

El proceso perverso de la evaluación de la docencia en las universidades: Un balance inicial y apuntes para mejorarlo*

José María García Garduño¹

El proceso perverso de la evaluación de la docencia en las universidades: Un balance inicial y apuntes para mejorarlo
José María García Garduño (2008)
Reencuentro
53:9-19

The perverse process of the valuation of the teaching in the universities: An initial rocking and begin to appear to enhance it
José María García Garduño (2008)
Reencuentro
53:9-19

Resumen

El propósito de este trabajo es analizar las limitaciones que tienen los cuestionarios de evaluación de la docencia y sus efectos en la profesión académica. Dichas limitaciones están asociadas con la complejidad de definir una docencia eficaz, la métrica que se emplea para medirla, su relación con los programas de incentivo y la validez de los cuestionarios de evaluación. Se proponen alternativas para que la evaluación de la docencia sirva al propósito de retroalimentar la docencia universitaria.

Palabras clave

Evaluación de la docencia, profesores universitarios, incentivos, validez de los cuestionarios de evaluación.

Abstract

The purpose of this handwork is to analyze the limitations between the teaching evaluation questionnaires and its effects in the academic profession. Such limitations are associated to the complexity of defining an efficient teaching, the metric used to measure it, its relation with the incentive programs and the validity of the evaluation questionnaires. The author proposes several alternatives for the teaching evaluation to serve the purpose of feeding back the university teaching.

Key words

Teaching Evaluation, University teachers, Incentives, Validity of the evaluation questionnaires.

A partir de las década de los noventa, la mayoría de las instituciones de educación superior del país implantó sistemas de evaluación de la docencia; dicha implantación no fue una decisión interna o espontánea. Bajo la plataforma del Programa para

la Modernización Educativa,² el gobierno federal promovió políticas de estímulos o pago por méritos. Posiblemente, esta política no sólo fue la adopción

² Poder Ejecutivo Federal (1989). *Programa para la modernización educativa*. México.

* Parte de este trabajo fue realizado en la Universidad Autónoma del Estado de Hidalgo con el apoyo de una beca PROMEP.

¹ Profesor Investigador de la Universidad Autónoma de la Ciudad de México, Plantel Cuauhtepc. josemariagarduno@yahoo.com.mx

de un sistema que ya se había implantado en los sistemas educativos de otros países, como Estados Unidos, y en algunas universidades mexicanas, privadas y públicas, sino también un medio para contrarrestar la inercia y conflictos originados por la demanda de instituciones de educación media y superior, de homologación salarial con la UNAM e IPN, y el abandono de la profesión docente, debido a la contracción salarial ocurrida en los años ochenta, la *década perdida*.

Junto con la política de incentivos a la profesión académica, o pago por méritos, apareció, de manera necesaria, una política de evaluación y rendición de cuentas del quehacer sustantivo de las instituciones y sus académicos. Desde luego, no todas las actividades sustantivas del trabajo académico: docencia, investigación y extensión o difusión de la cultura, se valoran por igual. En México, al igual que ocurre en otras latitudes, las funciones básicas del trabajo académico como la investigación tiene un peso mayor que la docencia en la evaluación del desempeño académico y, en consecuencia, en el otorgamiento de estímulos (Arbesú, Canales, *et al.*, 2004). Sin embargo, bajo diferentes ponderaciones, la evaluación del desempeño docente es un factor que influye en el monto de los estímulos académicos, en la mayoría de las instituciones universitarias mexicanas. Si bien en el programa de estímulos se consideran varios factores en la evaluación del desempeño docente, como la elaboración de programas y materiales didácticos, la parte más notable es la evaluación de la tarea docente por medio de las opiniones de los alumnos vertidas en los cuestionarios de evaluación al final de los periodos académicos. Esta asociación entre desempeño docente y su evaluación por parte de los estudiantes ha traído consigo un proceso perverso en las instituciones que lo han adoptado. Una de las acepciones de esta palabra, de acuerdo con el diccionario de la Real Academia Española, es “que corrompe las costumbres o el orden y estado habitual de las cosas”. Este trabajo tiene como objetivo analizar las consecuencias y debilidades de la evaluación de la docencia y proponer algunos lineamientos

iniciales para modificar ese sistema que se ha transformado en una práctica meramente burocrática y que poco ha aportado al mejoramiento de la calidad docente; por el contrario, como lo señalan diversos estudios internacionales, puede estar contribuyendo a debilitarla.

El programa de estímulos y su relación con el desempeño docente

Si bien los programas de incentivos o recompensa pueden resultar atractivos y benéficos para cierto tipo de profesiones o trabajos, una de sus limitaciones es que están asociados con la motivación extrínseca, con una recompensa. En esa medida, como señala Etzioni (1961), el poder remunerativo implica acatamiento calculado de la tarea. ¿Qué tanto me conviene hacer eso o aquello? ¿Cuánto dinero o puntos me dará? La remuneración está asociada con la motivación extrínseca (Ryan y Deci, 2000). La motivación extrínseca despoja al profesor de su poder de autodeterminación o autonomía, uno de los valores profesionales más enraizados en la profesión académica desde la universidad medieval.

Al respecto, analizando la situación de los estímulos, Ibarra y Porter señalan que:

“Al introducirse la variable económica en el modelo, como sucedió desde que implementaron las políticas de deshomologación, se provocó un desplazamiento de las finalidades sustantivas por las consecuencias materiales, produciéndose una gran confusión sobre las motivaciones del trabajo académico” (2007: 35)

Ello ha conllevado al profesor a actuar en el aula y otros ámbitos de su labor, en función de las políticas de incentivos o de estímulos, las cuales no siempre coinciden con lo que él o ella considera más importante para la enseñanza y el aprendizaje de sus alumnos. Otro factor que contribuye a la confusión o ineficacia de esa política es la métrica, ya que una condición necesaria para que el pago por incentivos funcione adecuadamente, es que

la métrica o medida del producto esté claramente especificada (Lazear, 2003); es difícil evaluar la docencia en la educación superior, debido a que es casi imposible identificar con claridad la métrica o medida de una docencia eficaz

El debate internacional sobre la evaluación de la docencia

Lo que se diga o investigue en Estados Unidos acerca de la evaluación docente, como en otras áreas de la vida económica y social, para bien o para mal, tiene repercusiones internacionales.³ De ahí la importancia de analizar lo que ocurre en ese país y en otros países como Canadá y Australia, donde la evaluación docente es una línea importante de investigación dentro de la educación superior. En los Estados Unidos, a principios del siglo pasado, surgió la evaluación del desempeño docente. El medio más popular y barato para llevarla a cabo ha sido el uso de cuestionarios en los que se solicita a los estudiantes que evalúen el desempeño del profesor. Su uso comenzó a popularizarse en las tres últimas décadas; el activismo estudiantil de los años sesenta (McKeachie, 1996) y la orientación al mercado que tienen las instituciones estadounidenses, hicieron proclive su popularidad. La evaluación se ha convertido en una práctica casi universal en el ámbito de la educación superior estadounidense. Asimismo su uso se popularizó en el ámbito internacional, a partir de las dos últimas décadas.

Esa popularidad no sólo fue en su empleo en las aulas universitarias y en los comités de promoción, sino también en la investigación. La evaluación de la docencia es el área más investigada de la educación superior. Feldman (1990) calculaba que ya a principios de la década pasada, se habían producido más de 2000 publicaciones sobre el tema; las décadas doradas de la investiga-

ción fueron los años setenta y los ochenta (Centra, 1993). Buena parte de la investigación estuvo dedicada a determinar la validez y confiabilidad de esos instrumentos, y su relación con otras variables, como rendimiento académico y características de los profesores (Marsch, 1987). Realmente, en un principio, su uso no implicaba gran conflicto para los profesores. El uso de los cuestionarios nació con un fin meramente formativo, los resultados tenían la finalidad de retroalimentar al profesor acerca de su desempeño; aunque una parte sustancial de los profesores estadounidenses y de otras latitudes, incluyendo la nuestra, se han resistido a su empleo porque sienten que es poco útil o injusto ser evaluados por sus alumnos, ha existido una aceptación o subordinación tácita a tal política.

En la medida en que los resultados de los cuestionarios de evaluación paulatinamente se fue empleando como criterio para otorgar definitividad y promoción en las universidades estadounidenses, las voces de protesta se empezaron escuchar con más vigor. Entre esas voces, destacan la de McKeachie (1979), psicólogo de la Universidad de Michigan, uno de los primeros en lanzar críticas a la valoración de la efectividad docente por medio de los cuestionarios. Desde los años setenta empezaron a surgir investigaciones aisladas que ponían en duda la validez de los cuestionarios, porque se consideraban medidas sesgadas e incompletas del desempeño docente. Aunque el grueso de la investigación no fue concluyente, los sesgos más conspicuos que se identificaron fueron a) la disciplina: profesores de las ingenierías y ciencias duras obtienen menores puntuaciones en los cuestionarios que las humanidades y ciencias sociales; b) la personalidad del instructor: los profesores entusiastas, extrovertidos, con sentido del humor obtienen mejores puntuaciones que los que no poseen o manifiestan esos rasgos de personalidad; c) etapa de los estudios que cursa el estudiante: los estudiantes de los primeros semestres valoran más que el profesor sea accesible, ameno, claro y organizado; los estudiantes avanzados valoran más el dominio del conocimiento por parte del docente (García, 2000a).

³ No debemos olvidar que sus creencias acerca de su destino manifiesto y la desunión de los mexicanos nos costaron la mitad del territorio en el siglo XIX. En el ámbito de la educación, el sistema de educación superior es el más importante y el que moldea o marca la pauta a otros sistemas de educación superior en varios sentidos, entre ellos la evaluación docente. Por ejemplo, véase Altbach, P.G. y Selvaratnam, V. (1997). *From dependence to autonomy. The development of Asian universities*. Boston: Dordrecht.

A mediados de la década de los noventa surgió una de las críticas más feroces en contra de ese sistema: Stone (1995) recoge y amplía las voces críticas de algunas investigaciones producidas en las décadas de los ochentas y noventas, para afirmar que existía inflación de las notas o calificaciones y que esto era determinado por la masificación de la matrícula, la tendencia de los estudiantes a elegir los cursos menos difíciles y el uso de los cuestionarios para evaluar la docencia. Otro artículo que dio mayor sustento a las críticas surgidas en los años noventa y que resume las realizadas en décadas anteriores es el de Damron (1996). El autor critica la validez de los cuestionarios para evaluar el desempeño docente y hace una revisión sobre la influencia de otras variables en la evaluación. Posteriormente, Haskell (1997) en una serie de tres artículos publicados en la revista electrónica *Educational Policy Analysis Archives* lanza un fiero ataque al uso de los cuestionarios de evaluación por considerarlos inválidos y atentatorios contra la libertad académica. El autor narra algunos casos de profesores que demandaron a sus instituciones ante los tribunales por dar conocer a los estudiantes los resultados de las evaluaciones, con el fin de que ellos eligieran cursos. Las críticas han crecido en tono e intensidad durante la presente década. Dos pequeños artículos son citados con mayor frecuencia en la literatura reciente. El primero es el de Ruskai (1997), profesora de matemáticas, en un artículo que “Las evaluaciones de la docencia por los estudiantes tienen que ser investigadas con más cuidado, sus deficiencias, limitaciones”. La autora agrega que “muchos profesores experimentados cuestionan la confiabilidad de las evaluaciones de los alumnos para medir la efectividad docente y les preocupa que puedan tener efectos contraproducentes tales como la inflación de las notas, desaliento de la innovación e inhibir a los profesores a imponer retos difíciles a los estudiantes. Estos problemas son de importancia para profesores de todas las disciplinas, pero debido a la familiaridad de los científicos matemáticos con la estadística les corresponde mayor responsabilidad; que los matemáticos deberían de insistir en que cualquier componente numérico del proceso de

evaluación deben tener estándares mínimos de validez estadística” (1997: 1)

Como podrá observarse, la crítica de la autora no es incendiaria, pero señala de manera concisa y clara las deficiencias, limitaciones y efectos de la evaluación docente. Las críticas a la evaluación docente se hicieron más frecuentes a principios de esta década. Sproule (2000) señala que el propósito original de la evaluación docente fue proveer de retroalimentación a los profesores acerca de su efectividad docente, pero después se les dio un uso sumativo para decidir la recontractación, pago por mérito, definitividad en el empleo y promoción de los profesores. El autor señala las deficiencias conceptuales y estadísticas en la determinación de la efectividad docente por medio de estos instrumentos.

En la década de los noventa se dio en los países anglosajones un descenso notable en la investigación empírica sobre la evaluación docente. En la presente década continuaron los análisis críticos y una recuperación de la investigación empírica sobre el uso de los cuestionarios de evaluación. Entre esas voces críticas destaca otro artículo publicado en la American Association of University Professors, una de las primeras asociaciones de profesores en el mundo. Gray y Bergmann (2003) en sólo tres páginas hacen una de las críticas más demoledoras y difíciles de rebatir al sistema de evaluación docente que impera en los Estados Unidos y proponen otras alternativas para realizarla. Las autoras señalan:

En los 70, las escuelas empezaron a requerir a los profesores que los estudiantes llenaran formas de evaluación y que las turnaran a la administración universitaria. Los administradores pronto descubrieron que tenían un arma para ser empleada en contra del 50 por ciento de los profesores; podían proclamar que la mitad de los profesores debajo de la media de cada departamento eran malos profesores (Gray y Bergmann, 2003: 1).

Las autoras agregan:

Los estudiantes dan malas evaluaciones a aquellos profesores cuyo acento difiere del de los estudiantes,

a aquellos que enseñan cursos temidos o considerados de poco valor, tales como estadística para psicólogos (2003: 2).

Depender en exceso de la valoración de los estudiantes también impide innovación en la disciplina y metodología. Un profesor sin nombramiento definitivo no se puede arriesgar a intentar nuevas maneras de enseñar que pueden mejorar el desempeño de los estudiantes si el profesor sabe que el viejo método producirá calificaciones arriba de la media.

Las autoras nos ofrecen un análisis que pocas veces se hace. Los cursos poco valorados o temidos causan evaluaciones desfavorables. La tiranía de los cuestionarios de evaluación no sólo atenta contra la libertad académica sino también impide la innovación en la docencia. ¿Por qué arriesgarse con métodos nuevos si los viejos producen buenas evaluaciones?

Los fines de la evaluación y la nueva generación de estudios de evaluación de la docencia

La presente década marcó un hito en la generación de estudios, no sólo por las críticas Gray y Bergmann (2003) y otros autores, sino también gracias a una nueva generación de estudios empíricos. En el siglo pasado, la investigación sobre la evaluación docente en los Estados Unidos fue llevada a cabo fundamentalmente por psicólogos. En esta década los investigadores más productivos sobre el tema de la evaluación docente provienen de las escuelas de economía, contaduría, administración y mercadotecnia. La mayor parte de la investigación no ha sido para defender la utilidad y validez de los cuestionarios de evaluación sino para criticarlos. Una de las preguntas cruciales que hacen sus detractores (Emery, Kramer y Tian, 2001) es si los estudiantes son clientes o productos de la enseñanza universitaria. Es un dilema que no tiene respuesta clara. Emery y colaboradores señalan que las instituciones preparan a los estudiantes para desempeñarse en el sector productivo; para ellos

los estudiantes son un producto. Sin embargo, al interior de las instituciones que emplean los cuestionarios de evaluación, no sólo las estadounidenses, los estudiantes son vistos como clientes. El verlos como clientes afecta, naturalmente, el proceso de evaluación, los autores citados señalan que ese proceso no sólo ha concedido demasiado poder a los estudiantes para decidir sobre aspectos que no siempre están cualificados para evaluar: la dependencia de los estudiantes para evaluar a los profesores y tomarlo como el juicio con un peso único o importante, puede conducir a una disminución de la calidad de la enseñanza. Como se ha visto, desde esta perspectiva, no es extraño encontrar que el profesor se preocupe más por complacer a los estudiantes que por enseñarlos de la manera que a su juicio resulte más conveniente. La respuesta al dilema descansa en las instituciones y sus profesores, quienes, hasta ahora, no han protestado como grupo.

Ninguna otra profesión tan antigua y establecida como la academia se ha visto sujeta a una evaluación constante de su desempeño profesional de manera tan burocrática y poco útil como la que se realiza a través de los estudiantes. El trabajo de los arquitectos, médicos, ingenieros, abogados u otras profesiones similares, es materia de certificación por sus pares, la cual puede ser llevada a cabo cada dos o tres años o más, pero en esas profesiones no se evalúa en sus miembros cada producto o servicio que realizan. Con la imposición de la evaluación docente, la certificación académica no la dan los pares sino los clientes (estudiantes); ellos pueden decidir si el profesor es suficientemente competente para recibir estímulos o ser considerado buen profesor. El mensaje tácito que está enviando la administración universitaria y los diseñadores de las políticas de la evaluación docente podría ser el siguiente.

Estimad@ alumn@ de primer ingreso:

Bienvenido a tu universidad. Nuestro propósito es hacer tu estancia lo más productiva y placentera posible. Hemos puesto mucho cuidado en selec-

cionar a los profesores más capaces y responsables; casi todos nuestros profesores cuentan con Perfil PROMEP y varios de ellos pertenecen al Sistema Nacional de Investigadores. Sin embargo, no estamos completamente seguros de su desempeño docente; si somos sinceros contigo, no confiamos mucho en ellos. Por tal motivo, a partir de este momento y hasta que termines tus estudios te suplicamos que evalúes el desempeño de tus profesores en cada curso que tomes. Tu opinión será muy tomada en cuenta para confirmar si en realidad tenemos una planta docente de calidad. Estamos conscientes de que no eres el mejor calificado para evaluar el desempeño de tus docentes, pero te queremos dar el gusto y trabajo de hacerlo con el fin de que, cuando tus expectativas, independientemente de que sean realistas o alcanzables, no se cumplan, le pongas al profesor la calificación que creas que se merezca. No te prometemos tomar tu calificación y recomendaciones en cuenta, pero tus evaluaciones nos servirán para presionar al profesor a hacer su trabajo como el que tú esperas y a nosotros para tener el control de los incentivos que le podemos otorgar. Entre más te complazca, aunque esa complacencia no te lleve a aprender mucho, más valoraremos al profesor. Será el único criterio que tomaremos en cuenta para calificarlo como un buen o mal docente. Gracias por tu participación. Atentamente

Tu universidad, con el patrocinio de SHCP, SEP, PROMEP, CIEES.

La certificación continua del trabajo del profesor que se ha impuesto en las universidades pervierte la relación profesor-alumno, ya que la labor del docente no debe ajustarse a lo que su experiencia y *expertise* le señalan sino que tiene que ajustarse a los criterios de satisfacción del usuario. En esta dualidad radica parte del problema. Los alumnos, en general, buscan un curso fácil de aprobar. Como un profesor americano (Damron, 1996) solía expresar: un aprendizaje sin dolor (*painless education*). El modelo docente universitario actual pasó de un extremo a otro. Del viejo autoritarismo en el que el profesor era quien tenía

la primera y última palabra, que podía hacer a cualquier estudiante el más desgraciado de la institución, hasta el otro extremo en el que se encuentra la situación que prevalece en la mayoría de las instituciones de educación superior: los estudiantes y la administración universitarias, no los pares ni la sociedad, son reconocidos como los únicos y más aptos para juzgar las competencias docentes de los profesores, sea de aquellos que acaben de ingresar a la institución o que otros que hayan sido distinguidos con las más altas preseas académicas. Son los dos extremos del autoritarismo; uno del docente y el otro de los estudiantes y la administración universitaria. Los profesores, como lo podría atestiguar cualquier académico, hemos sido víctimas y verdugos de esa educación autoritaria. Varios de nosotros hemos recibido un tratamiento injusto y autoritario, a veces con repercusiones de por vida, por parte de alguno de nuestros antiguos profesores. Asimismo, no es extraño que ya en nuestras funciones como docentes, voluntaria o involuntariamente, hayamos infligido algún tratamiento injusto a algún estudiante. Parte del problema es de tipo estructural, el enseñante, en su calidad de poseedor del conocimiento, ostenta un poder mayor que el aprendiz. Por otra parte, la docencia no sólo es una técnica, es también un arte y como tal, la parte emocional-personal juega un papel tan importante como la científica-técnica. En esta dualidad radica parte del problema.

En esa complejidad de la docencia que involucra el conocimiento, el método de enseñanza, la cultura y subculturas académicas de la disciplina(s) y la personalidad del profesor y del estudiante, se encuentra la dificultad de evaluar la docencia. Cada disciplina tiene una identidad propia y, en consecuencia, un estilo docente peculiar. Incluso, si este problema se pudiera superar por medio de una enseñanza uniforme en todas las disciplinas, faltaría uniformar el estilo docente de cientos de miles de profesores. Todos sabemos que no hay un profesor que enseñe exactamente de la misma manera que otro. Dado el trabajo artesanal que se hace en la docencia, el método es el maestro, hay tantas variantes y estilos docentes como maestros.

Posiblemente por estas razones es difícil evaluar la docencia. El hecho de uniformar la evaluación para todas las disciplinas, cursos y profesores, porque de otra manera es muy costoso, trae consigo una evaluación anodina del quehacer docente.

Los cuestionarios son instrumentos limitados y poco válidos para evaluar la docencia

Dada la dificultad que implica medir o apreciar los efectos de la docencia, lo que mejor pueden hacer los cuestionarios es dar una apreciación del índice de satisfacción del estudiante; para ello se ha optado evaluar, fundamentalmente, el proceso docente y pocas veces los resultados. Las limitaciones sobresalientes de estos instrumentos se relacionan con su validez, medir lo que pretenden medir, y las variables intervinientes en los resultados de la evaluación. Uno de los problemas de fondo de los cuestionarios de evaluación, es su indeterminación. Desde el punto de vista filosófico, la indeterminación se refiere a un modelo dado que no provee de una explicación unívoca de un cuerpo de datos (Sproule, 2002). Sproule señala que los cuestionarios de evaluación padecen de indeterminación porque sirven a diferentes propósitos: primero, proporcionar retroalimentación de los estudiantes a los comités de evaluación de profesores; el segundo es que en el proceso de evaluar se les pide a los estudiantes dar su apreciación del curso y de la clase a través de reactivos generales; el tercero tiene que ver con los estándares mínimos que la universidad fija arbitrariamente. Como señalan Gray y Bergmann (2003), si se usa un promedio, global de la institución o departamental, ¿ello significa que ese 50% de los profesores es incompetente? Parece que sí, a juicio de la administración universitaria.

A pesar de que la investigación realizada décadas atrás había señalado que los cuestionarios son medidas válidas y confiables para medir el desempeño docente (García, 2000b), y que los sesgos del instrumento no afectaban su validez global (García, 2000a), la investigación reciente

pone en serias dudas tal aseveración. Aquellos que defendimos la validez de los cuestionarios nos dejamos guiar por el grueso de la investigación internacional y la propia que demostraba la estabilidad en los juicios de los estudiantes acerca del desempeño de sus profesores. En una investigación realizada en México con 1391 profesores de tiempo completo y asignatura que impartieron clase a 8745 grupos durante seis semestres escolares, se comprobó que estos instrumentos son consistentes a lo largo del tiempo (García, 2003). Marsh (2007), en una de sus más recientes investigaciones con una muestra de 195 profesores concluye que a lo largo de 13 años y 6024 cursos no existe variación significativa en el desempeño docente en la muestra estudiada. El problema no es del instrumento, cuando se emplea con fines formativos; los problemas de validez ocurren cuando los cuestionarios se emplean con fines diferentes. De acuerdo con Onwuegbuzie y colaboradores (2007), los más graves problemas son su validez de criterio, contenido y de constructo. La validez de criterio se refiere a la relación que tiene otro criterio o medida. Los autores citan la investigación clásica de Cohen (1981) que reportó una relación de .43 entre los resultados de los cuestionarios y el desempeño académico. Los autores citados señalan que los estudiantes que alcanzan el más alto nivel de desempeño tienden a dar el crédito al docente independientemente si dicho crédito es justificable, no siempre es el caso. Onwuegbuzie y colaboradores (2007) señalan que si bien los cuestionarios pueden tener validez aparente (o de *facie*) en la medida en que sus *ítems* pueden parecer a quien responde relevantes o interesantes de contestar, no existen suficiente evidencias de que los *ítems* de los cuestionarios tengan suficiente validez de contenido, ya que no tienen definido con suficiente claridad el concepto de efectividad docente.

Si uno revisa cualquier cuestionario de evaluación encontrará que tienen un gran parecido; un cuestionario de evaluación de la docencia por los estudiantes contiene reactivos que tienen que ver con las siguientes áreas: la organización y

planeación de la clase, el dominio del conocimiento, la comunicación con los estudiantes y su disponibilidad para atenderlos, y la manera de evaluar el aprendizaje. En México, poco se ha hecho énfasis en los resultados del aprendizaje y en la inclusión de reactivos que evalúen globalmente el curso y al profesor. Es común que este tipo de instrumentos incluyan preguntas insulsas como: si el profesor usó recursos tecnológicos, si llegó temprano, si recomendó usar la biblioteca, si recomendaría el curso a otros estudiantes o comparar al profesor con otros profesores del programa o la institución (Scriven, 1995). Tales atributos que se quieren medir no tienen relación directa, al menos no se ha probado, con que la eficacia docente. Aunque existen esfuerzos de algunos investigadores que propugnan la medición de los resultados del aprendizaje (Cashin, 1990), en México no es común encontrar este tipo de reactivos; desde luego incluirlas mejoraría en algo la validez, pero su inclusión no garantizaría que esas formas de evaluación la incrementaran sustancialmente.

Otros factores que atentan contra la validez del instrumento son las expectativas de los estudiantes. La investigación reciente ha comprobado consistentemente que existen diferencias entre las expectativas iniciales acerca del curso y su percepción final (Comm y Mathaisel, 2006). Las expectativas de obtener una alta calificación o nota final en el curso están inversamente relacionadas con la calificación que los estudiantes otorgan a los profesores en los cuestionarios como lo han demostrado diversos estudios recientes. Jhonson (2002) señala que la probabilidad de que un estudiante le otorgue al profesor un alto puntaje en el cuestionario se incrementa por tres si en lugar de darle un 80% le da un 100% de calificación. Clayson, Frost y Sheffet (2006) señalan que es una prueba de efecto de correspondencia recíproca: si el profesor da altas notas, los estudiantes también. Esta relación perversa no sólo provoca humillaciones injustas a los profesores como lo señalan acertadamente Gray y Bergmann (2003) sino también una inflación de las notas o calificaciones. Un estudio realizado en las universidades de la provincia de Ontario,

Canadá señala que con excepción de matemáticas, las calificaciones aumentaron significativamente entre 1973-74 y 1993-94. Los cursos de inglés y biología tuvieron las mayores ganancias). Estudios como el citado son abundantes en la literatura anglosajona. Sería interesante realizar estudios similares en México, a partir de la aplicación de los cuestionarios de evaluación.

No sólo las expectativas de los estudiantes atentan contra su validez, también el *locus* de control, constructo creado por Rotter (Morris, 1997). Existe un *locus* externo, el cual determina que atribuyamos los sucesos personales fundamentalmente a eventos o personas externas; por el contrario el *locus* interno atribuye los sucesos fundamentalmente a la responsabilidad interna del individuo, él o ella se sienten capaces de controlar su propio destino. La investigación realizada por Grimes y colaboradores (2004), encontró que los estudiantes que tienen un *locus* de control externo otorgan más bajas puntuaciones a sus profesores en los cuestionarios que aquellos que tienen *locus* de control interno. Por otra parte, el anonimato también afecta los resultados en las evaluaciones. El estudio de Blunt (1991) señala que las respuestas anónimas de los estudiantes son significativamente más altas en: interés por la materia, actitud del docente y su sentido del humor. No parece extraño, los estudiantes cuando saben que el cuestionario no es anónimo temen expresar juicios severos en contra de los profesores, por las posibles represalias que pueden tomar.

A manera de conclusión

Los sistemas de evaluación de la docencia implican una fuerte carga administrativa para las instituciones de educación superior. Su carga, costo y escasa validez para evaluar el desempeño docente de decenas de miles de profesores, independientemente de su estatus y experiencia, y lo poco que aportan al mejoramiento de la calidad, hacen injustificable a todas luces la continuidad de la política vigente de evaluación de la docencia. Es oneroso procesar evaluaciones de miles de cursos realizadas por dece-

nas de miles de alumnos-materia en cada periodo escolar sin que tengan un fruto específico, más allá del control y otorgamiento de incentivos. Por otro lado, los profesores se ven sometidos a una especie de certificación continua de sus habilidades docentes, algo inusitado en otras profesiones, y los alumnos obligados a responder tantos cuestionarios comocursos tomados durante sus estudios. No es conveniente ni para los profesores que se ven sometidos a un escrutinio constante e injustificado de su actividad docente ni para los estudiantes que toman el trabajo como una rutina aburrida y poco útil.

¿Qué se podría hacer para mejorar el sistema de evaluación actual?

A pesar de las evidencias internacionales y nacionales que señalan la poca utilidad de la evaluación docente, la evaluación llegó para quedarse, debido a que está ligada a los programas de incentivos económicos y al control administrativo del trabajo docente. Sin embargo, el sistema se podría mejorar si se partiera de una premisa fundamental: servir a un propósito formativo, de mejora y no sumativo de control y asignación de estímulos. Naturalmente, la administración universitaria requiere de algún tipo de control sobre las tareas del docente y su realización; puede, y está en su derecho de, tener un control administrativo, si el profesor asiste a impartir clase y si tiene las credenciales para ellos, pero el control o rendición de cuentas del desempeño académico, como ocurre en otras funciones académicas, debe descansar en los pares, no en la administración, ni mucho menos en los estudiantes. Si se dejara de ligar el resultado de los cuestionarios al programa de estímulos, los resultados de la evaluación podrían servir de alguna manera como medida de retroalimentación de la labor docente. Desde esta consideración se podrían implementar las siguientes variantes:

- 1) Evaluar no más de un curso al azar cada semestre de los que impartan los profesores noveles o de reciente ingreso a la institución (Hoyt y Pallet, 1999).
- 2) Los profesores titulares serán sometidos a evaluación de un curso, seleccionado al azar, cada dos o tres años (Hoyt y Pallet, 1999).
- 3) Cada semestre, a mediados del periodo escolar, al menos en uno de los cursos que impartan, los profesores harán preguntas a los alumnos, como las que sugieren Bastick (2002); el método lo denomina el autor: *alineación de las habilidades, competencia profesional y actitudes*. Consiste en tomar la opinión de los estudiantes a mitad del curso. El instrumento consta de dos columnas; en la primera se pregunta qué tanto (expresado en porcentaje) el curso está poniendo énfasis en 1) habilidades, 2) entendimiento y 3) actitudes; se agrega una pregunta final para indagar qué tanto el alumno está disfrutando el curso y qué tanto esperaba del mismo cuando inició. En la segunda columna se le pide al estudiante que exprese qué tanto debería poner énfasis y qué tanto se ha cumplido de lo que esperaba del curso, en las áreas de habilidades, entendimiento y actitudes. La diferencia entre ambas columnas le permite al profesor obtener información para retroalimentar su labor. Estos resultados se pueden discutir con grupos de colegas o colegios académicos.
- 4) Otra opción es videogravar a los profesores que se presten para ello o bien que se decida en el colegio académico. Los profesores videogravados pueden recibir retroalimentación de sus colegas. Ya es hora de sacar de la secrecía la manera en que enseñamos; no se trata de artes de alquimia o prácticas de sociedad secretas sino de una actividad profesional que tiene que realizarse, como lo hacen en otras profesiones: a los ojos de otros colegas. Se ha comprobado que la videogravación es un manera poderosa de retroalimentación de la docencia en la educación superior (Penny y Coe, 2004).

Bibliografía

- Altbach, P. G. y Selvaratnam, V. (1997). *From dependence to autonomy. The development of Asian universities*. Boston: Dordrecht.
- Arbesú, Ma. Isabel, Canales, Alejandro, Crispín, María Luisa, Cruz, Isabel, Figueroa, Alma, y Gilio, Ma. del Carmen (2004). "Las políticas y usos de la evaluación de la docencia en la educación superior: planteamientos y perspectivas". En: Rueda, M. y Díaz Barriga, F. (Coordinadores). *La evaluación de la docencia en la universidad. Perspectivas desde la investigación y la intervención profesional*. México: CESU-UNAM y Plaza y Valdés Editores, p.203-257.
- Bastick, Tony (2002). *In-course optimization of teaching quality*. Trabajo presentado en la Reunión Anual de la Southwest Educational Association (SERA), Austin, Texas
- Blunt, Adrian (1991). "The effects of anonymity and manipulated grades on student ratings of instructors". *Community College Review*, Vol. 18, (4), p. 48-54. (Ciudad?)
- Cashin, William E. (1990). *Student ratings of teaching: Recommendations for use*. Kansas State University, Center for Faculty Evaluation and Development. Disponible en: http://www.idea.ksu.edu/papers/Idea_Paper_22.pdf
- Centra, J. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Clayson, Dennis, Frost, Taggart y Sheffet, Mary (2006). *Academy of Management Learning*, Vol. 5 (1), p 52-65.
- Cohen, P. A. (1981). "Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies". *Review of Educational Research*, 51, (3) p. 448-458.
- Comm, Clare y Mathaisel, Dennis (2006). "Employing 'gap análisis' to measure classroom outcomes". *Journal of Marketing Education*, Vol. 28, (3), p. 254-264.
- Damron, J.C. (1996). *Instructor personality and the politics of the classroom*. Disponible en: www.mankato.msus.edu/dept/psych/Damron_politics.html
- Emery, C. Kramer, T. y Tian, R. (2001). "Customers vs. products: Adopting an effective approach to business students". *Quality Assurance in Education*, Vol. 9 No. 2, p. 110-115.
- Etzioni, Amitai (1961). *Complex organizations*. Nueva York: The Free Press.
- Feldman, K. A (1990). "An afterword for "The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisectional validity studies". *Research in Higher Education*, 30, p. 583-645.
- García Garduño, José María (2000a) "¿Qué factores extra clase afectan la efectividad docente en la educación superior? Síntesis de la investigación producida sobre los sesgos en los cuestionarios de evaluación de la docencia". *Revista Mexicana de Investigación Educativa*, 5, p. 303-325. Disponible en: <http://www.comie.org.mx/v1/revista/portal.php?idm=es&sec=SC03&&sub=SBB&criterio=ART00289>
- García Garduño, José María (2000b). "Las dimensiones de la efectividad docente, validez y confiabilidad de los cuestionarios de evaluación de la docencia: síntesis de investigación internacional". En: Rueda, M. y Díaz-Barriga, F. (Compiladores), *Evaluación de la docencia*. México: Paidós, p. 41-62.
- García Garduño, José María (2003). "Profesores universitarios y su efectividad docente. Un estudio comparativo entre México y Estados Unidos". *Perfiles, Educativos*, 25 (100), p. 42-55. Disponible en: <http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=13210004&iCveNum=5898>
- Gray, Mary y Bergmann, Barbara (2003). "Student teaching evaluation: Inaccurate, demeaning and misused". *Academe Online*, September-October. Disponible en: <http://www.aaup.org/AAUP/pubs-res/academe/2003/SO/Feat/gray.htm>
- Grimes, Paul, Millea, Meghan J. y Woodruff, Thomas (2004). "Grades-who's to blame? Student evaluation of teaching and locus control". *Journal of Economic Education*, Vol. 35, (2), p. 129-147.
- Haskell, Robert (1997). "Academic freedom, tenure, and student evaluation of faculty: galloping polls in the 21st century". *Educational Policy Analysis Archives*, Vol. 5, No. 6. Disponible en : <http://epaa.asu.edu/epaa/v5n6.html>

- Hoyt, Donald, P. y Pallett (1999). *Appraising teaching effectiveness: beyond student ratings*. IDEA Paper # 36. IDEA Center. Disponible en: http://www.idea.ksu.edu/papers/ideapaper36_appendixD.pdf
- Ibarra, Eduardo y Porter, Luis (2007). "El debate sobre la evaluación: del homo academicus al homo economicus". *Reencuentro* No. 48, abril, p. 35. Disponible en: http://148.206.107.10/biblioteca_digital/estadistica.php?id_host=2&tipo=ARTICULO&id=901&archivo=367901pcr.pdf&titulo=El%20debate%20sobre%20la%20evaluaci%F3n:%20del%20homo%20academicus%20al%20homo%20economicus
- Johnson, Valen (2002). "Teacher course evaluations and student grades: An academic tango". *Chance*, 15, (3). Disponible en: www.amstat.org/publications/chance/153.johnson.pdf
- Lazear, Paul (2003). "Teacher incentives". *Swedish Economic Policy Review* 10, 179-214. Disponible en: <http://www.sweden.gov.se/content/1/c6/09/54/30/5cf67673.pdf>
- Marsh, H. (1987). "Student evaluations of university teaching. Research findings, methodological issues, and directions for future research". [número completo] en: *International Journal of Higher Education*, 1, (3). Número completo.
- Marsh, Herbert (2007). "Do university teachers become more effective with experience? A multilevel growth model of student's evaluation of teaching over 13 years". *Journal of Educational Psychology*, Vol. 99 (4), p. 775-790.
- McKeachie, Wilbert (1979). Student ratings of faculty: A reprise. *Academe*, 65, p. 384-397.
- McKeachie, Wilbert, J. (1996). "Students ratings of teaching". En: England, J., Hutchings, P. y McKeachie, W., *The profesional evaluation of teaching*. American Council of Learned Societies, Ocassional Paper No 33. Disponible en: www.acls.org/op33.htm
- Morris, Charles (1997). *Psicología*, 9a ed. México: Pearson.
- Onwuegbuzie, A. Witcher, A. Collins, K. M.T., Filer, J., Wiedmaier, C. y Moore, C. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-method analysis. *American Educational Research Journal*, 44 (1), p. 113-160.
- Penny, Angela y Coe, Robert (2004). "Effectiveness of consultation on student ratings feedback. A meta-analysis". *Review of Educational Research*, Vol. 74, (2), p. 215-253.
- Ruskai, Mary B. (1997). Evaluating student evaluations. *Notices*, Vol., 44. (3). Disponible en: <http://www.ams.org/notices/199703/page2.html>
- Ryan, R. y Deci, E. (2000). "Intrinsic and extrinsic motivations: Classic definitions and new directions". *Contemporary Educational Psychology*, 25, p. 54-67.
- Scriven, Michael (1995). "Student ratings offer useful input to teacher evaluations". *Practical Assessment & Evaluation*, 4 (7). Disponible en: <http://pareonlines.net/getvn.asp?v=4&n=7>
- Sproule, Robert (2000). "Student evaluations of teaching: A methodological critique of conventional practices". *Educational Policy Analysis Archives*, Vol. 8, No 50. Disponible en <http://epaaa.asu.edu/v8n50.html>
- Sproule, Robert (2002). "The underdetermination of instructor performance by data from the student evaluation of teaching". *Economics of Education Review*, 21, p. 287-294.
- Stone, J. E. (1995). "Inflated grades, inflated enrollment and inflated budgets". *Educational and Policy Analysis Archives*, Vol. 11, (11). Disponible en: <http://epaa.asu.edu/epaa/v3n11.html>



