias por azar. En el caso de la medicina, si un paciente tiene varias determinaciones mayores a 126 mg% de glucemia es porque es diabético, siempre y cuando la confiabilidad del método sea alta (cercana a 1).
   Recuerden el examen final clásico de Medicina en donde se toman decisiones sobre la competencia de los alumnos a partir de un solo caso. ¿Qué porcentaje de falsos negativos y positivos tendrá? Posiblemente muy alto ya que adolece de falta de confiabilidad intercasos (obviamente es un solo caso) e inter observador (que se hace patente en las llamadas mesas examinadoras con especialistas o profesionales de distintas áreas y sin una estructuración mínima del examen).
b. La objetividad de un examen se refiere al grado en el que observadores con el mismo entrenamiento en el uso del instrumento de evaluación obtienen los mismos resultados (alta confiabilidad inter observador). Los exámenes de ítems de elección de opciones múltiples (los conocidos *multiple-choice*) permiten que la opinión de los observadores no altere los puntajes. Cuando se utilizan estos exámenes, la confiabilidad de los puntajes no se ve afectada por los procedimientos de corrección. En los exámenes más habituales (como preguntas a desarrollar) la confiabilidad puede verse amenazada por la baja correlación inter observador. Sin embargo, esto no quiere decir que deberían abandonarse en aras de adoptar exámenes con una mayor confiabilidad, sino que se deberían utilizar como parte de la estrategia de enseñanza y más que nada por su impacto educacional (ver más adelante).

En la tabla 1, se puede observar el uso de exámenes en cuanto a su confiabilidad de acuerdo al tipo de decisiones a adoptar según sus resultados.

Tabla 1: Demandas de Confiabilidad y naturaleza de la Decisión (Modificado de Linn R & Gronlund N[3]).

Alta confiabilidad es necesaria cuando:
- La decisión es importante.
- La decisión es final (en un curriculum, una carrera).
- La decisión es irreversible.
- La decisión no se puede confirmar.
- La decisión está relacionada con individuos.
- La decisión tiene efectos duraderos (por ejemplo, selección de residentes).

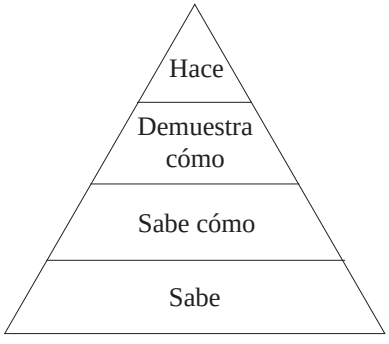
Baja confiabilidad es tolerable cuando:
- La decisión es de menor importancia.
- La decisión se toma en estadios tempranos del currículo (evaluación formativa).
- La decisión es reversible.
- La decisión se puede confirmar por otras fuentes.
- La decisión está relacionada con grupos.
- La decisión tiene efectos temporarios (por ejemplo, revisar un tema del curso).

2. *Validez*: se refiere a diferentes dimensiones que se resumen con el término validez. Así, podemos hablar de diferentes tipo de validez[3,5-6]:
   a. Validez de concepto (construct): muestra hasta qué punto el método mide lo que se pretende que mida. Si se pretende medir el nivel de hemoglobina de un paciente y el método mide el recuento de leucocitos, entonces el método no es válido ya que no se podrá establecer si el paciente está anémico o no. En Educación, si se pretenden medir las habilidades comunicacionales para entrevistar un paciente, un examen estructurado en base a ítems de verdadero-falso no es válido ya que no indicará claramente si el estudiante tiene las habilidades comu-

nicacionales para la entrevista clínica.

b. Validez de criterio: muestra si los resultados de un examen son concordantes con los resultados de otras pruebas ya validadas (es raro encontrar un *Gold-Standard* en educación), o si estos resultados predicen el desempeño futuro de los alumnos. Por ejemplo, en Canadá se realizó un estudio donde se observó que los médicos que obtenían mayores puntajes en un examen de certificación, tenían una práctica de mejor calidad que las de los que obtenían puntajes más bajos[8-9].
c. Validez de contenido: demuestra el grado de "muestreo" de todos los aspectos que se pretende medir. A modo de ejemplo, se puede decir que para que los instrumentos de evaluación de la competencia clínica tengan validez de contenido, es posible utilizar la Pirámide de Miller. El educador George Miller (1990)[7] definió un modelo para la evaluación del competencia profesional como una pirámide de cuatro niveles (ver fig. 2). En los dos niveles de la base se sitúan los conocimientos (saber) y decir cómo aplicarlos a casos concretos (saber cómo). En el nivel inmediatamente superior (mostrar cómo), se ubica a la competencia cuando es medida en ambientes "in vitro" (simulados) y donde el profesional debe demostrar todo lo que es capaz de hacer. En la cima se halla el desempeño (hace) o lo que el profesional realmente hace en la práctica real independientemente de lo que demuestre que es capaz de hacer (competencia).

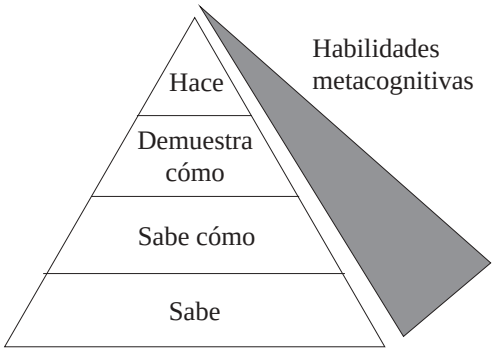
Figura 2: La pirámide de Miller G[10].



A nivel de post-grado, existe evidencia de que lo que el médico es capaz de hacer en el nivel del demuestra cómo (evaluación de habilidades clínicas en medios "simulados" de la vida real) no predice el desempeño en el nivel de "hace" (lo que realmente hace en la vida real, o sea, la competencia)[11]. Dicho de otra manera, es posible que sea un excelente especialista en el "laboratorio", pero con desempeño "pobre" en la realidad. Esto obedece al hecho de que los actos y las decisiones que se toman en el día a día en la práctica real están influidos por numerosos factores que están "controlados" en el nivel de la evaluación de la competencia pura. Tales factores pueden ser agrupados como aquéllos relacionados con el individuo (salud física y mental del profesional, el estado cognitivo en el momento de la consulta, relaciones con otras personas, incluida la familia) y aquellos factores relacionados con el sistema en el cual se desempeña el profesional (tales como programas obligatorios del Estado para aplicar en el cuidado de los pacientes, características de la organización del sistema de Salud, forma de pago, condiciones de la atención, etc.). Como se puede concluir, existe una compleja interacción entre competencia, desempeño y los factores individuales y sistémicos en el cuidado de los pacientes.

Por otra parte, la evaluación de los alumnos debería ser capaz de "medir" las llamadas habilidades "metacognitivas" que definen la percepción que el individuo tiene de por qué sabe lo que sabe (fig. 3)[2,12]. Con este nombre se conocen a las estructuras del conocimiento que son capaces de "diagnosticar" y "prescribir" tratamientos cuando existen carencias en el conocimiento. Es el sistema de monitoreo del conocimiento y del estilo que cada uno tiene para adquirirlo. Sin estas habilidades desarrolladas se vuelve dificultoso seguir aprendiendo, o sea, el desarrollo profesional continuado se vuelve dependiente de la oferta de cursos (aprendizaje centrado en el docente, o más bien, en los proveedores de educación continuada) más que de la demanda de conocimientos por parte de los especialistas en relación a resolver problemas que surgen del cuidado médico (aprendizaje centrado en el estudiante) cuando este sistema de monitoreo no está desarrollado. Existen pruebas para estimularlas y evaluarlas tales como el portafolio y el plan de Desarrollo Personal.

Figura 3: Las habilidades metacognitivas (modificado de Van der Vleuten K[2]).



3. Impacto educacional: "La evaluación guía el aprendizaje de los estudiantes"[2]. Esta afirmación ha llegado a convertirse en un nuevo paradigma en la educación como se mencionó anteriormente. Esto significa que los estudiantes (en este caso, los que quieren certificarse) buscan "pasar" los exámenes y orientarán sus esfuerzos en ese sentido más que en "aprender". La integración de los conocimientos y la adquisición de determinadas competencias se verán más o menos estimuladas de acuerdo al tipo de examen que se aplique. La administración de pruebas donde se tenga que demostrar cómo estimulará la adquisición de competencias relacionadas con ese dominio. Si la prueba es del tipo de examen de selección de opciones múltiples, el aprendizaje se orientará al dominio de sabe y a lo sumo de sabe cómo, en desmedro de demuestra cómo o hace.

   "La evaluación guía el aprendizaje a través de su formato"[2]. Las tareas solicitadas en las pruebas de evaluación deben ser lo más cercanas posibles a las que los estudiantes realizarán en la práctica real. Una situación descripta en la literatura mostró cómo la evaluación de habilidades a través de una lista de cotejo muy detallada y descontextualizada de una situación real (por ejemplo, revisar una rodilla, ver una radiografía de tórax), llevó a la trivialización de la habilidad: los alumnos comenzaron a memorizarlas para pasar el examen.

   La evaluación debe guiar el aprendizaje a través de la información brindada a los alumnos[4]. La evaluación también debe ser un ejercicio de aprendizaje a través del *feedback* de los resultados del examen y comprometiendo a los alumnos en planes de mejoramiento de su propio aprendizaje, individualmente o aún mejor, en grupos de reflexión. Además, los resultados de los exámenes son un buen insumo para un programa de mejoramiento de la calidad del programa de formación.
4. Aceptabilidad: se refiere a que tanto los evaluados como los evaluadores acepten los métodos de evaluación propuestos para el examen de certificación. Pueden aparecer resistencias a nuevas iniciativas relacionadas con la cultura del país, el modelo educacional (centrado en el docente vs. centrado en el estudiante), sistemas de entrenamiento de postgrado, etc.
5. Costo: una buena evaluación es definitivamente costosa. Los nuevos métodos de evaluación de las competencias (como puede ser el ECOE, portafolios y la observación de entrevistas reales) son métodos válidos para la evaluación de competencias en los niveles de demuestra cómo y hace. Sin embargo, su implementación es costosa por el despliegue logístico y económico que requieren. Esto puede limitar su aplicación a pesar de cumplir con todas las demás características.

   Sin embargo, invertir en evaluación es invertir en enseñanza y aprendizaje: una buena evaluación facilitará un buen aprendizaje.

De particular importancia es la estructura regulatoria, que es el marco en el que toda la información de los exámenes es recolectada. Si la hemoglobina de un paciente es baja y se asume que el paciente está anémico, esto no significa automáticamente que sufre de deficiencia de vitamina B12. Para aclarar el diagnóstico son necesarias otras pruebas, las cuales deberían estar orientadas en la misma dirección para facilitar el diagnóstico. Sin embargo, algunos diagnósticos no son tan fáciles de hacer, ya que a veces los resultados son contradictorios o no conclusivos. La tarea de suma importancia en estos casos es considerar toda la información, valorarla, considerar la sensibilidad y especificidad de cada una para establecer un diagnóstico y comenzar un tratamiento. Para la enfermedad "incompetencia" se aplican los mismos criterios. Los juicios deben estar basados en información de diferentes pruebas, prolijamente recolectada e interpretada. Esto se consigue con un programa de evaluación con una estructura regulatoria cuidadosamente diseñada.

CONSECUENCIAS PARA LA APLICACIÓN EN LOS EXÁMENES DE TODOS LOS DÍAS

1. No basar las evaluaciones en exámenes cortos: para que los exámenes sean suficientemente confiables, se debe usar una muestra amplia del mismo contenido. Las preguntas que son de desarrollo extenso producen en general exámenes menos confiables que las preguntas que permiten respuestas más cortas. Una prueba que está basada en un solo caso es poco confiable.
2. No basar los exámenes en un solo método: cada examen tiene sus propias fortalezas y debilidades. Ningún instrumento aisladamente es la panacea: el "cuadro" general sólo puede ser visto por un programa de evaluación organizado alrededor de diversos métodos cuidadosamente seleccionados y combinados, con diferente validez, confiabilidad, impacto educacional y aceptabilidad.
3. Combinar los resultados: si se utilizan varios métodos, entonces los resultados de los exámenes pueden ser combinados de dos maneras diferentes, con o sin compensación[13]:
   a. Combinación sin compensación significa que el estudiante tiene que aprobar todas las pruebas individuales para pasar el curso.
   b. Combinación con compensación total significa que el estudiante puede pasar si la media total de los puntajes es suficiente, es decir que podría sacarse

un 10 en un contenido y un 2 en otros y aún así pasar.

En la combinación con compensación parcial, la media global determina si el estudiante puede pasar, pero para cada prueba individual un puntaje mínimo tiene que ser obtenido. En el caso anterior, el estudiante no pasaría. En general, los métodos sin compensación llevan a un número alto de resultados falsos negativos (tienen la enfermedad “incompetencia” pero pasan), debido a que los déficits de confiabilidad de cada método se combinan y este error puede ser acumulativo a través de todas las pruebas de la combinación. El estudiante podría aprobar cada contenido con el mínimo puntaje, situación en la que la probabilidad de falso negativo es mayor dada la falta de discriminación por debilidad en la confiabilidad del método.

En cambio, la combinación con compensación total puede llevar a que los alumnos puedan decidir abandonar el estudio de algunos de los contenidos y compensar con el puntaje en otros llevando a un indeseable efecto negativo en el estilo de aprendizaje. La compensación parcial es un compromiso entre ambos métodos y es el que mejor refleja nuestro trabajo clínico. Aún cuando una prueba sea anormal, concluiremos que el paciente está sano si los otros parámetros son normales, pero si uno de los resultados de laboratorio está realmente muy desviado de lo "normal" (verdadero positivo), entonces asumiremos que algo malo puede estar pasando con el paciente y padecer la enfermedad ("incompetencia médica").

En el próximo número, se revisarán diferentes métodos y cómo se podrían combinar de acuerdo a las bases conceptuales descriptas en este documento.

En conclusión, la evaluación de los conocimientos no es tan compleja como parece ni tan simple como estamos acostumbrados a "padecerla". Es preciso incluir desde el inicio del diseño curricular un sistema de evaluación que sea coherente con las competencias deseadas y que oriente el aprendizaje de los alumnos en la misma dirección. Una combinación de métodos con impacto en los diferentes niveles de la pirámide de Miller[7] y con suficiente confiabilidad y validez, asegurarán a los responsables de los programas de educación, así como a los alumnos y la sociedad, que los egresados de estos programas no padecen la enfermedad "incompetencia médica". Tal vez, si en el caso de Brasil se hubieran seguido algunos de estos principios, el famoso panadero analfabeto no habría ingresado en la Facultad.

REFERENCIAS

1. Newble D, Jaeger K. The effect of assessments and examinations on the learning of medical students. Med Educ 1983, 7: 165- 171.

2.Van der Vleuten CPM. The assessment of professional competence: development, research and practical implications. Adv Health Sci Educ Theory Pract 1996, 1: 41- 67.

3. Brailovsky C. Educación Médica, Evaluación de las competencias. En: Aportes para un Cambio Curricular en Argentina. Buenos Aires: OPS - Facultad de Medicina, UBA; 2001.

4. Linn R y Gronlund N. Measurement and Assessment in Teaching. New Jersey: Prentice – Hall; 2000.

5. Van der Vleuten CPM. Validity of final examinations in undergraduate medical training. BMJ 2000, 321: 1217-1219.

6. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to their development and use. 2 ed. New York: Oxford University Press; 1995.

7. Van der Vleuten CPM, Swanson DB. Assessment of clinical skill with standarized patients: state of the art. Teaching and learning in medicine (1990).

8. Freidzon E. How dominants are the professions? En: Professionalism Reborn: Theory, Prophecy and Policy. Cambridge: Polity Press; 1994.

9. Tamblyn R, Abrahamowicz M, Brailovsky CA y col. The association between licensing examination scores and medical practice. JAMA 1998, 280: 989-996.

10. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990, 65(9s): S63 – S67.

11. Rethans JJ, Norcini JJ, Baron-Maldonado M y col. The relationship between competence and performance: implications for assessing practice performance. Med Educ 2002, 36: 901-909.

12. Venturelli J. Educación Médica: Nuevos enfoques, metas y Métodos. Washington DC: OPS; 1997.

13. Muijtjens AM, van Vollenhoven HM, van Luijt S y col. Sequential testing in the assessment of clinical Skills. Acad Med 2000, 75: 369-373.