

Educación médica, evaluación de las competencias

Prof. Dr. Carlos A. Brailovsky*

Voy a desarrollar algunas reflexiones sobre la visión actual, contemporánea, de algunos aspectos de la educación médica y en particular de la evaluación de la competencia. No se trata, por supuesto, de un curso teórico sobre la evaluación, sino simplemente de presentar ciertas ideas respecto de lo que debe ser, de lo que se está haciendo en este momento y de la importancia de tener en cuenta distintos métodos a utilizar para evaluar la competencia, en particular, la competencia clínica. Trataré de poner en contexto ciertos conceptos actuales respecto de la evaluación en instituciones como una Facultad de medicina y en particular, relacionar la evaluación con la competencia.

Para comenzar me parece necesario precisar qué es la competencia, puesto que si no se define lo que es la competencia, es decir lo que se desea medir, es imposible hacerlo. La competencia es un constructo complejo, multifacético, multivariado, multidimensional, a menudo en relación con una situación multidisciplinaria, en particular en el campo de la medicina. Lo que hace que la investigación sobre la evaluación de la competencia clínica o de los mecanismos del razonamiento clínico sean tan fascinantes para los que investigan en estas áreas, es la complejidad misma del tema que se aborda, ya que siendo tan complejo, crea desafíos enormes, en particular en lo que concierne a su evaluación.

Deseo insistir sobre dos elementos que me parecen primordiales: 1) sólo se puede evaluar lo que se puede observar; 2) siempre hay que identificar y definir los objetivos que los estudiantes deben alcanzar al final de una actividad de formación. Es indispensable explicitar la meta a alcanzar para lograr desarrollar los instrumentos necesarios para poder evaluar de la mejor manera posible el alcance de esos objetivos. Aunque parezca sorprendente, este aspecto de la enseñanza se olvida muy a menudo. 3) No porque tienen algunos instrumentos a mano o porque sean familiares con algunos de ellos, su utilización está justificada. Una regla fundamental de la psicometría es que se deben utilizar los instrumentos pertinentes a la competencia que se desea evaluar, o como ahora está

* MD, MA (Ed), MCFP (Hon)

muy de moda, de las “competencias”. Personalmente prefiero hablar de habilidades; un conjunto de habilidades forman la competencia.

Cuando se habla de competencia y de la importancia de su evaluación, no está de más presentar una pequeña historieta derivada de las aventuras de un “turista” noruego en uno de sus viajes de conquista a Francia. Recorriendo la campiña francesa, nuestro amigo noruego encuentra un granjero con quien entabla conversación. En un determinado momento le pregunta: ¿Es muy difícil ser granjero? ¡Para nada!, contesta el otro. Y agrega: Sobre todo *si se conoce bien* la profesión. La conversación continúa y el granjero agrega: Bueno, se hace tarde y tengo que ir a *ordeñar* mis gallinas. Esta historia nos permite cuestionarnos acerca de la importancia de una evaluación objetiva y *externa* de la competencia autoproclamada.

Figura 1



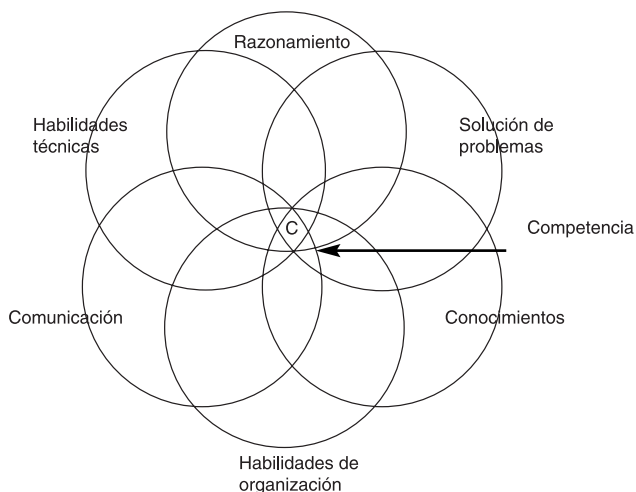
Esta historieta parece ser una anécdota divertida. Sin embargo es interesante hacer el paralelo con situaciones reales. El granjero podría ser un médico y la pregunta sería: ¿Es difícil ser médico? Y la respuesta: “...no, especialmente cuando se conoce bien su profesión...”, y su comentario de competencia autoproclamada, es decir que lo que él cree que es su competencia, no corresponde más a su realidad que lo que era para el granjero.

Ahora bien, ¿qué es la competencia? Se han propuesto muchas definiciones, algunas más operacionales que otras. Probablemente la más simple es: “La capacidad de hacer alguna cosa o la capacidad de cumplir una tarea.” Se trata de una definición simple, pero tan general que es difícil imaginar evaluarla de manera realista. Schon, en 1987, propuso un modelo que todavía se utiliza en ciertos medios: “El modelo de la racionalidad técnica”. Este autor define la competencia como “*la capacidad de resolver problemas seleccionando, aplicando y ajustando los conocimientos para afrontar problemas en contextos estables*”. A pesar de que es un enfoque que se utili-

za en ciertos medios, pienso que se trata de un modelo muy reduccionista que pretende aplicarse a situaciones que no evolucionan, lo que hace que desde mi punto de vista no pueda aplicarse a las ciencias de la salud. La definición de Kane (1992) es mucho más dinámica e interesante. Este autor define la competencia profesional, para un individuo dado, como: “...el grado de utilización de los conocimientos, las habilidades y el buen juicio asociados a la profesión, en todas las situaciones que se pueden confrontar en el ejercicio de la práctica profesional.” Aquí encontramos una nueva dimensión que es fundamental: el buen juicio, la capacidad de razonar para tomar decisiones. Además, Kane introduce otra noción importante: “todas las situaciones de la práctica profesional”, es decir, el aspecto dinámico del constructo. Si deseamos esquematizar la definición, podríamos decir que: la competencia profesional representa la capacidad de un profesional de utilizar su buen juicio así como también los conocimientos, habilidades, actitudes asociadas a la profesión para solucionar los problemas complejos que se presentan en el campo de su actividad profesional. Esta enumeración contiene todos los elementos que se deben “medir” cuando se desea evaluar la competencia. A partir de aquí, hay que decidir si se deben evaluar los elementos de manera independiente o global. Sin embargo, cualquiera sea la decisión, los elementos deben estar siempre contextualizados, asociados a una situación profesional. No hay que olvidar que la competencia no existe al exterior de un contexto particular y profesional.

Esta definición corresponde al constructo mencionado al comienzo, es decir, un modelo multidimensional y multifacético, lo que plantea desafíos importantes para su evaluación. En particular, puesto que un constructo es por definición un concepto, *no es accesible de manera directa*. Se pueden definir indicadores de buena o mala competencia y a partir de la “medida” de los indicadores, inferir que la competencia corresponde a un nivel determinado.

Como se observa en la figura 2, la “competencia” es un constructo que es accesible de manera parcial, a través de aspectos que se pueden medir, como los conocimientos, las habilidades técnicas, las actitudes, la capacidad de solucionar problemas, el razonamiento, los conocimientos, la capacidad de comunicar (con los pacientes o con los colegas, etc). El problema mayor que se plantea es que midiendo los aspectos accesibles se tiene una visión muy parcial del cuadro global. No sólo se dejan de lado las interacciones simples entre elementos de la competencia, sino que además tampoco se observan las interacciones múltiples entre ellos. Inferir niveles de competencia a partir de la medida de aspectos parciales de la misma es muy impreciso. Desgraciadamente, la mayor parte de los instrumentos que se utilizan actualmente para evaluar la competencia clínica no permiten más que una visión parcial de la misma.

Figura 2

¿Por qué hay que evaluar la competencia?

Hay muchas razones para evaluar la competencia de los estudiantes o de los profesionales. Entre ellas podemos enumerar las siguientes:

1. La evaluación forma *parte integrante* del proceso educativo
2. Para *certificar* la competencia (evaluación sumativa)
3. Para dar *retroacción* (evaluación formativa)
4. Para *evaluar* los programas de formación
5. Para *confirmar* los valores y los estándares de la profesión
6. Por la responsabilidad e imputabilidad frente a la sociedad

Las cuatro primeras conciernen muy particularmente a las instituciones de formación profesional, colegios o universidades. No hay duda de que es necesario saber cómo y a dónde han llegado los estudiantes al final de un proceso de formación; por lo tanto, hay que evaluarlos.

Las dos últimas conciernen a los colegios médicos encargados del control de la calidad del ejercicio profesional. Efectivamente, los colegios deben certificar que los profesionales están en condiciones de ejercer la profesión de manera autónoma, segura y de acuerdo con criterios preestablecidos.

¿Qué preguntas hay que hacer cuando se trata de evaluar la competencia?

1. ¿Por qué evaluar?

2. ¿Qué evaluar?
3. ¿Quién evalúa?
4. ¿Cómo se evalúa?
5. ¿Cuándo se evalúa?
6. ¿Para quién se evalúa?

Son preguntas que parecen evidentes y sin embargo, cuando se trabaja con docentes que preparan exámenes, muy a menudo no tienen una idea muy clara de las razones de la evaluación. Respecto de la primera pregunta, las respuestas podrían ser: para dar retroacción a los estudiantes, con fines de toma de decisión, porque es “obligatorio”, con fines de promoción, etc. La segunda pregunta concierne al contenido de lo que se evalúa: lo que se dijo en clase, lo que está en los apuntes, lo que aparece en los libros de texto recomendados, ¡el “yo se” lo que los alumnos “deben saber”!, etc., pero como a menudo los *objetivos de evaluación* no han sido definidos al comienzo del curso, las respuestas son en general muy vagas. La tercera pregunta se presta a grandes controversias. No necesariamente es el docente a cargo del curso quien debe evaluar los estudiantes. Dependiendo del instrumento que se decida utilizar, pueden ser colegas, personas exteriores al sistema, etc. En cuanto a cómo se evalúa, la respuesta depende del instrumento que aparezca como más apropiado y congruente con las estrategias pedagógicas utilizadas y con lo que se desea medir. Si se ha tratado de desarrollar el sentido crítico y el razonamiento científico de los estudiantes, las preguntas escritas a elección múltiple *no representan* el mejor modelo de instrumento a utilizar. La quinta pregunta está en relación con el momento ideal de la evaluación: exámenes parciales luego de cierto período, exámenes semanales, exámenes terminales. Finalmente, hay que tener en cuenta la diferencia entre una evaluación con fines de promoción o de titulación al interior de una institución de enseñanza, un concurso por un puesto, un examen de certificación, etc.

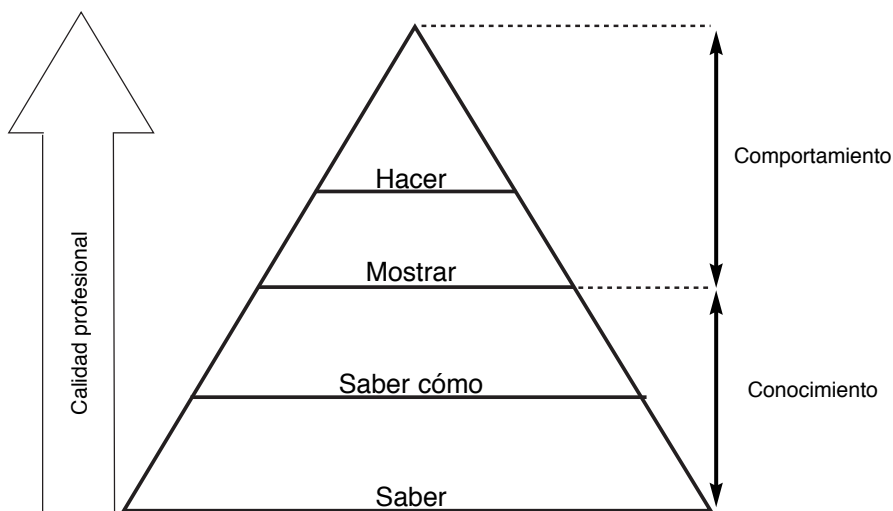
Durante el resto de la presentación, aunque los criterios de elección y de control de calidad de los instrumentos de medida son comunes a todas las evaluaciones que se deseen efectuar, el acento será puesto en particular sobre *la evaluación de la competencia clínica*.

¿Dónde se está actualmente respecto de la evaluación de la competencia?

Georges Miller ha desarrollado un modelo de competencia profesional representado por una pirámide compuesta de varios niveles. En la base se sitúan los conocimientos (el saber) sobre los que se apoya la com-

petencia (el saber cómo). A un nivel superior se encuentra el desempeño (mostrar cómo) y finalmente la acción en la práctica real (el hacer). Este modelo resume bien el constructo de la competencia y permite operacionalizar su evaluación y en particular la elección de los instrumentos de medida.

Fig. 3



Modificade de Miller G.E., The assesment of clinical skills/competence/performance. Academic Medicine (Supplement) 1990, 65: S63-S67

Pero habiendo definido la competencia e ilustrado los estratos que la componen, de acuerdo al modelo de Miller, debemos mencionar las características indispensables de los instrumentos que se utilizan para medirla, entre otras, la confiabilidad, la validez, el impacto educativo, la aceptabilidad por los distintos usuarios y los costos eventuales. Estas propiedades caracterizan la utilidad de un instrumento de medida y pueden resumirse en la siguiente ecuación.

$$U = \left\{ \underset{(p)}{F} + \underset{(p)}{V} + \underset{(p)}{E} + \underset{(p)}{A} \right\} \times \underset{(p)}{\frac{1}{C}}$$

donde U es la utilidad del instrumento que es igual a la suma de la proporción de fidelidad o confiabilidad (F), más la proporción de validez

(V), más la proporción del impacto educativo (E), más la proporción de aceptabilidad (A), el todo multiplicado por la inversa del costo.

La confiabilidad

Se trata de un concepto complejo, que se puede definir y medir de distintas maneras, pero del que se pueden describir ciertos atributos.

1. Es la expresión cuantitativa de la reproducibilidad con la que un instrumento mide el mismo atributo, el mismo constructo.
2. Representa la medida en que los ítems individuales se comportan de manera similar (co-varían) al interior de un examen.
3. Es la constancia con la que un examen mide lo que se supone que debe medir.

La confiabilidad de un instrumento, que se representa con un coeficiente llamado alfa de Cronbach, depende de *la dimensión* del muestreo. En un proceso de evaluación es fundamental explorar (muestrear) muchos contenidos y aspectos de los problemas o de las situaciones clínicas que se desean evaluar. No sólo hay que preocuparse de la dimensión de la muestra (el número de situaciones), sino que, cuando se eligen distintos problemas hay que evitar que los contenidos se superpongan demasiado a través de los distintos problemas clínicos. La situación ideal es aquella en la que hay mucho contenido independiente y parte de contenido que es común. Por otra parte es muy importante muestrear bien las áreas de interés. ¿Qué quiere decir eso?

Cuando el área de interés a evaluar es muy extensa, si tenemos pequeñas muestras, enfrentamos una situación desventajosa, tanto desde el punto de vista de la psicometría como del de validez de contenido. Sobre todo cuando se trata de un constructo de gran complejidad como es la competencia clínica. Cuando se evalúa la competencia clínica, se produce un fenómeno muy bien descrito en la literatura que se llama “la especificidad de la situación clínica”. Este fenómeno implica que el desempeño de cierta calidad frente a una situación clínica no permite predecir la calidad del desempeño frente a otra situación. Es decir que un buen candidato frente a una situación clínica dada, puede ser un mal candidato frente a una situación clínica distinta. Lo que señala, por ejemplo, que hay que tener muestreos relativamente numerosos, tanto en contenidos como en dimensiones a medir cuando se evalúa una materia tan extensa como Clínica Médica. Esto quiere decir que no se debe, de ninguna manera, tomar la decisión de aprobar o aplazar a un estudiante evaluándolo en una sala de hospital con un caso clínico único, al lado de la cama de un enfermo.

Es legítimo y defendible utilizar este tipo de evaluación durante las pasantías y para la evaluación continua de los estudiantes durante la formación clínica, pero su uso es a proscribir totalmente en el momento de los exámenes terminales.

Uno de los problemas que se plantea con este tipo de examen es la falta de confiabilidad de las evaluaciones y de concordancia entre los distintos evaluadores. D. Swanson, responsable de los exámenes de evaluación de la competencia clínica del National Board of Medical Examiners de los Estados Unidos, ha demostrado en un trabajo publicado en 1987 que los exámenes orales clásicos, aun aquellos de una duración de cuatro a seis horas y con 4 a 8 situaciones clínicas, no consiguen mostrar un coeficiente de confiabilidad superior a 0,47. Hay que comparar este coeficiente con el coeficiente mínimo aceptable para tomas de decisión. ¡Este coeficiente no debe ser menor que 0.70! De toda evidencia, el problema es mucho más agudo cuando se utiliza un solo caso puesto que es imposible, desde el punto de vista estadístico, medir su coeficiente de confiabilidad, que es prácticamente nulo. Esto se debe al hecho de que la confiabilidad de un examen depende de la varianza del mismo. El número de casos y de dimensiones de la competencia que se miden aumenta en general la varianza de los scores.

2. La validez

Es otro aspecto de la calidad de un instrumento de medida de la competencia.

1. Muestra hasta que punto un examen mide “realmente” lo que se espera que mida.
2. El grado es el que un instrumento de medida cubre o muestrea en todos los aspectos de lo que se pretende medir. Está basada sobre la opinión de expertos.
3. Para la evaluación de la competencia clínica, la validez de contenido debe basarse sobre:
 - las habilidades que se desean evaluar;
 - el contenido clínico, que debe estar en relación estrecha con la esfera de peritaje esperada;
 - el instrumento utilizado, que debe ser congruente con la estrategia pedagógica utilizada y la habilidad que se desea medir.

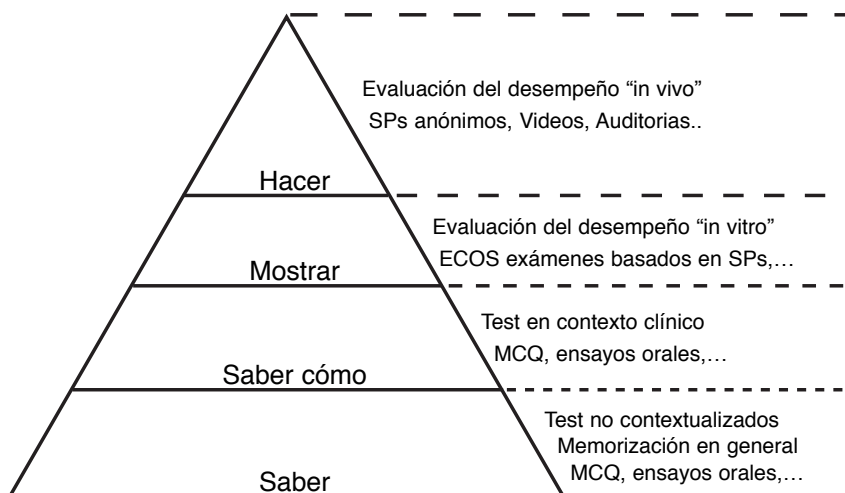
Existen distintos aspectos de la validez, pero en esta presentación los aspectos que nos interesan son: la validez de contenido y la congruencia de

los instrumentos de evaluación con los aspectos de la competencia que se desean evaluar.

Se puede decir que la validez de los instrumentos de medida está en relación con los estratos de la pirámide de Miller (fig. 2). No hay duda que la calidad profesional global se incrementa a medida que se trepa la pirámide. La complejidad taxonómica aumenta desde los simples conocimientos hasta la acción, la actividad en la vida real. Tener *conocimientos* (saber) no significa *saber explicar* cómo utilizarlos (decir lo que se debe hacer). Y..., *decir lo que se debe hacer* no implica *saber desempeñarse*; y saber desempeñarse en una situación de evaluación no implica necesariamente *actuar* con sabiduría y profesionalismo en la vida real, es decir, de manera autónoma con verdaderos pacientes. Por lo tanto, si queremos categorizar estos niveles de la pirámide, los dos inferiores (en la base) representan los conocimientos, mientras que los dos superiores, mostrar y hacer, representan los comportamientos. Esto condiciona distintas maneras de abordar los dos aspectos. En el lenguaje de la evaluación se dice que la validez de la misma está en relación con la trepada de la pirámide: cuanto más alto se sube en la pirámide más válido será el modelo de evaluación que se debe utilizar.

3. ¿Cómo se trepa la pirámide? O de la buena utilización de los instrumentos de medida

Figura 4



El *saber* clásicamente se mide antes de medir la calidad de los desempeños. En general se utilizan exámenes de elección múltiple, aunque también se pueden utilizar “ensayos” o redacciones cortas o aun exámenes orales, pero todas estas evaluaciones se practican de manera descontextualizada y miden principalmente la memorización de hechos. *El saber cómo* se puede medir con modelos clínicos compuestos por preguntas contextualizadas, es decir a partir de una situación clínica simple, una pequeña viñeta pone en contexto “real” la situación. Los desafíos son mucho más complejos cuando se va trepando la pirámide; cuando queremos evaluar el *mostrar* (desempeño), enfrentamos situaciones taxonómicas de otro nivel, mucho más complejas, en las que no sólo se describe *cómo*, sino que *se muestra cómo*. Se entra en una dimensión totalmente diferente de la evaluación. Lo que precede muestra los errores potenciales que pueden cometerse cuando se infiere la presencia de competencia a partir de exámenes que miden conocimientos y dejan de lado la medida de desempeños.

Daré un ejemplo que apoya lo anterior. Con uno de mis colegas hicimos una experiencia que nos permitió efectuar observaciones interesantes, confirmando a qué punto es importante distinguir la etapa de “saber cómo” de la de “mostrar cómo” antes de inferir a partir de los resultados observados el nivel de competencia del “estudiante” evaluado. Con mi colega de estudios de doctorado en psicometría diseñamos un protocolo de investigación aprovechando el hecho de que él es profesor de educación física y entrenador de un equipo universitario de esquí. Yo me presté como cobayo de la experiencia. Debo aclarar que a pesar de vivir en un país de largos inviernos y mucha nieve, nunca he esquiado. Para nuestra experimentación, durante la primera etapa, mi colega me familiarizó, desde el punto de vista teórico, con técnicas de base y avanzadas de esquí. Durante la segunda etapa de la experiencia, los estudiantes de su grupo (n=31) me hicieron pasar un examen oral sobre técnicas de transferencia de peso para tomar curvas en el slalom. No necesito insistir en el hecho de que mis resultados fueron muy buenos y los “examinadores” consideraron que podía ser categorizado como ¡experto! Por suerte no se les planteó ir un poco más allá y pasar a la etapa siguiente: “Bueno, ahora vamos a la pista de esquí donde usted va a ponerse los esquís y nos va hacer una demostración de cómo efectúa la transferencia de peso para mantener una buena velocidad durante una bajada.” Estos resultados no son anecdóticos puesto que formaron parte de las experiencias que compusieron el trabajo de tesis de mi colega.

Este ejemplo nos lleva al nivel siguiente de la trepada de la pirámide, a la evaluación del desempeño “in vitro” en la que utilizan distintas situaciones que imitan lo que ocurre en la vida real, pero en un contexto homogéneo y controlado. Para la evaluación de los desempeños, se utilizan distintas técnicas y modelos. En este momento, el más popular y proba-

blemente el mejor estudiado y validado es el Examen Clínico Objetivo Estructurado (ECO), que en la literatura de habla inglesa se representa con la sigla OSCE. Este instrumento de evaluación de la competencia clínica está compuesto de una serie de casos clínicos muy bien estructurados, en los que se utilizan grillas de evaluación estandarizadas y pacientes actores “simulados”, quienes juegan el papel correspondiente. Para ello aprenden un guión que les permite actuar y comportarse como verdaderos pacientes.

Finalmente, termina la trepada y se llega a la cúspide de la pirámide, es decir, la evaluación de los profesionales en la práctica real, frente a verdaderos pacientes. Una manera de proceder es utilizando la observación directa en consultorios donde se dispone de espejos unidireccionales de manera de interferir lo menos posible en la consulta. Por supuesto, por razones éticas, se necesita el acuerdo de los pacientes que consultan. Las consultas se pueden grabar en video y más tarde pueden ser vistas por los estudiantes o los residentes de manera individual o en presencia del supervisor de manera de analizar lo que ha pasado durante la consulta. Otra manera de proceder es utilizando pacientes simulados anónimos, que piden consulta sin que los médicos a evaluar sepan quiénes son. Por supuesto, por las mismas consideraciones éticas, los médicos que reciben estos “pacientes” en consulta, aunque no puedan identificarlos, han debido aceptar prestarse a este tipo de evaluación. Para completar las evaluaciones de la competencia, se utilizan auditorías en las que se analizan legajos o historias clínicas en relación con patologías prototípicas de manejo clínico y se encuentra a un cierto número de pacientes que corresponden a las historias analizadas.

La conclusión en esta etapa es que la autenticidad educativa o profesional representa el punto fundamental de la evaluación, al interior y a través de los niveles de la pirámide.

Figura 5

De la mala utilización de un buen instrumento



Un último elemento a retener es que en general los instrumentos no poseen un valor intrínseco que indique que son mejores o peores que otros. Es sobre todo la manera de utilizarlos lo que causa problemas.

4. Impacto educativo

La última etapa de análisis de los instrumentos de medida está representada por la evaluación del impacto educativo de los mismos, es decir, de qué manera influyen sobre el proceso de enseñanza – aprendizaje. Ya en 1984, Fredericksen mencionaba que: *“One of the most potent methods to change the way students learn and the way faculty teach is to change the test”* (El método más poderoso para cambiar la manera de estudiar de los estudiantes y la manera en que los profesores enseñan, es cambiar la manera de evaluar.), en *“The Real Test Bias”*, *American Psychologist*. Y Georges Miller, considerado como el padre de la pedagogía médica y antiguo director del Departamento de Educación Médica de la Universidad de Chicago propuso que: *“evaluation drives curriculum”* (la evaluación orienta el currículo). En otras palabras, es a partir de lo que se desea alcanzar, los objetivos de enseñanza y de aprendizaje, que se prepara el currículo y las estrategias pedagógicas correspondientes necesarias para el alcance de los objetivos de formación; y que se decide de los métodos que permitirán medir el alcance de esos objetivos. Por lo tanto, los contenidos de los programas, los objetivos de aprendizaje y de evaluación y los métodos que se utilizan para medir el alcance de los mismos están interconectados. Sin embargo, la investigación actual en psicometría muestra que probablemente la evaluación no arrastra el currículo, sino que orienta lo que el estudiante aprende y cómo lo aprende. Aparece cada vez más claro que la evaluación dirige el aprendizaje de los estudiantes. ¿De qué manera? A través de los contenidos y del formato de la evaluación.

Si, por ejemplo, se utilizan estrategias como el aprendizaje por problemas o las discusiones en pequeños grupos para desarrollar el pensamiento crítico de los estudiantes y luego se los evalúa con exámenes de “multiple choice”, se enfrenta una incongruencia total entre el instrumento que se utiliza y el enfoque pedagógico utilizado con los estudiantes. Hace poco tiempo, evaluando los exámenes escritos de certificación utilizados por una organización profesional, descubrí que utilizaban un grupo de preguntas “multiple choice” para medir la calidad de la comunicación de los candidatos. Los responsables del examen pretendían que si la pregunta plantea: ¿qué preguntaría usted en esta situación? Y el candidato elige la respuesta: “si su mamá se murió joven”, ¡están evaluando comunicación!

Es importante ser consciente del formato del instrumento que se utiliza en las evaluaciones por diferentes razones:

1. Se puede pensar que se evalúa lo que no se está evaluando.
2. Los estudiantes adaptarán el estudio al tipo de formato que se utiliza.
3. Es importante para la credibilidad del sistema que el instrumento sea congruente con los métodos de enseñanza utilizados.

Pienso que lo que hay que retener de esto es que la evaluación tiene un impacto educativo muy importante, que orienta el aprendizaje y que hay que utilizarla de manera estratégica para reforzar el aprendizaje. En Canadá, la evaluación no se utiliza con fines punitivos sino con la finalidad de ayudar y guiar a los estudiantes a través del proceso de maduración y formación. El rol del docentes es ser guías del aprendizaje. El de los estudiantes es ser responsables de su aprendizaje y del alcance de los objetivos que se les han propuesto.

5. ¿Cuáles son las tendencias de la educación médica en Canadá?

En Canadá se utilizan distintos modelos y enfoques pedagógicos, en particular en las facultades de medicina: aprendizaje en situación (situacional), aprendizaje centrado en proyectos, aprendizaje basado en problemas, aprendizaje a través del descubrimiento, aprendizaje totalmente centrado en el estudiante; se utiliza también toda una serie de estrategias llamadas “aprendizaje auténtico” y aprendizaje basado en pacientes. No todos los enfoques se utilizan en todas las Facultades de Medicina del país (hay 16 en Canadá), pero quiero subrayar que cuando se reflexiona acerca de cambio curricular hay que reflexionar sobre todos los aspectos pedagógicos implicados en el proceso de cambio. Se trata de una dimensión de gran importancia puesto que cambios en el vacío no llevan a ningún lado, más que a “frustraciones institucionales”. Un cambio curricular es algo global y no puntual; requiere una reflexión profunda acerca de todo el sistema que se quiere modificar así como del que se propone para remplazarlo. No se trata simplemente de cambiar o modificar contenidos, horarios de clases o de trabajos prácticos.

En este momento, en nuestro medio hay un verdadero desplazamiento de paradigmas: del objetivismo al constructivismo.

El objetivismo, paradigma “casi” universal, postula que el conocimiento científico:

1. es verdadero;
2. es absoluto;

3. se recoge metodológicamente;
4. está compuesto de hechos objetivos.

El objetivismo ha sido el paradigma dominante de manera casi absoluta durante muchos años. Sin embargo, en este momento se observa un desplazamiento gradual de paradigma hacia el constructivismo.

El constructivismo postula que el conocimiento:

1. está construido por el que aprende;
2. se basa en la comprensión del que aprende;
3. el que aprende construye las estructuras cognitivas a través de interacciones.

Este desplazamiento de paradigma es algo muy importante puesto que es la base de los estudios e investigaciones que se hacen sobre los modelos del razonamiento clínico. La base de la visión moderna del desarrollo del pensamiento clínico reposa sobre el modelo de redes conceptuales del pensamiento y sus modificaciones e interacciones para la resolución de problemas y la toma de decisiones.

En resumen, cuando se plantea un cambio curricular, hay que tener en cuenta toda una serie de elementos que se pueden agrupar en la categoría “factores de cambio”. Cuando se habla de factores de cambio, hay que pensar en las teorías educativas que han sido aplicadas y validadas. Las teorías educativas tienen en cuenta insumos cognitivos, afectivos y de regulación. No se debe aplicar el modelo de despertarse por la mañana y decir: tengo una idea, y tratar de reinventar la rueda. La pedagogía es una ciencia sólida que se apoya en modelos teóricos, demostraciones e investigación. Cuando se introducen cambios no sólo hay que respetar los modelos teóricos, sino también justificar si se tienen en cuenta las exigencias y las expectativas de la sociedad en la que funcionan las instituciones. Las visitas de acreditación de Facultades de Medicina en Canadá, exigen que se presente un estudio de entorno social y que se explique de qué manera los programas de la Facultad se adaptan a las necesidades del medio social. Por ejemplo, en este momento la atención ambulatoria y la cirugía de un día sin internación han tomado gran importancia en nuestro medio. Los responsables de la acreditación deben ser informados de las medidas previstas para adaptar los programas de formación a la nueva situación. Los responsables de los programas de formación no vivimos en el interior de una bola de marfil, aislados de la sociedad que nos rodea de manera a permitirnos decir: “esto es lo bueno para la sociedad”, sino que tenemos la obligación de formar profesionales que respondan a las necesidades de la sociedad.

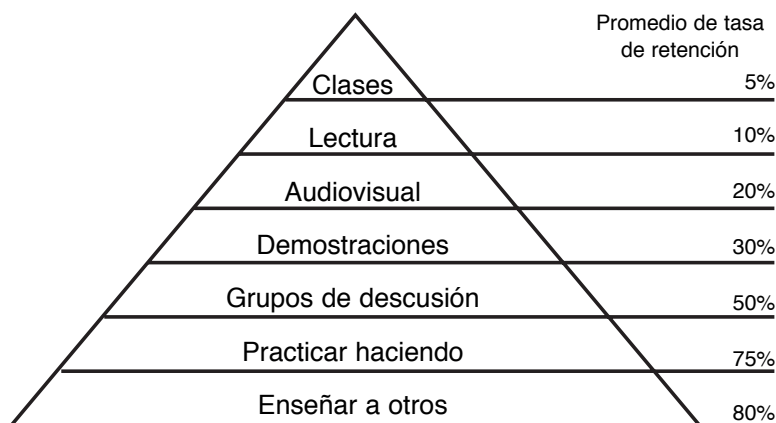
Métodos pedagógicos

Antes de terminar, deseo hacer algunos comentarios sobre métodos pedagógicos y algunos otros aspectos de la evaluación de los estudiantes.

Cuando se habla de cambio curricular, además de la pirámide de la competencia y de la evaluación hay que tomar en consideración la pirámide del aprendizaje. Esta pirámide pone en relación los métodos utilizados y la tasa de retención de conocimientos y conceptos.

Cuando se observa esta pirámide (fig.6), ciertos aspectos de la misma nos cuestionan: un teórico no induce más que 5% de retención de lo que se ha dicho en clase, mientras que ejercicios prácticos inducen alrededor de 75% de retención. Las otras intervenciones o actividades se sitúan en tasas de retención intermedias. Es interesante subrayar que la mejor manera de aprender es enseñar a otros (80% de retención). El corolario fácil sería postular que es el profesor quién más y mejor aprende a través de la preparación de una clase. ¡Y quizás sea cierto! En realidad, esto abre una avenida interesante cuando se habla de cambios curriculares; se trata de incorporar actividades en las que los alumnos jugarán un rol activo preparando charlas, seminarios, de los que serán responsables en distintos momentos.

Figura 6



Fuente: National Training Laboratories, Bethel, Maine, USA

¿Hacia dónde se está yendo en evaluación?

Tal como lo he mencionado, la evaluación orienta el aprendizaje, y el currículum condiciona lo que se debe evaluar y cómo evaluar. No hay que

olvidar que la evaluación es parte integrante del modelo educativo y debe ser congruente con él. Es necesario poner mucho énfasis en todas las nuevas habilidades que no se tenían en cuenta en los currículum clásicos: la importancia de aprender a aprender; los modelos de auto evaluación; el desarrollo de autocritica frente a lo que se hace; los modelos de liderazgo, movilidad en el interior de grupos y dinámica de grupos. Este último aspecto a adquirido gran importancia puesto que cada vez más nuestros jóvenes egresados tendrán que trabajar en el interior de grupos multidisciplinares compuestos por enfermeras, médicos, farmacéuticos, clínicos, psicólogos o trabajadores sociales, es decir, todo lo que está en relación con la dinámica de grupos, trabajo en equipo, liderazgo, ha adquirido gran importancia. No hay que olvidar el desarrollo de la capacidad reflexiva y del razonamiento. Esto implica mucha creatividad y ser capaces de enfrentar desafíos bastante complejos para desarrollar buenos instrumentos de evaluación, de manera de poder inferir a partir de los resultados obtenidos con ellos, la presencia y la calidad de las dimensiones mencionadas.

Una de las mayores dificultades con que se enfrenta la evaluación de la competencia es la “medida” de la meta cognición. Ya no se trata de medir el saber, el saber cómo o el mostrar cómo, sino medir una “tercera” dimensión de la pirámide, las meta habilidades, es decir, la comprensión de cómo se ha llegado al resultado. No es suficiente que se diga: “este es el resultado”, sino que hay que poder explicar como se ha llegado a él. Estos desafíos han estimulado la aparición de nuevas técnicas de evaluación:

1. auto evaluación,
2. evaluación por los pares,
3. cuadernos en los que se anotan las actividades cotidianas,
4. simulación de procesos de aprendizaje,
5. simulación de procesos de evaluación,
6. evaluación de resultados,
7. evaluaciones de tipo auditorías. La lista precedente, por supuesto, no es exhaustiva. Un problema que se enfrenta con estos enfoques es que todos los métodos son cualitativos o, a lo sumo, semi cuantitativos, lo que complica la toma de decisión objetiva e imparcial.

Debo insistir en el hecho de que todo lo que concierne a la evaluación de competencias representa un compromiso, un consenso entre responsables; en este tema no existe una receta única ni un instrumento ideal. Si tenemos en cuenta que la utilidad de un instrumento está representada por proporciones de confiabilidad, validez e impacto educativo,

la proporción de cada uno de los elementos estará ajustada a la finalidad del mismo.

1. Frente a una situación de selección de alumnos, es necesario disponer de un instrumento que tenga gran confiabilidad y que la confiabilidad sea buena en la zona de corte o de decisión que debe ser relativamente alta, ya que se quiere seleccionar el mejor o los mejores candidatos. Tiene que ser válido, pero el aspecto “impacto educativo” es totalmente secundario. Los resultados no serán utilizados para dar retroacción a los candidatos.
2. Frente a un examen final, la situación es relativamente similar. Se necesita buena confiabilidad, la validez siempre tiene que estar presente, el impacto educativo, una vez más, es relativamente secundario. Sin embargo, hay que tener en cuenta que los alumnos ajustarán la manera de prepararse en función del formato del examen. En este caso, se espera que la confiabilidad sea buena en la zona de una nota de corte intermedia. Si la confiabilidad es buena en un nivel muy alto, el riesgo es de aplazar 80 o 90% de los estudiantes. Desde el punto de vista de una buena evaluación final, se espera que 85 a 90% de los estudiantes aprueben la materia. La finalidad de la evaluación al final de un curso es de verificar el nivel de los estudiantes y no de aplazar al 80% de ellos. Si el curso ha sido bien estructurado, bien organizado, los objetivos de aprendizaje bien explícitos y los objetivos de evaluación bien definidos, se puede esperar que el 100% de los estudiantes aprueben. Con tasas de aplazos superiores a 25% es fundamental que los responsables del curso se planteen muchas preguntas acerca del mismo o de la pertinencia del examen.
3. Finalmente, en situación de evaluación continua durante la formación de los estudiantes (*in-training evaluation* de la literatura inglesa), el ejemplo clásico es el residente en un servicio hospitalario. En este caso, los instrumentos que se utilizan no tienen ninguna confiabilidad; lo importante es que estén estructurados de manera de tener un impacto educativo importante puesto que se utilizan sobre todo para hacer una evaluación formativa y para dar retroacción a los estudiantes. En este caso la confiabilidad es secundaria, la validez es importante y el impacto educativo, fundamental.

Cuando se observa el currículo en su globalidad, es necesario que el conjunto de los instrumentos de evaluación que se utilizan sean confiables, válidos y que posean un buen impacto educativo.

Como conclusión podemos decir que la evaluación constituye un aspecto del diseño pedagógico global de una institución que se debe utilizar estratégicamente por sus efectos educativos.

1. Se necesita mucha más investigación sobre los efectos educativos de la evaluación y probablemente menos investigación sobre sus propiedades psicométricas. No hay duda que hay que conocer la calidad psicométrica de los instrumentos de evaluación; sin embargo es necesario explorar otros aspectos de los mismos.
2. La investigación debe ser por definición rica contextualmente si se quiere obtener una imagen correcta de la calidad de los instrumentos de medida.
3. La utilidad de un instrumento depende del contexto en que se utiliza, puesto que un instrumento que funciona bien en un contexto puede funcionar mal en otro.
4. Es fundamental, para realizar un buen trabajo en evaluación, cubrir toda la pirámide de competencia. Si los evaluadores se limitan a los dos estratos inferiores de la pirámide estarán infiriendo erróneamente a partir de los resultados observados. Es necesario que estén bien sensibilizados a este problema. Muy a menudo se cree estar evaluando niveles superiores de la competencia cuando en realidad no es lo que se está haciendo.
5. No hay que temer basarse más a menudo en datos cualitativos y de juicio profesional, ya que la calidad global de las evaluaciones depende mucho más de un programa integral de evaluación que de instrumentos individuales. Si se quiere hacer una reforma curricular y no se aborda esta problemática de manera vigorosa, la reforma no podrá consolidarse.

Finalmente, el éxito del desplazamiento de los paradigmas está en manos de los que conciben y construyen los instrumentos de evaluación. Si el cuadro conceptual forma parte de “los viejos paradigmas”, las evaluaciones no podrán ser buenas. La calidad de los instrumentos sólo depende de quienes los crean y utilizan, no de quienes son evaluados con ellos.

Bibliografía

- Brailovsky CA, Grand'Maison P. "Using evidence to improve evaluation: A comprehensive psychometric assessment of a SP-based OSCE licensing examination", *Advances in Health Sciences Education*, 2000; 5:207-219.
- Brailovsky, C.A., Charlin, B., Beausoleil, S., Coté, S., & Van der Vleuten, C. "Measurement of Clinical Reflective Capacity Early in Training as a Predictor of Clinical Reasoning Performance at the End of Residency: An Exploratory Study on the Script Concordance Test", *Medical Education*, 2001 février (in press).
- Brailovsky, C.A., Grand'Maison, P. & Lescop, J. "A Large-scale Multicenter Objective Structured Clinical Examination for Licensure", *Academic Medicine* (1992) 67:S37-S39.
- Brailovsky, C.A., Grand'Maison, P. & Lescop, J. "Construct validity of the Quebec licensing examination standardized patient-based OSCE", *Teaching and Learning in Medicine* (1997) 9:43-50.
- Brailovsky, C.A., Miller, F & Grand'Maison, P., "L'évaluation de la compétence dans le contexte professionnel", *Service Social*, 1999,47:171-189.
- Brailovsky, C.A. & Grand'Maison, P. "Using evidence to improve evaluation: A comprehensive psychometric assessment of a SP-based OSCE licensing examination", *Advances in Health Sciences Education*, 2000,5:207-219.
- Brown J.B, Rainsberry P, Handfield-Jones R, Brailovsky CA. "The certification examination of the College of Family Physicians of Canada: IV) Simulated Office Orals", *Can Fam Physicians* 1996;42:1539-1548.
- Charlin, B., Roy, L., Brailovsky, C. & van der Vleuten, C. "How to build a script concordance test to assess the reflective clinician", *Teaching and Learning in Medicine* 2000,12:189-195.
- Grand'Maison, P., Brailovsky, C., Miller, F., Emond, J-G., Nasmith, L., "Impact of a performance-based licensing examination on teaching and learning in Family Medicine", *Academic Medicine*, 2001 (accepted).
- Grand'Maison, P., Brailovsky, C.A. & Lescop, J., "Content validity of the Quebec licensing examination OSCE: assessed by practicing physicians", *Can. Fam. Phy.* (1996) 42:254-259.
- Grand'Maison, P., Brailovsky, C.A., Lescop, J. & Rainsberry, P. "Using Standardized Patients in Two Licensing/Certification Examinations: Comparison of Two Tests in Canada", *Family Medicine* (1997) 29:27-32.
- Grand'Maison, P., Lescop, & Brailovsky, C.A. "Canadian Experience with Structured Clinical Examinations", *CMAJ*, (1993) 148:1573-1576.

- Grand'Maison, P., Lescop, J., Rainsberry, P. & Brailovsky, C.A. "Large-scale use of an objective, structured clinical examination for licensing family physicians", *CMAJ* (1992) 146:1735-1740.
- Grand'Maison P., Lescop J., Brailovsky C.A. "Large-scale use of an objective structured clinical examination for licensing family physicians", *Can. Med. Ass. J.* 1992;46:1735-1740.
- Handfield-Jones R., Brown J.B., Biehn J.B., Rainsberry P., Brailovsky C.A. "Certification examination of the College of Family Physicians of Canada. Part 3: Short Answer Management Problems", *Can Fam Physicians*, 1996;42:1353-1361.
- Harden R., Glesson F. "Assessment of clinical competence using an objective structured clinical examination (OSCE)", *Medical Education* (1979) 13:41-54.
- Kane M.T. "The assessment of clinical competence", *Evaluation and the Health Professions* (1992) 15:163-82.
- Kerlinger F.N. *Foundations of Behavioral Research*, 3rd edition, New York: Holt, Rinehart and Winston Inc, 1986.
- Linn R.L., Baker E.L., Dunbar S.B., "Complex, performance-based assessment: Expectations and validation criteria", *Educational Researcher*, 1991; 20:15-24.
- Miller, G., "The assessment of clinical skills/competence/performance" *Academic Medicine* (1990) 65:S63-S67.
- Newble D., Jeager K. "The effect of assessments and examinations on the learning of medical students", *Medical Education*, 1983;17:165-171.
- Shepard L.A., "The role of assessment in learning culture", *Educational Researcher*, 2000, 29:4-14.
- Stillman P.L., Haley H.L., Regan M.B., Philbin M.M., "Positive effects of clinical performance assessment program", *Academic Medicine*, 1991; 66: 481-483.
- Streiner D.L., Norman G.R., *Health Measurement Scales. A practical guide to their development and use*, 2nd edition, New York: Oxford University Press, 1995.
- Tamblyn, R., Abrahamowicz, M., Brailovsky, C.A., Grand' Maison, P., Lescop, J., Norcini, J., Girard, N. & Haggerty, J.L., "The association between licensing examination scores and medical practice", *JAMA* (1998) 280:989-996.
- Van Der Vleuten C.P.M., "The assessment of professional competence: developments, research and practical implications", *Advances in Health Sciences Education*, 1996; 1:41-67.