



Para citar este artículo, le recomendamos el siguiente formato:

Luna, E. y Torquemada, A. (2008). Los cuestionarios de evaluación de la docencia por los alumnos: balance y perspectivas de su agenda. *Revista Electrónica de Investigación Educativa, Especial*. Consultado el día de mes de año, en: <http://redie.uabc.mx/NumEsp1/contenido-lunatorquemada.html>

---

## Revista Electrónica de Investigación Educativa

Número Especial, 2008

### Los cuestionarios de evaluación de la docencia por los alumnos: balance y perspectivas de su agenda<sup>1</sup>

### Students' Evaluations of University Teaching: Balance and Perspectives of their Agenda<sup>2</sup>

Edna Luna Serrano (1)  
[eluna@uabc.mx](mailto:eluna@uabc.mx)

Alma Delia Torquemada (2)  
[almatorquemada@yahoo.com.mx](mailto:almatorquemada@yahoo.com.mx)

1 Instituto de Investigación y Desarrollo Educativo  
Universidad Autónoma de Baja California

Carretera Tijuana-Ensenada km 103, 22890  
Ensenada, Baja California, México

2 Centro de Investigaciones en Ciencias Sociales  
Universidad Autónoma del Estado de Hidalgo  
Carretera Pachuca-Actopan km 4, 42160  
Pachuca de Soto, Hidalgo, México

(Recibido: 13 de agosto de 2008; aceptado para su publicación: 10 de septiembre de 2008)

## **Resumen**

Este trabajo tiene el propósito de examinar, de manera crítica, los hallazgos de la literatura en relación con la aplicación y uso de los cuestionarios de evaluación de los estudiantes sobre la efectividad de la enseñanza, con el fin de orientar su utilización en el marco de un sistema que posibilite la mejora de la docencia universitaria. Consta de seis apartados. El contexto de la discusión; evaluación de la eficacia docente; aspectos metodológicos: confiabilidad y validez; el uso de los cuestionarios y mejora de la eficacia docente. Concluye a favor de utilizar los cuestionarios como una fuente de información del sistema con requisitos de uso y administración derivados de la investigación y experiencia práctica, y de establecer el vínculo entre los resultados de la evaluación docente y el diseño de modalidades de formación docente.

*Palabras clave:* Evaluación del docente; método de evaluación; personal académico docente.

## **Abstract**

This work has the purpose of critically evaluate the findings of literature in relation to the administration and use of the teacher's evaluation instruments in which the students can qualify the efficiency of teaching. The discussion may help orient the use of this kind of instruments within the frame of an evaluation system to improve teaching at college level. The work includes six sections: the discussion context; evaluation of teacher's efficacy; methodological aspects, i.e, reliability and validity; the use of instruments and the improvement of teacher's efficacy. It is concluded that evaluative instruments are useful as a system's information source with use and administration requirements derived from research findings as well as empirical data. Also, establishing the link between evaluation results and the design of new modalities of teacher's formation is discussed.

*Key words:* Teaching evaluation, evaluation methods, college faculty.

## **Introducción**

Un acuerdo generalizado en la universidad, es que los docentes tienen un papel determinante en la mejora de la calidad de la enseñanza. En el mismo sentido, se reconoce la necesidad de desarrollar sistemas de evaluación del desempeño del profesor que contribuyan a su profesionalización y, con ello, a la mejora de la formación profesional.

En México como en otros países, la evaluación de la docencia se realiza en respuesta a las demandas sociales hechas por diferentes audiencias con necesidades heterogéneas de evaluación. Hasta finales de la década de los ochenta, la evaluación docente se condujo principalmente por los requerimientos institucionales de información sobre la calidad de la enseñanza y para retroalimentar las estrategias de formación docente (Arias, 1984; Luna, 2002). A partir de 1990, con la implementación generalizada de los programas de compensación salarial, la evaluación de la docencia se sistematizó como uno de

los indicadores integrados a estos programas, con lo que se atendió principalmente los requerimientos de control administrativo (Canales y Gilio, en prensa). En la actualidad, es factible ubicar diversas peticiones: las de docentes y estudiantes sobre sistemas justos y pertinentes vinculados a la mejora de la enseñanza; las de autoridades para la toma de decisiones administrativas, asignación de cursos, promociones y estímulos económicos; y las gubernamentales de rendición de cuentas sobre la calidad de la docencia (Luna, 2004; SEP, 2007).

La evaluación docente con base en los puntajes de opinión de los estudiantes, es la estrategia de evaluación más utilizada en Norteamérica, Europa y Asia, y al mismo tiempo es la más estudiada (Theall y Franklin, 2000; Seldin, 1993). No obstante, en la actualidad se acumulan evidencias sobre el mal uso de los puntajes de los estudiantes, lo que da cuenta del déficit en las buenas prácticas y la persistente disparidad entre los hallazgos de la investigación y los procedimientos implementados en las instituciones. En este contexto de utilización generalizada de los cuestionarios, se presentan varias situaciones paradójicas: su aplicación no atiende los aspectos básicos que la literatura señala como fundamentales para validar su empleo; los docentes manifiestan una creciente incomodidad por su uso, en la toma de decisiones administrativas; y se sobreestima o minimiza su utilidad.

Este documento tiene el propósito de examinar de manera crítica, los hallazgos en la literatura relacionados con la aplicación y uso de los cuestionarios de evaluación de los estudiantes sobre la efectividad de la enseñanza, para orientar su utilización en el marco de un sistema que posibilite la mejora de la docencia universitaria.

## **I. El contexto de la discusión**

En el marco de la evaluación educativa, se ubica la discusión sobre los cuestionarios de evaluación de la docencia por los estudiantes; los principios generales que guían su diseño, aplicación y uso son los de este ámbito del conocimiento, en particular los relativos a la evaluación de la docencia. Un requisito fundamental para los procesos de evaluación de la docencia, es que los resultados sean considerados dignos de crédito por los docentes, los estudiantes, los administradores y por la comunidad en general. De acuerdo con House (1998), la credibilidad de las conclusiones de la evaluación se basa en la utilización de una metodología científica que asegure la confiabilidad y validez de los procesos y la independencia política y el profesionalismo de los evaluadores. La evaluación requiere ser considerada también equitativa: los involucrados (docentes, estudiantes y administradores) esperan percibir que los programas de evaluación reflejan sus intereses y puntos de vista (Good, 1996). Asimismo, se demanda que los sistemas de evaluación de la docencia sean congruentes con la filosofía institucional para apoyar el logro de su misión (McLaughlin y Pfeifer, 1988).

Actualmente, la evaluación de la docencia con base en los cuestionarios de opinión de los estudiantes, es reconocida como la estrategia de evaluación con mayor conocimiento conceptual y empírico acumulado en el contexto universitario (McKeachie, 1996). Los principales propósitos asociados a su utilización se relacionan con medidas de la efectividad de la enseñanza en la toma de decisiones administrativas, con el diagnóstico y retroalimentación de los maestros para mejorar el proceso de instrucción, y con la investigación sobre la enseñanza (Marsh y Dunkin, 1997). Por ello, se considera que los puntajes son útiles para los docentes, los estudiantes y los administradores.

### **Evaluación de la eficacia docente**

La búsqueda de respuestas sobre la enseñanza efectiva ha generado la existencia de tres perspectivas de investigación: la que se concentra en los productos de la enseñanza; la relacionada con los procesos que realiza el profesor; y la que establece la relación entre procesos y productos, a fin de indagar los vínculos entre lo que los profesores hacen y cómo cambian los estudiantes (Abrami, d'Apollonia y Rosenfield, 1997). En general, los cuestionarios de evaluación de la docencia por los estudiantes se alinean a una de las tres aproximaciones.

Los puntajes son una medida directa de un producto de la actividad docente, denominada satisfacción de los estudiantes con la enseñanza. Para muchos, medir la satisfacción de los estudiantes respecto a la docencia es razón suficiente para utilizar los cuestionarios. Por otro lado, los puntajes de los estudiantes no ofrecen una medida directa de productos valiosos de la enseñanza como pueden ser los aprendizajes, expectativas, creencias, actitudes, valores e interés hacia el objeto de estudio de los estudiantes. De manera general, los cuestionarios se utilizan como una medida indirecta de la mayoría de los productos de la enseñanza, porque proveen un criterio básico en los juicios que infieren la medida en que el docente afecta la enseñanza. Lo anterior no justifica las malas prácticas caracterizadas por utilizar los cuestionarios como instrumentos para cubrir todos los ángulos de una actividad tan compleja como la docencia (Rueda, Elizalde y Torquemada, 2003).

Los cuestionarios pioneros se diseñaron con base en los componentes que los expertos consideraron los más importantes para la enseñanza, susceptibles a la observación y juicio de los estudiantes. De esta manera, se elaboró el cuestionario tradicional de Remmers y otros instrumentos (Centra, 1993). En la actualidad, la teoría clásica sobre el desarrollo de cuestionarios de evaluación por los alumnos propone combinar tres estrategias:

- Deducir los reactivos de un análisis lógico de los componentes de la efectividad docente, tomando en cuenta la retroalimentación de los profesores y estudiantes, así como los propósitos de la evaluación.

- Realizar análisis empíricos mediante el empleo de técnicas estadísticas como análisis factorial, diseños multirrasgo-multimétodo y modelamiento de ecuaciones estructurales.
- Fundamentar los instrumentos de evaluación en una teoría de la enseñanza y aprendizaje (Marsh y Dunkin, 1997).

Se reconoce que la mayoría de los cuestionarios desarrollados y publicados, atendieron los primeros dos aspectos e ignoraron elementos centrales de la teoría del aprendizaje que les subyace (Luna, 2000).

En los cuestionarios, la enseñanza se asume como una actividad compleja y multidimensional; por lo general, los agrupamientos de reactivos proceden de los análisis factoriales a los que fueron sometidos los instrumentos, y reflejan las diferentes dimensiones de la efectividad de la enseñanza, a juicio de los estudiantes. Por ello, a pesar de que no existe consenso sobre las características de un docente de calidad, el estudio de las dimensiones relacionadas con la efectividad docente, las identifica y clasifica de acuerdo con su importancia para el aprendizaje. Así, se ubican como las más importantes: claridad y entendimiento; estimulación del interés por la materia; preparación y organización del curso; logro de los objetivos planteados; y motivación a los estudiantes para tener alto rendimiento (Feldman, 1997). El supuesto comúnmente aceptado es que estas dimensiones y los reactivos que las incluyen, forman parte de las habilidades genéricas que todo docente debe presentar.

Quienes critican el tipo de enseñanza que la mayoría de los cuestionarios promueven de manera implícita, por lo general aducen dos argumentos. Por un lado, que éstos fomentan una pedagogía conservadora sustentada en el modelo de transmisión del conocimiento, ya que en la época en la que fueron desarrollados, las clases tradicionales eran la norma. Por otro lado, la necesidad de incorporar los cambios contextuales propios de cada situación educativa, por ejemplo, las nuevas prácticas instruccionales derivadas de los planes de estudio por competencias, aprendizaje colaborativo y la educación a distancia (Centra, 1993; Theall y Franklin, 2000). Por ello, desde principios de los años noventa se habla de las limitaciones de incluir únicamente las dimensiones asociadas a la efectividad docente en los cuestionarios y de la necesidad de incorporar las particularidades del contexto pedagógico, como una manera de considerar y responder a la filosofía y misión de la institución.

Alejados de la tradición metodológica del campo, en México muchos de los cuestionarios actualmente en uso, carecen de rigor metodológico en su elaboración. Omiten los componentes validados por la literatura con las consecuentes desventajas, el alejamiento del propósito central de origen: identificar a los profesores que presentan los elementos asociados a la enseñanza efectiva y al aprendizaje, así como a quienes carecen de ellos. Además, los administradores agregan reactivos con un marcado interés de control administrativo o determinados por las exigencias de programas

especiales de compensación salarial. Así, un elemento crítico a considerar, es la calidad técnica del diseño del cuestionario que asegure su confiabilidad, validez y utilidad.

## **II. Aspectos metodológicos básicos: confiabilidad y validez**

La utilidad práctica y teórica de los cuestionarios, depende del cumplimiento de los estándares psicométricos de diseño y aplicación del instrumento. En las décadas de los ochenta y noventa, la investigación se orientó a estudiar la confiabilidad y validez de los cuestionarios como una medida de la efectividad docente; se demostró que los puntajes son confiables, estables y relativamente válidos en través de diferentes escenarios educativos (Marsh, 1987; Abrami, d'Apollonia y Cohen, 1990). Cabe hacer notar que no es un tema vigente en las publicaciones especializadas recientes.

Tradicionalmente, confiabilidad y validez son conceptos opuestos de un continuo de generalización. Así, confiabilidad refiere la generalización de los puntajes y representa el mismo constructo colectado bajo situaciones similares; validez significa la generalización de los puntajes con diferentes constructos que se hipotetiza están relacionados, por ejemplo, puntajes y aprendizaje de los estudiantes (Marsh y Dunkin, 1997). Los análisis empíricos tienen el propósito de determinar la confiabilidad de los resultados y la validez de las interpretaciones hechas sobre la base de las evaluaciones de los estudiantes.

### **Confiabilidad**

Las dos medidas de confiabilidad más utilizadas son la consistencia de la medida, considerada como el acuerdo relativo entre los estudiantes en una clase o curso, y la estabilidad de los puntajes a través del tiempo, tomada como el acuerdo entre los estudiantes en relación con un mismo instructor, con el mismo instrumento en diferentes momentos (Braskamp y Ory, 1994).

En el medio académico persiste una serie de cuestionamientos sobre la confiabilidad de las evaluaciones de los estudiantes referentes a la efectividad de la enseñanza. Sin embargo, muchos de ellos carecen de referentes de investigación que los sustenten. Entre los más comunes se encuentran los que debaten a los estudiantes como jueces fiables; aducen, por ejemplo, que los estudiantes no pueden hacer juicios sobre la enseñanza o el docente debido a su inmadurez, o que los estudiantes no son capaces de reconocer la enseñanza efectiva hasta después de haber demostrado dominio de la materia o haber egresado. En cuanto a los hallazgos de investigación, se muestra de manera consistente que la opinión de los estudiantes no se modifica pasados varios años, incluso después de haber egresado (Overall y Marsh 1980; Marsh, 2007). No obstante, la polémica ha derivado en la aceptación de que en algunos rubros de la competencia docente los estudiantes no son los mejores jueces, como es el caso de opinar sobre la pertinencia de la bibliografía, donde los colegas

cuentan con mayores elementos para emitir una valoración (Luna y Rueda, 2001).

Los estudios que investigan la correlación entre los puntajes y los efectos del docente en una asignatura en particular, tienen como fin conocer en qué medida los puntajes de un docente se correlacionan en dos diferentes cursos, o conocer la correlación entre los puntajes de una asignatura enseñada por diferentes profesores en dos ocasiones diferentes. La conclusión es que los puntajes reflejan principalmente la efectividad del instructor más que la influencia del curso, y que las evaluaciones de los estudiantes están en función del profesor que enseña y no del curso que se evalúa (Gilmore, Kane y Naccarato, 1978; Marsh, 1981). Estos resultados apoyan la confiabilidad de las evaluaciones de los estudiantes, y al mismo tiempo, fundamentan la sugerencia de que los puntajes dados a un docente deben ser promediados para incrementar la generalización. Por ello, las decisiones de tipo administrativo deben estar basadas en un número adecuado de cursos diferentes, por lo menos cinco series de dos cursos distintos (Marsh y Dunkin 1997).

### **Estabilidad**

Las dos aproximaciones más comunes en los estudios de estabilidad y cambios en las evaluaciones de los estudiantes revisan lo estable de los promedios de los puntajes a través del tiempo y la persistencia de las diferencias individuales. Uno de los trabajos más comprehensivos, es el de Marsh (2007), que estudió a 195 docentes evaluados de manera continua durante un periodo de 13 años (6,024 cursos, 30.9 asignaturas por docente). Encontró estabilidad de los puntajes en el tiempo, tanto en cursos de licenciatura como posgrado, en docentes noveles, y en quienes difieren sustancialmente en relación con su efectividad docente. Asimismo, las diferencias individuales fueron altamente estables a través del tiempo. Estos resultados refutan de manera contundente a los programas que asumen que la retroalimentación derivada de los puntajes es suficiente para mejorar la enseñanza.

### **Validez**

Los estudios de validez aportan evidencias que justifican las inferencias realizadas con base en los resultados de un instrumento (Pedhazur, 1991). La validez de los puntajes otorgados por los estudiantes al profesor se estudia desde dos perspectivas: como una medida que muestra la opinión de los estudiantes acerca de la calidad de la instrucción, y como una forma adecuada de reflejar la efectividad de la enseñanza (Abrami, d'Apollonia y Cohen, 1990).

En la consideración de que la enseñanza efectiva es un constructo hipotético difícil de validar por la imposibilidad de utilizar un criterio único, los estudios de validez de los puntajes de los estudiantes se realizan desde la aproximación de validación de constructo (Marsh y Roche, 1997; Marsh y Dunkin, 1997). En estos, la tendencia es relacionar los puntajes de los estudiantes con diferentes

indicadores como una manera de aportar evidencias de validez. El indicador por excelencia para inferir la efectividad de la enseñanza es el aprendizaje de los estudiantes, sin embargo, tiene la enorme dificultad que representa la gran cantidad de factores externos al maestro que intervienen en los procesos de enseñanza y aprendizaje.

Los estudios de validez a través de cursos simultáneos, tienen como fin establecer la relación entre el aprendizaje de los estudiantes y los puntajes obtenidos por el profesor; una conclusión a la que se llegó es que las dimensiones correlacionan de manera diferenciada (Cohen, 1987). Feldman (1997), con base en 32 estudios producidos entre 1949 y 1989, concluye que los coeficientes de correlación más altos se encuentran en: preparación y organización del curso, .57; claridad en la instrucción, .56; logro de los objetivos planteados, .49; percepción de los resultados de la instrucción por parte del alumno, .46; estimulación del interés por la materia, .38; y motivación a los estudiantes para tener alto rendimiento, .38.

En el tenor de acumular evidencias de validez, otras fuentes con relaciones significativas y consistentes con los puntajes, son las auto-evaluaciones de los docentes (Feldman, 1989) y los registros hechos por observadores externos (Murray, 1983). Una conclusión generalizada en el ámbito de la investigación sobre los cuestionarios, es que se cuenta con las evidencias de validez necesarias para justificar su utilización, sin embargo no son suficientes, por ello, es deseable que la investigación futura amplíe el rango de criterios de validez.

En los estudios de validez, es de especial relevancia investigar el impacto de los factores independientes al docente que afectan la evaluación de los estudiantes al profesor. Aunque al presente no existe consenso acerca de la definición de sesgo en los puntajes, Feldman (1997) los define como uno o más factores que directamente e inapropiadamente influyen en el juicio de los estudiantes sobre la evaluación de un curso.

Las características inherentes a una asignatura y situaciones contextuales ajenas al docente con las cuales se identifican efectos moderados de sesgo son: el tamaño del grupo (Cashin y Sixbury, 1993); la dificultad del curso (Cashin, 1988); el interés por la materia (Marsh, 1987); y las expectativas de calificación (Howard y Maxwell, 1982). Estos resultados sugieren que es inconveniente realizar comparaciones entre todos los profesores, pero es pertinente entre quienes comparten condiciones similares, por ejemplo igual tamaño de grupo - pequeño 1 a 14 alumnos; mediano 15 a 34; grande 35 en adelante (Cashin y Sixbury, 1993)- y motivación de los estudiantes.

### **El uso de los cuestionarios**

La literatura sobre la evaluación de la enseñanza por los estudiantes en la década de los noventa, alertó sobre los problemas de uso ligados al cuestionario estudiantil (Poissant, 2004). Los análisis de las prácticas de evaluación en



países como Estados Unidos, Canadá y México concluyen que por lo general se privilegian los propósitos administrativos sobre los de mejoramiento de la actividad (Saroyan, 2003; Luna, 2002). Así, el examen del funcionamiento de los sistemas indica que con un mismo procedimiento se argumenta lograr propósitos de control y de mejoramiento, con un discurso donde prevalecen los fines de mejora de la calidad docente sobre los de control (Luna, 2002). La realidad demuestra que la evaluación orientada al control se concibe como útil para el mejor funcionamiento institucional, sin embargo, no implica que ayude al desarrollo profesional individual. Al mismo tiempo, se requiere tener presente que los propósitos de la evaluación afectan los métodos.

En la consideración de que el mal uso y abuso de los cuestionarios puede llegar a invalidar la confiabilidad y validez de los instrumentos, a partir de la década de los noventa se insistió en la necesidad de ubicarlos en el contexto de un sistema de evaluación docente (Theall y Franklin, 2000). De acuerdo con Cashin (1996), en la literatura especializada del campo existe acuerdo acerca de los principios generales que deben guiar el diseño y la operación de los sistemas de evaluación de la docencia:

- Definir los propósitos de la evaluación.
- Clarificar el uso de los resultados.
- Incorporar de manera significativa a los participantes.
- Buscar la manera de considerar la historia y características del contexto particular de enseñanza en el procesamiento y análisis de los resultados.
- Producir datos y medidas técnicamente aceptables.
- Cuidar la confidencialidad de los resultados, a la luz de los intereses de los docentes y estudiantes.
- Diferenciar entre la evaluación de la enseñanza y la evaluación del trabajo académico en general.
- Instrumentar estrategias de retroalimentación y acciones para el mejoramiento.
- Incorporar múltiples fuentes de información, ya que la complejidad de la enseñanza y de los diferentes contextos educativos provoca que los cuestionarios y listas de cotejo sean insuficientes para valorar la docencia.
- Combinar el desarrollo profesional con la evaluación.
- Revisar el sistema de manera periódica.

En cuanto la utilización de los cuestionarios se cuenta con una serie de estrategias derivadas de los hallazgos de investigación, entre las cuales cabe destacar:

- La aplicación y administración estandarizada del cuestionario. Establecer un procedimiento tipo mediante el cual los estudiantes responden el cuestionario en periodos establecidos del conocimiento de la comunidad

en general. Esto incluye que los involucrados, docentes y estudiantes, conozcan el uso que se dará a los resultados (e.g. Seldin, 1993).

- La oportunidad y frecuencia de la evaluación depende de los propósitos. Si los cuestionarios son utilizados para retroalimentar al profesor con el fin de mejorar la enseñanza, deben ser contestados en el primer tercio del periodo escolar para dar oportunidad al profesor de realizar ajustes a sus prácticas; si son destinados a tomar decisiones de tipo administrativo, su aplicación se sugiere en las últimas dos semanas, pero nunca a la par de los exámenes o al final de ellos (e.g. Luna, 2002).
- La credibilidad de los resultados incrementa si los cuestionarios son contestados por un mínimo de estudiantes, 75% de los registrados en el curso (e.g. Centra, 1993).
- La confidencialidad. El sistema requiere asegurar a todos los involucrados en el proceso, confidencialidad en el manejo de la información, de tal manera que ni estudiantes ni docentes puedan sentirse amenazados por cómo se utiliza la información (e.g. Neumann, 2001).
- Las preguntas del cuestionario. El número de preguntas depende de los propósitos del cuestionario; se requiere mayor cantidad de preguntas cuando los propósitos son formativos. De manera independiente a los propósitos, se sugiere dar la oportunidad a los estudiantes a través de preguntas abiertas, de expresar con sus palabras los aspectos a mejorar (Seldin, 1993).
- Análisis de la información. El proceso de análisis, interpretación y reporte de los resultados requiere ser realizado por profesionales de la evaluación que atiendan las necesidades de los docentes, estudiantes y la institución (e.g. Theall y Franklin, 2000).
- El reporte de resultados. En la consideración que el reporte de resultados es la única retroalimentación que se puede asegurar disponible para el docente, se sugiere que éste contenga la información necesaria para interpretar los puntajes, o tener disponible un manual de interpretación (Neumann, 2000).

### **Mejora de la eficacia docente**

Diferentes autores señalan el poder orientador que tienen los procesos de evaluación; específicamente, argumentan que un programa adecuado para evaluar la docencia, que utiliza los cuestionarios contestados por los estudiantes, puede guiar al mejoramiento de la enseñanza (Marsh, 2001). De acuerdo con Centra (1993), los procesos de evaluación de la docencia logran apoyar el mejoramiento de la enseñanza a condición de cumplir cuatro criterios: proporcionar al profesor información nueva, ser información valorada por el docente, suministrar estrategias al profesor para mejorar su desempeño y que el docente se encuentre motivado para realizar mejoras. En el mismo sentido, para Seldin (1993) la utilidad de los cuestionarios depende de dos factores: de la motivación personal de los docentes para mejorar y de que el docente conozca el proceso para mejorar.

Al mismo tiempo, a nivel empírico se demostraron cambios en la efectividad de la enseñanza con base en la retroalimentación derivada de los cuestionarios; sin embargo, las modificaciones son mínimas cuando ésta consiste solo en la presentación escrita de los resultados de los cuestionarios, y se incrementa considerablemente cuando va acompañada de una asesoría personal (L'Hommediu, Menges, y Brinko, 1990). Los datos proporcionados por los puntajes de los estudiantes proveen la información base para delimitar las habilidades a desarrollar en los programas de formación, en la medida en que reportan las fortalezas y debilidades del quehacer docente. Sin duda, la información derivada de los cuestionarios debería formar parte de un diagnóstico general de necesidades, que podría ser complementado con otras metodologías que exploren a detalle necesidades particulares (Luna, Cordero y Galaz, 2007). Una asignatura pendiente es establecer el vínculo entre evaluación de la docencia y programas de formación pedagógica.

### **III. Conclusiones**

La evaluación de la docencia basada e en los cuestionarios de apreciación de los estudiantes, es la estrategia más utilizada y que acumula mayor conocimiento derivado de la investigación. No obstante, la incomodidad de los docentes sobre los puntajes y el déficit en las buenas prácticas son signos de la persistente disparidad entre el mundo de la investigación y las prácticas institucionales. Por lo general, los cuestionarios se utilizan para la toma de decisiones administrativas y para orientar la mejora de la enseñanza; una responsabilidad ineludible de los usuarios es atender los aspectos básicos que la literatura señala como fundamentales para validar su empleo.

Los cuestionarios son únicamente una medida que muestra la opinión de los estudiantes acerca de la calidad de la instrucción, por tal motivo, de ninguna forma se justifica su uso como instrumentos que cubren todas las manifestaciones de una actividad tan compleja como la docencia.

A partir de la década de los noventa, se insiste en la necesidad de utilizar los cuestionarios como parte de un sistema de evaluación docente, a los que se suma una serie de requerimientos básicos que van de la delimitación explícita de los propósitos de la evaluación, al manejo metodológico y administrativo que asegure la credibilidad en el proceso, a las acciones dirigidas para el mejoramiento de la docencia.

En la actualidad, el diseño de los cuestionarios enfrenta el reto de incorporar los cambios contextuales propios de cada situación educativa con el fin de ser congruentes con la filosofía y la misión institucional. Entre estos cambios destacan: las peculiaridades de los planes de estudio por competencias, la educación a distancia y las particularidades del contexto pedagógico por ejemplo el nivel educativo (licenciatura o posgrado). De no hacerse de esta forma, corren el riesgo ser obsoletos.

Un objetivo fundamental de la evaluación de la docencia, es contribuir a la profesionalización del profesorado, para ayudar a la mejora de la formación profesional de los estudiantes. Los puntajes de los estudiantes tienen un enorme potencial para contribuir al logro de esta meta, lo cual es factible en la medida en que se vinculen los resultados de los cuestionarios con estrategias de formación docente; este es uno de los retos más significativos en el presente.

## Referencias

Abrami, P. C., d'Apollonia, S. y Cohen, P. A. (1990). Validity of student ratings of instruction what we know and what we do not. *Journal of Educational Psychology*, 82 (2), 219-231.

Abrami, P. C., d'Apollonia, S. y Rosenfield, S. (1997). The dimensionality of student ratings of instruction: what we know and what we do not. En R. Perry y J. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 321-367). Nueva York: Agathon Press.

Arias, G. F. (1984). El inventario de comportamientos docentes (ICD): un instrumento para evaluar la calidad de la enseñanza. *Perfiles Educativos*, 4, 14-22.

Braskamp, L. A. y Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass Publishers.

Canales, A., y Gilio, M.C. (en prensa). La actividad docente en el nivel superior ¿diferir el desafío? En M. Rueda (Coord.), *La evaluación de los profesores como recurso para mejorar su práctica*. México: IISUE-UNAM.

Cashin, W. E. (1988). *Student ratings of teaching. A summary of research*. (IDEA reporte No. 20). Kansas: Kansas State University. Division of Continuing Education. (No. de servicio de reproducción de documentos ERIC ED302-567).

Cashin, W. E. (1996). *Developing an effective faculty evaluation system*. (IDEA reporte No. 33). Kansas: Kansas State University. Center for Faculty Evaluation and Development. Division of Continuing Education. (No. de servicio de reproducción de documentos ERIC ED395-536).

Cashin, W. E. y Sixbury, G. R. (1993). *Comparative data by academic field*. (IDEA report técnico No. 8). Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Centra, J. A. (1993). *Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey Bass

Cohen, P. A. (1987). *A critical analysis and reanalysis of the multisection validity meta-analysis*. Trabajo presentado en la Annual Meeting of the American Educational Research Association de 1987, Washington, DC.

Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, 30, 137-194.

Feldman, K. A. (1997). Identifying exemplary teachers and teaching: evidence from student ratings. En R. Perry y J. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 368-395). Nueva York: Agathon Press.

Gilmore, G. M., Kane, M. T. y Naccarato, R. W. (1978). The generalizability of student ratings of instruction: Estimates of teacher and course components. *Journal of Educational Measurement*, 15, 1-13.

Good, T. (1996). Teaching effects and teacher evaluation. En J. Sikula (Ed.), *Handbook of research on teacher education* (6a. ed.), (pp. 617-665). Nueva York: Macmillan.

House, E. R. (1998). Acuerdos institucionales para la evaluación. *Perspectivas*, XXVIII (1), 123-131.

Howard, G. S. y Maxwell, S. E. (1982). Do grades contaminate student evaluations of instruction? *Research in Higher Education*, 16, 175-188.

L'Hommedieu, R., Menges, R. J. y Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback. *Journal of Educational Psychology*, 82, 232-241.

Luna, S. E. (2000). Aspectos implícitos sobre la enseñanza reflejados en los cuestionarios de evaluación de la docencia. En M. Rueda y F. Díaz-Barriga (Eds.), *Evaluación de la docencia. Perspectivas actuales* (pp. 63-83). México: Paidós.

Luna, S. E. (2002). *La participación de docentes y estudiantes en la evaluación de la docencia*. México: UABC-Plaza y Valdés.

Luna, S. E. (2004). Los cuestionarios de evaluación de la docencia por parte de los alumnos: recomendaciones para su utilización. En M. Rueda y F. Díaz-Barriga (Coords.), *La evaluación de la docencia en la universidad* (pp. 98-121). México: UNAM-Plaza y Valdés.

Luna, S. E., Cordero, A. G. y Galaz, F. F. (2007, septiembre). *Los resultados de la evaluación docente y su uso para el diseño de modalidades de formación de los profesores*. Trabajo presentado en el I Congreso Internacional Nuevas Tendencias en la Formación Permanente del Profesorado, Barcelona, España.

Luna, S. E. y Rueda, M. (2001). Participación de académicos y estudiantes en la evaluación de la docencia. *Perfiles Educativos*, XXI (93), 7-27.

Marsh, H. W. (1981). The use of path analysis to estimate teacher and course effects in student ratings of instructional effectiveness. *Applied Psychological Measurement*, 6, 47-60.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Education Research*, 11 (3), 253-288.

Marsh, H.W. (2001). Students' evaluations of university teaching. Consultado el 26 de septiembre de 2007 en: [http://apps.uws.edu.au/uws/edc/seeg/SETs\\_HerbMarsh\\_presentation\\_2001.pdf](http://apps.uws.edu.au/uws/edc/seeg/SETs_HerbMarsh_presentation_2001.pdf)

Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, 99 (4), 775-790.

Marsh, H.W., y Dunkin, M. J. (1997). Students' evaluations of university teaching: A multidimensional perspective. En R. Perry y J. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 241-320). Nueva York: Agathon Press.

Marsh, H. W. y Roche, L. (1997). Making students' evaluations of teaching effectiveness effective. The critical issues of validity bias and utility. *American Psychologist*, 52 (11), 1187-1197.

McKeachie, W. J. (1996). Students ratings of teaching. En J. England, P. Hutchings y W. McKeachie (Eds.), *The professional evaluation of teaching*. American Council of Learned Societies, Occasional Paper No. 33. Disponible en: [www.acls.org/op33.htm](http://www.acls.org/op33.htm)

McLaughlin, M. y Pfeifer, R. (1988). *Teacher evaluation: Improvement, accountability and effective learning*. New York: Teachers College Press.

Murray, H. G. (1983). Low inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal Educational Psychology*, 71, 856-865.

Neumann, R. (2000). Communicating Student Evaluation of Teaching Results: Rating interpretation guides. *Assessment and Evaluation in Higher Education*, 25 (2), 121-134.

Neumann, R. (2001). A Decade of Student Evaluation of Teaching: Case Study of an Evolving System. *Journal of Institutional Research* 10 (1) 96-111. Consultado el 28 de septiembre de 2006 en: <http://www.aair.org.au/jir/May01/Neumann.pdf>

Overall, J. U. y Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology* 72, 321-325.

Pedhazur, E. J. y Schmelkin, L. P. (1991). *Measurement, design, and analysis: an integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Poissant, H. (2004). La evaluación de la enseñanza universitaria. Algunas sugerencias para optimizar su utilización. En M. Rueda (Coord.), *¿Es posible evaluar la docencia en la universidad? Experiencias en México, Canadá, Francia, España y Brasil* (pp. 85-110). México: ANUIES-UABJ.

Rueda, B. M., Elizalde, L. L. y Torquemada, A. (2003). La evaluación de la docencia en universidades mexicanas. *Revista de la Educación Superior, XXXII* (3), 71-77.

Saroyan, A. (2003). Evaluation of teaching for improvement: a myth or a reality? En M. Rueda, F. Díaz Barriga y M. Díaz Pontones (Eds.), *Evaluar para comprender y mejorar la docencia en la educación superior*, (pp. 151-160). Col. Cultura Universitaria, Serie Ensayo no. 71. México: UAM-UNAM-UABJO.

Secretaría de Educación Pública (2007). *Programa Sectorial de Educación 2007-2012*. (pp. 9-23). Gobierno de México: Autor.

Seldin, P. (1993). The use and abuse of student ratings of professors. *The Chronicle of Higher Education*, 39 (46) 40.

Theall, M. y Franklin, J. (2000). Creating responsive student ratings systems to improve evaluation practice. *New Directions for Teaching and Learning*, 83, 95-107.

---

<sup>1</sup> Este trabajo se realizó gracias al financiamiento otorgado por el Consejo Nacional de Ciencia y Tecnología al proyecto 49052.

<sup>2</sup> This work has been realized sponsored by el Consejo Nacional de Ciencia y Tecnología for the 49052 project.