

Evaluando la calidad de la universidad, particularmente su docencia

ROBERT STAKE* | GLORIA CONTRERAS P.** | ISABEL ARBESÚ***

Evaluar la calidad de la universidad es una tarea sumamente compleja, considerando los múltiples factores susceptibles de evaluación, así como la intención o concepción (paradigma) de la que se parte para evaluar. Este artículo se centra en torno a cuatro ejes fundamentales: evaluación formativa y sumativa; las problemáticas que impiden que se lleve a cabo un tipo de evaluación comprensiva; y la responsabilidad que implica evaluar, considerando aspectos como la rendición de cuentas de los méritos, debilidades y preocupaciones en la universidad, a partir del quehacer docente. La tarea de evaluar constituye una necesidad actual, no sólo por su valoración social como un tipo de *medición* de la calidad, sino como parte intrínseca de un proceso de mejora necesario y constante, tanto de la institución universitaria como de sus docentes.

Palabras clave

Evaluación formativa
Evaluación sumativa
Universidad
Calidad
Docencia

Recepción: 12 de mayo de 2011 | Aceptación: 15 de junio de 2011

- * Ph.D., Princeton University, 1958. Profesor de Educación y Director del Center for Instructional Research and Curriculum Evaluation (CIRCE) de la Universidad de Illinois. Líneas de investigación: formación de profesores primarios y secundarios de Matemática y Ciencias; desarrollo de enseñanza con equidad; educación ambiental; servicios sociales urbanos. Publicaciones: (2008, en coautoría con L.M. Rosu), "The Assessment of Teaching and Learning in USA: Reflections on school quality and the role of teacher in public schools", *Romanian Pedagogical Journal*, vol. 1, núm. 1, pp. 49-66; (1986), *Quieting Reform: Social science and social action in an urban youth program*, Champaign, IL, University of Illinois Press. CE: stake@illinois.edu
- ** Doctora en Ciencias de la Educación por la Pontificia Universidad Católica de Valparaíso, Chile. Profesora del área de Evaluación del Instituto de Educación de la Pontificia Universidad Católica de Valparaíso. Actualmente directora del programa de Magister en Educación de la Escuela de Pedagogía de la Pontificia Universidad Católica de Chile. Publicaciones: (2010), "Diseño y aplicación de un sistema de evaluación del desempeño docente con fines formativos: la experiencia de la Pontificia Universidad Católica de Valparaíso, Chile", *Revista Iberoamericana de Evaluación Educativa*, vol. 3, núm. 1e, pp. 179-191; (2010), "Diagnóstico de dificultades de la evaluación del aprendizaje en la universidad: un caso particular en Chile", *Educación y Educadores*, vol. 13, núm. 2, pp. 219-238. CE: gcontrer@ucv.cl
- *** Doctora en Educación, profesora e investigadora de la Universidad Autónoma Metropolitana, unidad Xochimilco. Miembro del Sistema Nacional de Investigadores, la Red (nacional) de Investigadores sobre Evaluación de la Docencia (RIED) y la Red Iberoamericana de Investigadores sobre Evaluación de la Docencia (RIIED). Miembro investigador del Observatorio sobre la Didáctica de las Artes con sede en la Universidad de Barcelona. Sus áreas de investigación y docencia tienen que ver con el estudio y la evaluación del modelo pedagógico denominado sistema modular que se imparte en la Unidad Xochimilco de la UAM; con el análisis, la evaluación de la docencia y la enseñanza en educación superior; y con la metodología de investigación cualitativa. Publicaciones: (2010, con G. Argumedo), "Diseño e instrumentación de portafolios para evaluar la docencia", *Revista Observar*, núm. 4, pp. 28-44; (2010, con G. Argumedo), "El uso del portafolios como recurso para evaluar la docencia universitaria", *Revista Iberoamericana de Evaluación Educativa*, vol. 3, núm. 1e, pp. 134-146. CE: isabel.arbesu@gmail.com

Las universidades son organizaciones complejas, integradas por sectores parcialmente autónomos y facultades aún más autónomas. La tradición y las reglas propician la homogeneidad de sus actividades; sin embargo, la mayoría de las autoridades, académicos y estudiantes gozan de una cultura de libertad en el trabajo. Vincent Tinto (1995) señala que es difícil encontrar una sola cultura académica, sino que más bien encontramos diferentes culturas, diferentes prácticas de evaluación y variados grupos de académicos y estudiantes.

DIVERSAS EVALUACIONES EN LA UNIVERSIDAD

Aunque los cimientos de las edificaciones se desgastan con el tiempo, el colapso sucede cuando se les añaden nuevas estructuras; lo mismo sucede con la enseñanza computarizada y colaborativa: la investigación en la universidad busca abrirse paso con un sistema de financiamiento que muchas veces resulta controvertido.

Para algunos estudiantes la universidad es, de cierta manera, un club campestre, no tanto por las piscinas, sino por su exclusividad. Las promesas y expectativas de la universidad varían considerablemente de un lugar a otro, por lo que el estudio de cada caso reclama sensibilidad y cuidado.

Evaluar la calidad¹ del trabajo en la universidad, sus programas y servicios, es una responsabilidad compartida por una amplia variedad de interesados: los miembros del directorio o consejo administrativo, las autoridades centrales y sectoriales, los estudiantes y sus familias, y, por supuesto, los académicos y sus sindicatos; también los medios de comunicación, etc. Todos ellos evalúan lo que la universidad hace, aunque gran parte de la evaluación que realizan es compleja e invisible, dado que la mayor parte de ésta es informal y está cimentada en la introspección y

experiencia personal, además de que está enmarcada en la larga tradición y reputación de las instituciones universitarias (Dressel, 1971).

La calidad global de cualquier universidad es irregular, condicionada y tan compleja como sus muchas funciones, aulas, bibliotecas y laboratorios. Otorgarle a una universidad el número XVII, situarla entre las diez primeras o calificarla como “muy buena”, es una extrema simplificación aun a pesar de la especificidad de los criterios o la profundidad de la información. Los *rankings* son muy difundidos porque la gente se siente cómoda con los estereotipos y el brillo que otorga la adulación narcisista.

Por lo anterior, la comparación entre universidades o departamentos sirve de poco al bien público o gerencial. Las evaluaciones más valiosas son aquellas con carácter formativo, que ayudan al desarrollo y que se destinan a evitar las disfunciones, a superar las debilidades y a dar forma a una nueva enseñanza, una nueva investigación y un servicio público renovado.

Tal como lo expresan sus autoridades, las universidades tienen su propio carácter, una misión, una salud organizacional, una ética, una singularidad, diversidad e integridad. Cada departamento, programa profesional y área, y cada individuo de la comunidad universitaria contribuye consciente, o inconscientemente, a ese carácter. Prestar atención cuidadosa a los eventos, al flujo de los datos operativos y al estado de ánimo de la comunidad son buenos indicadores pero dirigir las más grandes y las más pequeñas funciones de la universidad depende menos de indicadores formales y más de la experiencia y la introspección.

Si bien hay muchas posibilidades de mejorar la evaluación de la calidad en las universidades locales y en las internacionales, existe un peligro considerable en representar la complejidad de sus funciones mediante

1 Calidad significa bondad, mérito y valor (Stake y Schwandt, 2006). No se limita a las ideas de *total quality management* (Ahire, 1997).

critérios y simples indicadores. Para combatir estas presunciones hay muchos temas que considerar, pero limitaremos este artículo a cuatro, uno de los cuales será la evaluación de la docencia.

Estructuraremos el artículo en cuatro partes cuyos propósitos son: 1) destacar la tensión entre la evaluación formativa y la sumativa en la universidad; 2) identificar las barreras que impiden una evaluación comprensiva en la universidad; 3) presentar una visión sinóptica de la responsabilidad de evaluar la docencia, entendida como un ejemplo de rendir cuentas de los méritos y debilidades de las funciones de la universidad; y 4) analizar el rol de la evaluación formal e informal en la mejora de las funciones universitarias, tales como la docencia.

EVALUACIÓN FORMATIVA Y SUMATIVA, FORMAL E INFORMAL

Evaluar es parte de un comportamiento normal y en la universidad casi todo es evaluado (Stake y Cisneros, 2000). La mayor parte de la evaluación es informal, incluso involuntaria, ya sea que la realicen las autoridades u otros.

Evaluación formal. En algunas ocasiones, y como parte de alguna rutina operativa, se hacen esfuerzos extraordinarios por observar, medir, registrar e interpretar el mérito y las deficiencias de las cosas. Cuando la evaluación llega a ser planificada, rutinaria, registrada y publicitada decimos que es formal. Muchas de las funciones de la universidad son evaluadas de manera formal e informal.

Una base para la evaluación formal debería determinar sus métodos, procedimientos, instrumentos y actividades. A menudo, para una evaluación de cualquier “evaluando” (es decir, la entidad o función que está siendo evaluada en la universidad), hay distintas bases, y éstas dependerán del uso que se le quiera dar a la evaluación. Por supuesto, el uso será diferente si se trata de evaluación sumativa o formativa (Contreras, 2010).

Evaluación sumativa. Cuando la evaluación se realiza con el objetivo fundamental de comprender la calidad de un evaluando, la llamamos sumativa. Pensamos que la persona o instancia a evaluar tiene un valor; sabemos que el valor depende en parte de su situación, de sus potenciales usos y de los usuarios en sí mismos, pero por el momento nos concentraremos en qué tan bueno o malo es. Cuando hablamos de la colaboración entre la universidad y la industria, o de sus programas de música, o de sus bibliotecas, usualmente estamos hablando de su calidad en sentido sumativo, en un tiempo y lugar específicos.

Evaluación formativa. Pero a menudo vemos al evaluando como algo que más bien se está desarrollando, cambiando, reorganizando para un nuevo uso, y recurrimos a la evaluación formativa. El jardinero poda el árbol, el asesor cambia el protocolo, el entrenador recluta atletas. Todos ellos están evaluando formativamente cómo mejorar el proceso. En la evaluación formativa no sólo se considera el uso a futuro, sino también las situaciones probables, las nuevas oportunidades, las barreras y los posibles desastres.

Sería erróneo pensar que la evaluación sumativa de la calidad universitaria no presta atención al cambio y a la utilidad; asimismo, está equivocado quien cree que la evaluación formativa no presta atención a los méritos del evaluando en su forma inmediata. Se acostumbra combinar la evaluación formativa y la sumativa, pero los diferentes propósitos requieren diferentes criterios, diferentes explicaciones a los informantes, y diferentes instrumentos.

Los procesos de acreditación, tanto de las universidades como de sus departamentos, realizados por asociaciones profesionales, constituyen primordialmente una evaluación sumativa, pero a nivel teórico y, a veces, funcional, es un proceso para mejorar la institución (Dressel, 1971). Algunos países confían considerablemente en la evaluación mediante acreditaciones; otros, en procedimientos estanda-

rizados de control de calidad establecidos por el gobierno central. El hecho de no obtener la acreditación puede adjudicarse a procedimientos de actuación atípicos más que a una verdadera carencia. Aún en evaluaciones bien fundadas, la validez de aquello que es formalmente acreditado es algo que hay que revisar. Cualquier expresión reduccionista (como “acreditado”) simplifica extremadamente las fortalezas y debilidades del evaluando.

A continuación presentamos una conversación hipotética que podría darse entre autoridades de una universidad a propósito de la acreditación institucional. La idea es mostrar que este proceso es bastante más complejo de lo que aparenta y que algunas veces oculta finalidades distintas a las declaradas explícitamente.

A1: Estamos pensando en cancelar nuestro contrato con dos de las tres agencias que nos están acreditando en este momento.

A2: ¿Y eso no dañará su reputación?

A1: No lo sé. Estamos de acuerdo con la mayoría de sus estándares y ser acreditados por esas instituciones puede disminuir los reclamos, pero es demasiado caro.

A2: ¿No aprenden ustedes de sus evaluaciones?

A1: Las evaluaciones refuerzan, sobre todo, lo que ya sabemos. Algunas veces aprendemos más de nuestros propios comités. El año pasado dos de estos comités verdaderamente estudiaron el problema.

A2: Los comités pueden ser inútiles y desgastan a los equipos.

A1: Y ayudan poco a nuestra reputación ante los medios de comunicación.

A2: Sugiero que encuentren mejores formas de usar los procesos de acreditación. Quizás su equipo muestra poco aprecio por la acreditación y por lo tanto la tratan más como un proceso político que evaluativo.

A1: No lo sé. Tenemos que reducir costos.

En este caso pareciera haber una mayor preocupación por la imagen política que por la información que pudiera prestar un proceso de esta naturaleza.

BARRERAS PARA LA EVALUACIÓN COMPRESIVA EN LA UNIVERSIDAD

En párrafos previos señalamos que la universidad es compleja y está lejos de ser homogénea. Aunque es casi imposible, seguimos evaluando y debiéramos hacerlo de manera comprensiva. En la siguiente sección mostraremos un ejemplo acerca de este tipo de evaluación en la docencia. La idea no es llegar a una evaluación reduccionista, sino más bien considerar sus múltiples partes, aspectos y funciones, y las diversas formas de conocerlas. En todas las instituciones se encontrarán obstáculos para la evaluación. El orden en que los presentamos no indica ningún tipo de jerarquía.

1. *Generalizada oposición a la evaluación formal.* A pesar de que las autoridades y estudiantes universitarios evalúan regularmente sus actividades de manera informal, por ejemplo, en conversaciones, a menudo no confían en que una evaluación formal sea verdaderamente objetiva. Esta actitud tiene muchos orígenes: la desconfianza habitual en la autoridad, una preocupación por la excesiva simplificación de las evaluaciones, la inexperiencia, las dudas sobre el uso de la información, entre otros.
2. *Incomprensión por parte de las autoridades de la necesidad de la evaluación formal.* Las autoridades frecuentemente asocian la evaluación formal con ciertas estrategias que implican mediciones, recolección de datos objetivos, análisis estadísticos e indicadores, todo con miras a controlar las actividades

- universitarias. A menudo deben ofrecer resultados rápidamente, por lo tanto se promueven técnicas que permitan recoger gran cantidad de información en poco tiempo.
3. *Falta de financiamiento y de preparación de parte de quienes evalúan.* A pesar de que la evaluación está siendo respaldada cada vez más por diferentes centros institucionales dedicados, al menos en parte, a la evaluación y al mejoramiento de la calidad, frecuentemente sus evaluadores tienen un limitado conocimiento de los procesos de evaluación. Las universidades, por otra parte, no tienen los recursos para desarrollar formalmente evaluaciones del conjunto de sus funciones.
 4. *Implicaciones políticas.* La evaluación no suele ser reconocida por su naturaleza política, por lo que tiene limitaciones en lo que va a ser estudiado, en las recompensas por resultados favorables y en la distribución controlada de la información (Cronbach, 1977).
 5. *Temor a la manipulación de los datos.* Las personas que proporcionan la información esperan que ésta se emplee con fines específicos, por ejemplo, para la investigación; muchas veces, sin embargo, se abusa de esta información, pues no significa lo mismo para una evaluación formativa que para una sumativa. Si las personas se dan cuenta de esto pueden malinterpretarlo y sentir desconfianza.
 6. *Efectos reactivos de la evaluación.* Involuntariamente, el proceso de evaluación modifica la conducta. Cuando un criterio llega a ser considerado importante para la toma de decisiones, la gente, consciente o inconscientemente —y, a veces en forma poco ética—, evita comportarse de manera tal que disminuya los puntajes y perjudique su situación.
 7. *Manejo ético de los datos personales.* La confidencialidad de la información es un tema complejo, especialmente para las personas evaluadas. La institución es respetuosa con la privacidad de la información personal de muchas maneras, pero en otras ocasiones, las autoridades usan información privada para respaldar decisiones contrarias a esas mismas personas.
 8. *Propaganda institucional.* La creciente necesidad de las personas y sectores, y de la universidad en su conjunto, de alardear de sus fortalezas y logros hace que se vuelva importante negar las propias debilidades y el incumplimiento de ciertos objetivos. De esta manera, la evaluación pierde una importante porción de información que necesita para su realización.
- En general, las barreras recién mencionadas son pertinentes a la evaluación de la docencia en general, y las que se exponen a continuación son especialmente evidentes cuando se trata de la evaluación de los profesores (que es el tema de la próxima sección de este artículo).
- 9a. *Concepciones de la docencia y de la evaluación de la docencia.* Cada profesor universitario tiene sus propias formas de enseñar y estas formas están relacionadas con su disciplina, con sus experiencias personales y profesionales previas, con contextos particulares, con determinados estudiantes, y algunos otros aspectos. Así, una universidad puede presentar diversas concepciones acerca de la docencia y su evaluación, pero también habrá significados comunes al respecto.
 - 9b. *Subvaloración de la docencia.* Aunque se dice que la docencia en la universidad es muy importante, ha sido tradicionalmente subvalorada en compa-

ración con otras funciones tales como la investigación y extensión académicas (Canales, 2003). Esta situación ha conducido, por una parte, al descuido y a la escasez de recursos para la docencia, y por la otra, a una falta de lineamientos rectores para su evaluación.

- 9c. *Indicadores inválidos.* Algunos indicadores para evaluar la calidad de la docencia pueden parecer válidos, tales como “claridad al hablar”, y para algunos grupos pueden correlacionarse positivamente con la efectividad de la docencia, pero desde el punto de vista de los docentes estos criterios no necesariamente son válidos para tomar decisiones administrativas.

Las barreras mencionadas pueden superarse, aunque son fuertes y están muy extendidas. Su estudio continuo debería ser parte de la investigación institucionalizada de la universidad. Los investigadores educativos podrían estudiar estas barreras de la evaluación docente, en la medida en que constituyen ejemplos parciales de cómo la evaluación se ve frustrada en todos los sectores de la universidad.

UN EJEMPLO:

LA EVALUACIÓN DE LA DOCENCIA

En esta sección nos centraremos exclusivamente en la evaluación de la docencia, que es una de las más importantes en la universidad. La evaluación de cada una de las funciones de la universidad tiene sus peculiaridades, pero las cuestiones que vamos a tratar aquí sobre la evaluación de la docencia también valen para el resto. Aunque se evalúa con el propósito de mejorar la calidad, este ejercicio también puede evitar que el trabajo docente sea visto de forma parcial y sirve igualmente para buscar las fallas con el fin de repararlas (Lehmann, 1975).

De acuerdo a los actuales procedimientos de evaluación formal, la docencia en su conjunto, o la de un departamento en particular, puede ser evaluada casi exclusivamente por la suma de la calidad de la docencia de cada profesor individual. El auto-estudio holístico del desempeño del personal docente y los comités de revisión son posibles, pero la evaluación directa está fuera de alcance. Lo que sí es posible llevar a cabo es la evaluación efectiva de la calidad docente del profesorado en forma individual.

La complejidad de la docencia. Una posible definición de las muchas posibles de docencia es la de una actividad humana compleja que se adecúa a **situaciones comunes**, pero a la vez siempre diferentes, y su consecuencia debería ser el aprendizaje de los estudiantes. Harry Broudy (1963) identificó los modos de enseñanza en la escuela de esta manera: didáctico, heurístico y filial. El modo didáctico se refiere a la enseñanza directa, en la que se imparten conocimientos y habilidades; la enseñanza indirecta se refiere a la disposición de circunstancias de tal manera que el estudiante llegue a ser mejor educado. El modo heurístico se refiere a la enseñanza de principios generales a partir de los cuales un conocimiento específico puede ser deducido. En el modo filial se establece una relación entre profesor y estudiante que resulta en una motivación y posiblemente también despierte en el estudiante el amor por aprender lo que el profesor enseña. La buena enseñanza incluye interpretación del currículo, motivación, programación, planeación de tareas, supervisión de las interacciones en el aula, evaluación, calificación, colaboración con los condiscípulos, etc. (Darling-Hammond y Youngs, 2002; Arbesú *et al.*, 2007; David, 2010).

Algunos profesores permiten la observación de sus clases, pero la mayoría del profesorado universitario prefiere, por tradición, privacidad, y puede rechazar la presencia de observadores incógnitos o monitores

electrónicos en el aula. Probablemente este tipo de privacidad contribuye a una práctica docente efectiva por cuanto los profesores se sienten seguros y protegidos en el ámbito de la “libertad de cátedra”; sin embargo, también puede esconder una práctica deficiente. Las autoridades universitarias y el personal administrativo interesados en mejorar la docencia tienen el derecho de conocer el contenido y los métodos de cada profesor, pero la supervisión presencial, constante o incógnita, violenta la privacidad del docente y puede convertirse en un impedimento para el mejoramiento de la enseñanza.

Enseñanza comunitaria. La idea tradicional de evaluación de la docencia corresponde a la evaluación de un docente autónomo en una sola aula. Una perspectiva adicional puede encontrarse en la práctica, al evaluar la contribución que puede hacer un profesor a la revisión y mejora de los programas docentes de su departamento o unidad, pues lo que los profesores pueden hacer por la integridad de toda la propuesta docente, no sólo de la suya, también es importante. Un docente carismático, un laboratorista innovador o un tutor comprometido, es decir, una estrella solitaria en el firmamento, a veces contribuye poco a superar los frágiles, erráticos, frívolos y caducos cursos del departamento; por ello, cuando se evalúa la docencia

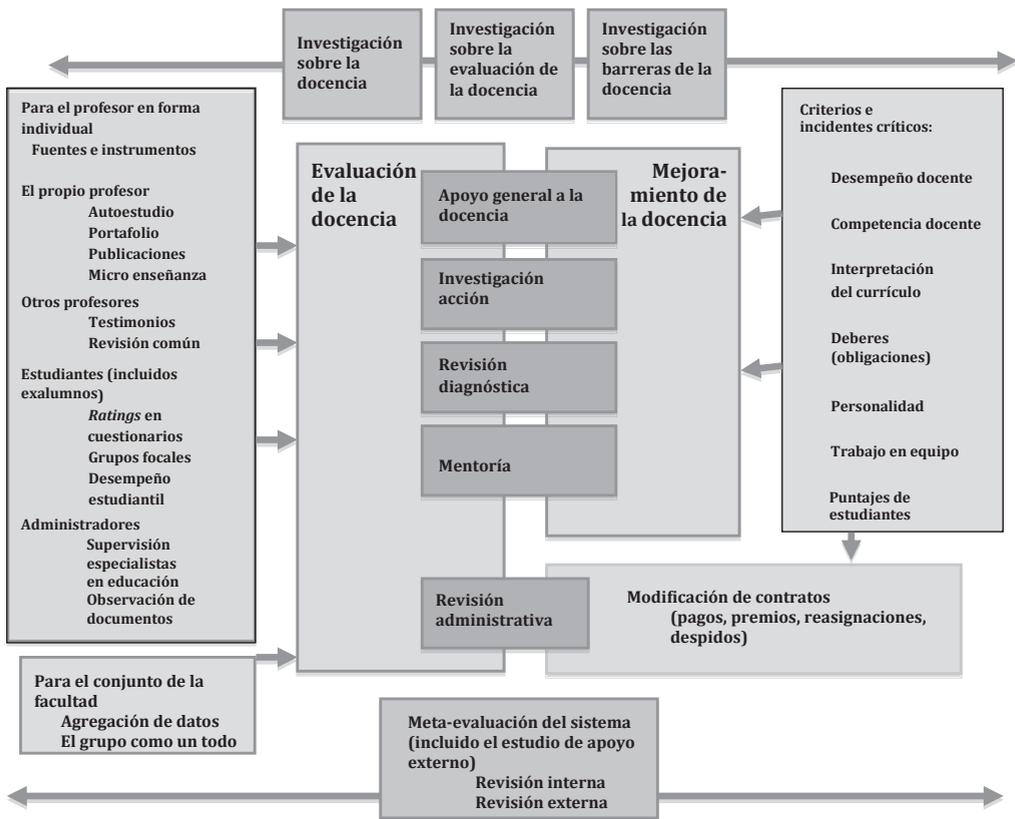
deben ser consideradas tanto las contribuciones individuales como las colectivas.

La colaboración entre los docentes de la universidad sobre temas de docencia no es nueva (Cruz, 1999), pero en la mayoría de los casos es más bien la excepción que la norma. Hablar de una facultad como “comunidad de práctica” nos remite a lo que algunos miembros del Instituto para la Investigación del Aprendizaje (Institute for Research on Learning) de Palo Alto (EUA) han trabajado. Nuestra colega Etienne Wenger (1991: 7) afirma al respecto:

... incluso aquellos que hablan de organizaciones de aprendizaje, de formación permanente o de la sociedad de la información, lo hacen generalmente en términos de aprendizajes individuales y procesos de información. La noción de comunidades de prácticas nos ayuda a romper este molde.

Una representación de la evaluación de la docencia. La Fig. 1 es una representación gráfica de la evaluación formal e informal de la docencia en la educación superior. Debe considerarse más como una visualización de la responsabilidad en la evaluación que como una guía. Presenta un ejemplo de rendición de cuentas de los méritos y deficiencias de las funciones de la universidad.

Figura 1. Representación de la evaluación formal e informal de la docencia en la educación superior



Los dos grandes rectángulos centrales representan la responsabilidad en la evaluación y mejoramiento de la docencia. Aunque están potencialmente integrados, se mantienen separados en este esquema porque son responsabilidades diferentes. A veces emergen como la responsabilidad correspondiente a la evaluación formativa-comprensiva, sin embargo la evaluación sumativa también sirve a la revisión administrativa de la docencia, como se indica más abajo. Esto no quiere decir que los mismos datos servirán a los propósitos formativo y sumativo; de hecho, los datos para mejorar la enseñanza no tendrían que estar inmediatamente a disposición de las autoridades universitarias porque puede haber mejores resultados si la información sobre las debilidades de los docentes es confidencial.

Los cuatro rectángulos verticales del centro representan cuatro áreas estratégicas de apoyo a la universidad: 1) proveer consejos para el mejoramiento; 2) promover el auto estudio; 3) proporcionar servicios de medición sobre la efectividad de la docencia; y 4) apoyar el tutelaje de la docencia por alguien más experimentado en enseñanza de esa disciplina. El cuadro inferior indica los distintos caminos que se siguen para la revisión administrativa del control de la calidad, cambiando ampliamente las prácticas docentes y modificando los contratos de los profesores.

Como veremos, la evaluación de la docencia comienza con el auto estudio reflexivo del profesor, pero después se amplía a otras fuentes y datos (Brubacher *et al.*, 2005). El rango de datos para la evaluación se muestra en el

lado izquierdo. Éstos provienen de distintas fuentes, habitualmente desde el docente que está siendo evaluado por sus superiores, pero también provienen de sus colegas de departamento, de los estudiantes y de los especialistas en docencia. El uso de una sola o dos de estas fuentes probablemente derivará en una visión estrecha de la calidad de la docencia. En el lado derecho están los criterios e incidentes críticos que serán descritos en la siguiente sección. Estas responsabilidades de la evaluación necesitan ser tomadas en un contexto de investigación y experiencia profesional, incluyendo lo que es pertinente a cada universidad en particular.

Otras responsabilidades importantes se muestran en los cuadros pequeños. Parte de la evaluación de la docencia (cuadro de abajo a la izquierda) pertenece al colectivo de profesores del departamento o facultad, ya sea estudiándolo como suma del desempeño docente individual de sus profesores o por su estudio como una unidad. Los tres cuadros en la parte superior indican la responsabilidad de desarrollar investigación en docencia y en la evaluación de la docencia, así como los obstáculos a la docencia (mencionados anteriormente). El último cuadro indica la necesidad permanente de apoyar la evaluación con meta-evaluación, es decir, con evaluación de la evaluación (Stufflebeam y Shinkfield, 2007).

Cabe señalar una vez más que ésta no es una guía de evaluación práctica, sino más bien la identificación de las responsabilidades compartidas por individuos y grupos, y por la universidad en su conjunto. Muchas de esas responsabilidades serán comprendidas por la evaluación informal. Los recursos insuficientes para la evaluación formal y para la asesoría en el mejoramiento de la docencia son factores que limitan el compromiso para realizar estas numerosas obligaciones; a pesar de esto, resultaría positivo reconocer el error de usar sólo uno o dos indicadores. Los criterios y los incidentes críticos deben ser comprensivos.

Criterios e incidentes críticos. Dos son las bases para evaluar la docencia, cada una proveniente de la perspectiva cuantitativa y la cualitativa, respectivamente: los criterios y los incidentes críticos. En un sistema sofisticado, a ambos se les debe prestar atención.

Un criterio es una propiedad distintiva o característica de algo, mediante la cual se puede juzgar o estimar su calidad, se puede hacer una clasificación o se puede tomar una decisión (Sadler, 2005). Un criterio a menudo es representado mediante una escala que va de menos a más. La buena evaluación formal es a menudo organizada alrededor de múltiples criterios y derivada de múltiples fuentes de información.

Un incidente crítico es un evento o condición por medio del cual la calidad puede ser juzgada (no hay que confundirlo con la técnica de incidentes críticos [Flanagan, 1954]). A veces un solo comportamiento o comentario crítico es tan importante que puede trascender las otras virtudes del evaluando y constituirse en la principal base para una determinada acción, por ejemplo el despido o, en el caso de una institución, la pérdida de la acreditación. En el año 2008, la Universidad de Iowa retiró a un profesor de sus clases y deberes después de ser acusado por un estudiante de haber tenido una conducta inapropiada (Jaschik, 2008); después de un largo periodo de aislamiento, ambos implicados se suicidaron. Las dos tragedias fueron consideradas al evaluar a la institución. Desde luego que otros sucesos más comunes y menos fatídicos pueden ser positivos, por ejemplo, un profesor que salga a defender a alguien. En un informe, este tipo de incidentes se representaría a través de la descripción detallada de esas experiencias puestas en contexto, más que mediante una medición.

La pertinencia de los criterios y de los incidentes críticos para la evaluación depende del propósito de la misma, de las personas que están evaluando, de la institución, de la disciplina que se enseña, etc. (Cisneros-Cohernour, 1997; Canales, 2003). Los criterios e incidentes

críticos pueden ser usados conjuntamente pero rara vez convergerán en representaciones invariables de la calidad de la docencia.

Los criterios e incidentes críticos más comunes para evaluar la docencia universitaria son:

1. Desempeño del profesor. Juzgar el desempeño implica juzgar lo que el profesor realmente hace.
2. Competencia del profesor. Evaluar la competencia implica valorar o medir lo que el docente es capaz de hacer. Por ejemplo, mediante una entrevista y la revisión de portafolios es posible evaluar la comprensión del docente sobre el tema que enseña.
3. Interpretación del currículo. Se juzga la docencia a través de la forma en que los profesores priorizan e interpretan los tópicos contenidos en el currículo. Aquí figura, por ejemplo, la capacidad de cubrir todo el plan de estudios y evitar ofrecer interpretaciones controvertidas de los temas.
4. Deberes. Los criterios acerca de qué es lo que el profesor está obligado a hacer son fijados *a priori* por contrato, ética profesional o costumbres. Aquí tenemos aspectos tales como la preparación y la asistencia a las clases (Scriven, 1995).
5. Personalidad. Los criterios se basan en ciertas características ideales del profesor, como la empatía, el sentido del humor o la relación con los estudiantes.
6. Trabajo en equipo. El trabajo en equipo es una competencia específica relacionada con el trato personal y el compromiso para el trabajo en colaboración con otros, particularmente con los miembros del departamento.
7. Rendimiento de los estudiantes. Una manera de juzgar la docencia es basarse en las calificaciones de los estudiantes, proyectos, desempeños o puntajes

en exámenes estandarizados. Se debe tomar en cuenta el valor agregado comparándolo con los puntajes que han obtenido los estudiantes a través de los años (Rosu y Stake, 2008). En este caso, la aptitud de los estudiantes es un factor importante a tener en cuenta.

Cualesquiera que sean los criterios o incidentes críticos usados, las personas responsables de la evaluación del trabajo docente deben compartir una manera de comprender lo que implica juzgar la docencia. Esto puede ser un problema debido a que la mayoría de los profesores de educación superior tiene un amplio conocimiento de sus respectivas disciplinas, pero no de los métodos pedagógicos (Troncoso y Hawes, 2006). En su disciplina, cada profesor tiene un lenguaje y una práctica distintivos pero no así en la docencia, y esto puede conducir a interpretaciones problemáticas sobre su calidad. Más aún, la docencia es altamente contextualizada, de manera que los profesores interpretarán el mismo criterio de distintas maneras porque los contextos (aulas, medios, laboratorios) son distintos.

Asesoría de especialistas. Es frecuente que las universidades dispongan de una oficina de servicios docentes, en particular para profesores jóvenes y extranjeros; por ejemplo, en la Pontificia Universidad Católica de Valparaíso (Chile) existe un centro cuyo propósito más importante es elevar la calidad de la docencia y de los aprendizajes de los estudiantes. Para este fin, el centro cuenta con diversas unidades que trabajan en evaluación de la docencia, calidad docente y desarrollo de nuevos currículos y procesos de acreditación. Algunos de estos servicios, como la formación docente, se ofrecen con carácter voluntario, pero algunas carreras han ido incorporando en forma creciente a sus profesores más jóvenes en este tipo de programas.

Algunas universidades públicas de la Ciudad de México tienen centros dedicados a evaluar y mejorar la docencia, con

especialistas que apoyan a los docentes. En algunas instituciones, como la Universidad Autónoma Metropolitana (UAM) unidad Xochimilco, no hay equipos de especialistas que evalúen la calidad de la docencia o que proporcionen formación docente. Los profesores son evaluados informalmente por administradores y colegas designados por el director de la división académica, y formalmente mediante un cuestionario aplicado a los estudiantes. Los resultados, con cierta frecuencia, se usan para otorgar algún tipo de compensación económica. Respecto de la evaluación formativa hay, en ciertas ocasiones, reuniones entre profesores para tratar sus problemas y, ocasionalmente, los investigadores educativos se reúnen para discutir distintas prácticas pedagógicas en la universidad. Adicionalmente, los profesores, generalmente en forma individual y en ciertas ocasiones en grupos, están involucrados en alguna forma de auto evaluación, algunos usando el portafolio docente.

Evaluación formativa de la docencia. La evaluación consiste, en este caso, en llegar a conocer la calidad del acto de enseñar, encontrando méritos y debilidades en cada aspecto de la docencia. Está diseñada para ayudar a los profesores a mejorar su trabajo diario y a cumplir los objetivos del curso. El aprendizaje de los estudiantes puede o no ser un indicador de la calidad de la docencia (Cisneros-Cohernour, 1997). Tal como están las cosas actualmente, los profesores son los actores más importantes en este proceso porque son los únicos que están en condiciones de observar y reflexionar sobre sus propias prácticas —a veces con la ayuda de especialistas o colegas— y decidir qué cambios se realizarán (Arbesú, 2006). En ocasiones, alguna administración preocupada por el cambio y por la protección de la privacidad apoya la evaluación formativa. La mayoría de las veces, sin embargo, las autoridades prefieren la evaluación sumativa. En otras, investigadores y autoridades coinciden en que la mejora de la educación se promueve a través de la reflexión

sobre la práctica y que esta reflexión generará, a su vez, conocimiento acerca de la docencia (Brubacher *et al.*, 2005; Rothstein, 2008; David, 2010).

La mayoría de los investigadores que estudian la evaluación formativa de la docencia están convencidos de que ésta es útil para: 1) cambiar los métodos de enseñanza; 2) comprometer a los estudiantes a mejorar la calidad del aprendizaje en el aula; 3) promover un espíritu de educación profesional permanente; 4) ayudar a la comunidad académica (estudiantes, profesores, autoridades y administrativos) a comprender las complejidades de las prácticas educativas, en particular la docencia (Cronbach, 1963). A pesar de que para los profesionales está claro cómo sustentar estos usos, las evidencias no son suficientemente sólidas como para convencer a las autoridades de suministrar los recursos necesarios para una evaluación formativa de la docencia. Los problemas que dicha evaluación enfrenta quedarán claros en la siguiente sección, que examina el mejoramiento en la universidad cuando se tiene una base de datos de la actividad evaluadora.

Informalmente, todos evaluamos la docencia. La evaluación informal se construye a través de percepciones y experiencias de las personas que participan y observan los procesos educativos: funcionarios públicos, profesores, estudiantes, autoridades, personal administrativo, etc. Este tipo de evaluación ocurre en todas las áreas de la universidad y en actividades curriculares y extracurriculares.

Formalmente, la evaluación de la docencia toma en cuenta ciertos objetivos y criterios que la universidad usa para evaluar la calidad docente. Los investigadores en educación han evaluado la docencia por muchos años, pero este campo de conocimiento está aún en construcción (Arbesú *et al.*, 2007; Ashcroft y Palacio, 2010). La evaluación de la docencia necesita métodos formales e informales, ya que ambos ayudan a los interesados a afinar la percepción sobre sus logros y problemas.

A través de este artículo hemos destacado la complejidad de la educación, de la docencia y, por lo tanto, de la evaluación. Ambos tipos de evaluación, formal e informal, son útiles para comprender las actividades docentes. En la siguiente parte se presentan las estrategias para pasar de la evaluación al mejoramiento de las funciones de la universidad. Se destaca, nuevamente, la complejidad de la universidad y de las personas que trabajan en ella.

EL ROL DE LA EVALUACIÓN EN LA MEJORA DE LAS FUNCIONES DE LA UNIVERSIDAD

Los defensores de la evaluación son optimistas al pensar que la recolección de información acerca de las funciones de la universidad llevará a su mejoramiento. A veces sucede, pero hasta este momento no existe una ciencia de la gestión universitaria ni una teoría de la educación superior que tenga características diagnósticas. Es posible aprender mucho acerca de cómo funciona la universidad sin comprender qué puede hacerse para mejorarla. De la evaluación, por ejemplo, el rector puede aprender que el Departamento de Ciencias Forestales está a merced de intereses comerciales, pero no encuentra ni el poder ni la estrategia para cambiar esa situación.

La Fig. 1 ilustra las responsabilidades de la evaluación formativa de la docencia, pero no muestra cómo el conocimiento emanado de la evaluación puede conducir al mejoramiento de la docencia. Se puede decir lo mismo acerca de todas las demás funciones de la universidad. El manejo de las universidades depende en gran parte de la experiencia e intuición, no obstante lo falibles que puedan resultar en ciertas situaciones.

Una evaluación formal de la mayoría de las funciones universitarias, en un sentido global y amplio, se encuentra por encima de las actuales capacidades de la universidad. Como se indicó anteriormente, una evaluación completa

de la docencia resulta prácticamente imposible, pero se puede mejorar la existente. Las funciones de la universidad tradicionalmente se evalúan integralmente de manera informal, pero algunas requieren evaluarse más formalmente. Los procedimientos informales son falibles, y lo mismo puede decirse de los formales. La validación y supervisión de responsabilidades son necesarias para la existencia y ampliación de los mecanismos de evaluación.

Lo que una buena evaluación revela no es sólo la calidad de las actividades, sino también su complejidad. Cuando se acepta que las relaciones por sí mismas no explican mucho, es más fácil oponerse a las soluciones simples. Cada función tiene complejidades educativas, financieras, políticas, ecológicas, estéticas, etc.; unas más evidentes que otras. La investigación conduce a la búsqueda de postulados sobre lo que podría intentarse y a la búsqueda de ejemplos ahí donde han sido puestos en práctica.

La declaración de la misión de una universidad usualmente no es de mucha ayuda, pues nos dice qué es aquello que ha tenido éxito en vez de expresar qué es lo que necesita mejorar. Sirve más para inspirar y entusiasmar que para establecer directrices. Hay problemas que apenas menciona y otros que omite totalmente. Se necesitan evaluaciones más realistas de los problemas, temas y obstáculos para una mejor gestión. Es difícil superar las debilidades en un ambiente de autocomplacencia e intereses mercadotécnicos. Las presiones para ofrecer una imagen comercialmente atractiva de la universidad han crecido de forma considerable, pero los buenos líderes otorgarán más atención a la reflexión escéptica, fresca y duradera. La calidad no queda suficientemente definida en un mosaico de premios y escalas. El valor real de una universidad es mucho más mundano. Los ideales optimistas y el criticismo escéptico son indicadores de salud institucional.

Estrategias para la evaluación formativa

Hay cuatro estrategias disponibles para pasar de la evaluación al mejoramiento real de las actividades universitarias; incluyen la docencia propiamente dicha pero se extienden a todas las funciones de la institución. Las cuatro son necesarias, pero probablemente no se les dará igual prioridad. En cada caso hay necesidad de atender la heterogeneidad de aquellos que tienen poder de decisión en la universidad y que controlan los recursos disponibles.

1. *La estrategia basada en objetivos.* Los responsables de la gestión y el funcionamiento de la universidad llegan a acuerdos en cuanto a lo que la universidad debe ser, y priorizan las debilidades que deben ser superadas y los adelantos a implementarse.
2. *La estrategia médica.* Los responsables de la gestión y el funcionamiento de la universidad llegan a acuerdos acerca de los aspectos en los que la institución no está a la altura de las expectativas previstas, lo cual conlleva la pérdida de oportunidades y de alguna manera daña a su comunidad, y priorizan las debilidades en las cuales trabajar.
3. *La estrategia de estudio sectorial.* Se otorga a los departamentos, unidades y otras colectividades apoyo creciente para el reconocimiento y superación de las debilidades para que lleguen a acuerdos acerca de qué aspectos se están trabajando sin problemas y cuáles de forma ineficaz o disfuncional.
4. *La estrategia de auto estudio.* Se otorga apoyo creciente para el reconocimiento y superación de las debilidades individuales de cada miembro del personal administrativo y del profesorado para el auto estudio y la coordinación de acciones de mejoramiento.

Sin importar cuán **detallada sea la investigación**, la estrategia para el mejoramiento no provendrá directamente de los indicadores o escalas. La investigación sobre la docencia puede ser prometedora, pero no diagnóstica. No hay más alternativa que recolectar una gran cantidad de información formal e informal para el análisis y luego especular sobre diferentes cuestiones, como por ejemplo, en dónde sería más conveniente destinar los recursos. A menudo los miembros del equipo pueden tener buenas sugerencias sobre los cambios que son factibles.

El todo y las partes de una universidad están diseñados con una singularidad característica, con bienes y contextos disímiles; las universidades difieren en el centralismo de su actividad rectora; es así que no puede esperarse que la evaluación y el mejoramiento vayan a ser los mismos, ni tampoco que estén visiblemente coordinados a través de toda la universidad. En cuanto a la función docente, el modelo de auto-estudio parece tener la mejor expectativa de éxito. A menudo, la función principal de la administración central es demostrar con convicción la necesidad de la evaluación y animar a los responsables de cada sector a trazar sus propias estrategias.

La universidad necesita evaluarse tanto formativa como sumativamente y reconocer que es una institución que sigue madurando y que sus virtudes y defectos son tan complejos que escapan a una medición fácil. Aun las mejores universidades tienen obstáculos por superar. La evaluación no es una garantía de protección contra el desorden ni un control de calidad del crecimiento, pero los esfuerzos conjuntos por buscar la calidad en la universidad pueden contribuir a una gestión razonada. El más grande indicador de calidad de una universidad puede ser la introspección de sus autoridades y de su personal administrativo y académico.

REFERENCIAS

- AHIRE, S.L. (1997), "Management Science-Total quality Management Interface: An integrative framework", *Interface*, vol. 27, núm. 6, pp. 91-105.
- ARBESÚ, M.I. (2006), *La práctica de la docencia modular: el caso de la unidad Xochimilco en la Universidad Autónoma Metropolitana*, México, UAM/Plaza y Valdés.
- ARBESÚ, M.I., F. Díaz Barriga, L. Elizalde, E. Luna, M.A. Rigo, M. Rueda y A.D. Torquemada (2007), "La evaluación de la docencia universitaria en México: un estado de conocimiento del periodo 1990-2004", en *Perspectiva Educacional*, núm. 48, pp. 27-58.
- ASHCROFT, K. y D. Palacio (2010), *Researching into Assessment and Evaluation in Colleges and Universities*, Londres, Routledge.
- BROUDY, H.S. (1963), "Historic Exemplars of Teaching Method", en N.L. Gage (ed.), *Handbook of Research on Teaching*, Chicago, Rand McNally, pp. 1-43.
- BRUBACHER, J.W., C.W. Case y T.G. Reagan (2005), *Cómo ser un docente reflexivo. La construcción de una cultura de la indagación en las escuelas*, Madrid, Gedisa.
- CANALES, A. (2003), "Docencia: tensiones estructurales en la valoración de la actividad", en M. Rueda, F. Díaz-Barriga y M. Díaz (eds.), *Evaluar para comprender y mejorar la docencia en la educación superior*, México, Universidad Autónoma Metropolitana/UNAM/Universidad Autónoma de Baja California, pp. 71-76.
- CISNEROS-Cohernour, E.J. (1997), "Trade-Offs: The use of student ratings results and its possible impact on instructional improvement", University of Illinois (mimeo).
- CONTRERAS, G. (2010), "Diseño y operación de un sistema de evaluación del desempeño docente con fines formativos: la experiencia de la Pontificia Universidad Católica de Valparaíso, Chile", *Revista Iberoamericana de Evaluación Educativa*, vol. 3, núm. 1e, en: http://www.rinace.net/riee/numeros/vol3-num1_e/art14.pdf (consulta: 20 febrero 2011).
- CRONBACH, L.J. (1963), "Course Improvement through Evaluation", *Teachers College Record*, núm. 64, pp. 672-683.
- CRONBACH, L.J. (1977), "Remarks to the New Society", *Evaluation Research Society*, núm. 1, pp. 1-3.
- CRUZ Souza, F. (1999), *Psicología comunitaria*, Buenos Aires, Lumen/Hvanitas.
- DARLING-Hammond, L. y P. Youngs (2002), "Defining 'Highly Qualified Teachers': What does 'scientifically-based research' actually tell us", *Educational Researcher*, vol. 31, núm. 9, pp. 13-25.
- DAVID, J.L. (2010), "Using Value-Added Measures to Evaluate Teachers", *Educational Leadership*, vol. 67, núm 8, pp. 81-82.
- DRESSEL, P.L. (1971), "Accreditation and Institutional Self-study", *North Central Association Quarterly*, núm. 6, pp. 277-287.
- FLANAGAN, J.C. (1954), "The Critical Incident Technique", *Psychological Bulletin*, vol. 51, núm. 4, s/pp.
- JASCHIK, S. (2008), "Accusations and Suicides", *Cronicle of Higher Education*, núm. 11, p. 14.
- LEHMANN, I.J. (1975), "Evaluating Teaching", en W.J. Gephart, R.B. Ingle y G. Saretzky (1975), *The Evaluation of Teaching: National Symposium for Professors of Educational Research*, Bloomington, Phi Delta Kappa, s/pp.
- ROSU, L.M. y R.E. Stake (2008), "The Assessment of Teaching and Learning in USA: Reflections on school quality and the role of teacher in public schools", *Romanian Pedagogical Journal*, vol. 1, núm. 1, pp. 49-66.
- ROTHSTEIN, J. (2008), *Student Sorting and Bias in Value Added Estimation: Selection on observables and unobservables*, Princeton, Princeton University/National Bureau of Economic Research.
- SADLER, R. (2005), "Interpretations of Criteria-Based Assessment and Grading in Higher Education", *Assessment & Evaluation in Higher Education*, vol. 30, núm. 2, pp. 175-194.
- SCRIVEN, M. (1995), "A Unified Theory Approach to Teacher Evaluation", *Studies in Educational Evaluation*, núm. 21, pp. 111-129.
- STAKE, R.E. y E.J. Cisneros-Cohernour (2000), "Situational Teaching Evaluation of Teaching on Campus", *New Directions for Teaching and Learning*, núm. 83, pp.51-72.
- STAKE, R.E. y T.A. Schwandt (2006), "On Discerning Quality in Evaluation", en I.F. Shaw, J. Greene y M. Mark (eds.), *Handbook of Evaluation*, Londres, Sage, pp. 404-418.
- STUFFLEBEAM, D.L. y A.J. Shinkfield (2007), *Evaluation Theory, Models, and Applications*, San Francisco, Jossey-Bass.
- TINTO, V. (1995), "El abandono en los estudios superiores: una nueva perspectiva de las causas del abandono y su tratamiento", México, UNAM/ANUIES, Cuadernos de Planeación Universitaria.
- TRONCOSO, K. y G. Hawes (2006), "A propósito de la evaluación por pares: la necesidad de semantizar la evaluación y las prácticas docentes", *Perspectiva Educacional*, núm. 48, pp. 59-73.
- WENGER, E. (1991), "Communities of Practice: Where learning happens", *Benchmarks*, otoño, pp. 6-8.