
EL AVANCE DE LA EVALUACIÓN EN MÉXICO Y SUS ANTECEDENTES

JOSÉ MARÍA GARCÍA GARDUÑO

En primer lugar, agradezco al COMIE, particularmente a Eduardo de la Garza, el haberme invitado a participar como comentarista en la primera sección temática de la RMIE dedicada a la evaluación. Antes de hablar sobre el artículo de Elisa Fernández y Edna Luna (2004), al cual debo la invitación, quisiera que se me consintiese comenzar con la presentación sucinta del panorama de la evaluación, en particular de dos de las áreas que aborda esta sección temática, evaluación de programas y de la docencia, y hacer un somero balance de las mismas dimensiones sobre su avance en México; posteriormente comentaré el tema y artículo de Fernández y Luna.

Hoy en día en el sector educativo, y en otros más, es muy común hablar de la evaluación. El desarrollo de la evaluación ha sido producto de dos disciplinas que, posteriormente, convergieron en el camino: la psicología y la educación. A la psicología, en particular a la psicometría, se le debe el desarrollo de los primeros tests de rendimiento como el Stanford Achievement Test en 1923. Sin duda los psicólogos como Edward L. Thorndike tuvieron mucho que ver en ese desarrollo. Sin embargo, es poco conocido que a Ralph Tyler se le debe la creación del término evaluación y, posteriormente el de *assessment*, para referirse a la estimación del desempeño escolar, y que este gran educador e investigador fue el primero que diseñó pruebas de rendimiento basadas en los objetivos del curso, tal y como se conocen el día de hoy (Cordero y García Garduño, 2004), a principios de la década de los treinta. La distinción entre *assessment* y evaluación es muy fina y, en ocasiones, hasta confusa. Por *assessment* los estadounidenses en-

José María García Garduño, Universidad Autónoma del Estado de Hidalgo. CE: josemariagarduno@yahoo.com.mx

tienden la evaluación sistemática, usualmente a través de pruebas del desempeño escolar; en cambio, evaluación es un término más amplio.

Tyler es considerado el padre de la evaluación educativa en todos los sentidos. No sólo acuñó el término de evaluación y creó la metodología para construir pruebas objetivas de rendimiento, a principios de la década de los treinta, sino que también ideó el primer método para evaluar programas educativos a raíz del estudio curricular más grande que se haya realizado en Estados Unidos: *El estudio de los ocho años (1933-1941)* del cual Tyler, junto con sus discípula y asistente Hilda Taba, fue el coordinador de evaluación de ese gran proyecto (Ridings, 1981). La evaluación de programas se inspiró en los mismos principios del método tyleriano para diseñar tests de rendimiento: evaluar en qué medida se han logrado los objetivos; estos métodos para valorar el aprendizaje y los programas estuvieron sin cuestionarse por más de 20 años. El cuestionamiento comenzó en 1963 con la publicación de un artículo de Lee Cronbach (otro discípulo de Tyler) sobre el mejoramiento de los cursos a través de la evaluación. Fue este gran psicólogo el primero en hablar sobre la evaluación formativa, sin usar ese término (Cronbach, 1983).

Posteriormente, a raíz de la *Guerra contra la pobreza* y el *Acta de educación elemental*, promulgadas en 1965 por el Gobierno Federal estadounidense, surge la necesidad de rendir cuentas sobre los miles de millones de dólares destinados a la educación y a otros programas sociales. Para llevar a cabo la tarea se contrataron especialistas en evaluación que determinarían la utilidad de los programas (Stufflebeam y Shinkfield, 1987). Estos especialistas pronto se dieron cuenta que el método tyleriano no proveía de toda la información necesaria sobre la marcha y utilidad de los mismos. Entre 1966 y 1971 surgen las principales aportaciones teórico-metodológicas sobre la evaluación de programas que han dado pie al desarrollo de otras modelos y variantes metodológicas: el modelo CIPP (contexto, insumo proceso y producto) de Stufflebeam, el orientado al consumidor de Scriven, el de la figura de Stake (de inspiración Tyleriana) y posteriormente su método cualitativo llamado *evaluación respondente*. Este último fue el primer modelo cualitativo basado en el estudio de casos.

En la presentación a la sección temática que comentamos, De la Garza (2004a) analiza los orígenes de los modelos naturalistas de evaluación en Inglaterra en los setenta. El autor provee de información poco conocida sobre los orígenes de la tradición naturalista. Stake fue participante de las

conferencias que describe De la Garza; los trabajos de Parlett y Hamilton, creadores del modelo de evaluación iluminativa y participantes en esa conferencia, fueron muy importantes para fundamentar su modelo de evaluación respondente.

A partir de la década de los setenta la evaluación empezó una etapa de profesionalización vertiginosa. En 1975 los evaluadores más connotados comenzaron a planear la construcción de estándares de evaluación que guiarán la práctica profesional; para ello fue creado el Comité Conjunto de Evaluación Educativa que publicó en 1981 los primeros estándares de evaluación de programas. En 1994 salió a la luz la segunda edición de los estándares (Joint Committee on Standards for Educational Evaluation, 1994), se dividen en cuatro grandes áreas: utilidad, viabilidad, propiedad (ética) y precisión (método) de la evaluación. El estándar más peculiar de la evaluación, y que la hace diferente de la investigación, es que la evaluación debe ser útil. Para realizar ese propósito es necesario que, además de entregar informes oportunos, el evaluador debe considerar las necesidades y expectativas del cliente y los grupos interesados en la evaluación. Cuando se hace sin que el evaluador se involucre con esos aspectos, considerando solamente o en gran medida las creencias y expectativas del evaluador, se le denomina evaluación de caja negra (*black box evaluation*). El Comité Conjunto recomienda que el proceso de metaevaluación, evaluación de la evaluación, se lleve a cabo bajo los estándares citados.

Podría decirse que las obras de Tyler, Scriven, Stufflebeam y Stake han sido las cumbre que han cobijado, inspirado y desarrollado el campo teórico-metodológico de la evaluación de programas. Todos ellos, con excepción de Tyler, están vivos y activos dentro del campo. Recientemente Scriven fue invitado por el Centro de Evaluación de la Universidad de Western Michigan, hasta hace poco dirigido por Daniel Stufflebeam, a dirigir el *Journal of Multidisciplinary Evaluation*. Parece ser que dedicarse a la evaluación está asociado con longevidad y productividad intelectual. Para comprobarlo es sólo cuestión de mirar la actividad de esos padres (vivos) de la evaluación.

La evaluación de la docencia

Al igual que la evaluación de programas, la de la docencia comenzó y se desarrolló en Estados Unidos, pero han seguido caminos paralelos que raramente se han entroncado. Herman Remmers de la Universidad de Indiana, creó el primer cuestionario de evaluación de la docencia (CEDA)

en 1927 (Centra, 1993). Remmers y su equipo fueron, prácticamente los únicos que investigaron sobre el tema durante los siguientes 25 años.

El uso de los CEDA comenzó a popularizarse a partir de finales de la década de los sesenta. En un principio, las instituciones universitarias de ese país promovían la aplicación voluntaria, posteriormente se hizo obligatoria. Hoy en día más de 90% de las instituciones estadounidenses evalúan la docencia por medio de esos cuestionarios (Seldin, 1993). El uso de los CEDA ha promovido el desarrollo de la investigación, principalmente en Estados Unidos y Canadá. Al respecto, se pueden distinguir al menos cuatro fases en el desarrollo: la primera corresponde a la época de Remmers; la investigación fue escasa y poco sistemática; la segunda la podríamos situar en la década de los setenta donde la mayoría estuvo dedicada a determinar la validez y confiabilidad de los CEDA; la tercera puede situarse en los ochenta; aquí los investigadores estuvieron más preocupados por relacionar los resultados de los CEDA con características de los alumnos, la disciplina y el profesor (García Garduño, 2000). A partir de los noventa, después de haber sido uno de los temas más estudiados de la educación superior en décadas anteriores, el interés de los investigadores comenzó a disminuir, pero algunos siguen elaborando sobre el tema. Otros, como Herbert Marsh, probablemente el investigador más importante desde los ochenta, están cambiando de objeto de investigación. Sus intereses recientes tienen que ver con el desarrollo del autoconcepto y su relación con constructor de la motivación.

El desarrollo de la evaluación en México: un proceso interrumpido

El trabajo de Martínez Rizo (2001) presenta un panorama muy completo del desarrollo de la evaluación del aprendizaje y de las pruebas en el medio internacional y en México. De acuerdo con este autor, el nacimiento de la evaluación educativa y concretamente de la del aprendizaje en México comenzó muy temprano, en 1936, con el Instituto Nacional de Psicopedagogía. Sin embargo, a decir del autor es en la década de los noventa con el lanzamiento del Examen de Habilidades y Conocimientos Básicos (EXHCOBA), la creación del CENEVAL en 1994 y el área de evaluación de la SEP en 1995, que se ha atendido de manera sistemática la evaluación del aprendizaje.

Puede colegirse que uno de los grandes problemas que se han tenido en el país y que ha impedido el desarrollo de la evaluación y la continuidad en los proyectos y programas es, por ejemplo, que en la evaluación del

aprendizaje los registros formales son escasos o inexistentes –salvo por los testimonios– sobre el gran salto hacia delante que se tuvo durante la segunda mitad de la década de los setenta, con la creación de la Dirección General de Evaluación del SEP, encabezada por Raúl Talán. En esa Dirección se formó un grupo, la mayoría de jóvenes psicólogos egresados de la UNAM y de la UIA, que se dedicó a elaboración científica de pruebas de rendimiento. Ese equipo, entre los que se cuentan Ángel López y Mestas, María del Rayo Jiménez, Carlos García Camacho, Enoch Tamariz, Armando Vásquez, María García y Graciela Galindo, tradujeron artículos sobre el tema, desarrollaron manuales de elaboración de reactivos y de análisis de ítems; construyeron y aplicaron pruebas de rendimiento a muestras nacionales. Sin embargo, el trabajo del grupo, el tiempo y el dinero invertido ha quedado perdido en algún archivo muerto de la SEP. La historia comenzó nuevamente a partir de la década de los noventa. A pesar de todo ello, es en la evaluación del aprendizaje donde México tiene los avances importantes en esta materia. Existen tres instituciones importantes dedicadas al cultivo de la disciplina: el Instituto de Investigación y Desarrollo Educativo de la UABC (IIDE), el CENEVAL y el INEE.

En materia de evaluación de programas podría afirmarse que, en comparación con el desarrollo internacional y con el de otros países de la región, es una de las áreas que en México ha tenido menos desarrollo, a pesar de que se publicó en español uno de los primeros textos dedicados al tema, me refiero a la obra de Carol Weiss (1975): *Investigación evaluativa*.

Ese nivel escaso de desarrollo se manifiesta, por ejemplo, en que la mayoría de los trabajos de evaluación publicados se basan en el modelo de caja negra; que poco se preocupa de la utilidad de la evaluación y toma en cuenta los intereses de los usuarios del programa; básicamente, se lleva a cabo bajo la orientación tyleriana de los objetivos. Asimismo, no adopta los métodos y estándares derivados de los enfoques más importantes de la evaluación de programas, o emplea términos como evaluaciones *cuantitativas o cualitativas*, cuando en el escenario internacional se ha dejado de emplear esa terminología, la cual fue derivada de la pugna entre lo cualitativo *versus* lo cuantitativo acaecida en las décadas de los setenta y ochenta. En el campo de la evaluación, esa pugna concluyó a principios de los noventa. De acuerdo con los estándares del Comité Conjunto (Joint Committee, 1994), los informes de evaluación deben incluir tanto información cuantitativa como cualitativa.

En materia de la profesionalización de la evaluación, algunos países de la región muestran avances más notables que México, por ejemplo, en 1989 se creó la Asociación Centroamericana de Evaluación, con sede en Costa Rica. En el 2002 se crearon organizaciones profesionales de evaluación en Brasil Colombia y Perú (Chianca y Youker, 2004). En México, a pesar del dinero y esfuerzo dedicado no se ha formado ninguna asociación profesional.

Evaluación de la docencia

La excepción ha sido la *evaluación de la docencia*. En esta área México ha hecho progresos notables. El uso de los cuestionarios comenzó casi a la par en que inició su uso sistemático en Estados Unidos, a finales de la década de los sesenta. El introductor en México fue el padre Ernesto Meneses Morales, sj, de la Universidad Iberoamericana. En 1971, en la UIA se crea el primer cuestionario mexicano de evaluación docente. En esa década en la UNAM se hacen varios trabajos de investigación sobre el tema. Durante la segunda mitad de los ochenta ciertas universidades públicas y privadas comenzaron a emplear CEDA de manera sistemática en la evaluación de las tareas docentes. En 1984 Arias Galicia publicó, por primera vez, las características psicométricas de un instrumento de evaluación docente.

A partir de los noventa, a la luz de los programas de estímulo, los cuestionarios de evaluación comenzaron a ser un tema común y aceptado dentro de las universidades públicas. A mediados de esta época se crea el Grupo Interinstitucional de la Evaluación de la Docencia, al que pertenecen dos de las autoras de la sección temática publicados en este número. En este grupo, que se formó bajo el liderazgo de académicos de la UNAM, están involucradas, según la numeralía de Eduardo Garza (2004b), ocho instituciones de educación superior (seis públicas y dos privadas).

Durante los ocho años de vida, el Grupo Interinstitucional de Evaluación de la Docencia ha producido una cantidad notable de libros, tesis, números temáticos y artículos sueltos, además de haber realizado varias reuniones internacionales sobre el tema. Al respeto se está preparando un artículo de investigación que analiza el desarrollo y producción de este grupo.

Uno de los dos trabajos de este grupo presentado en el reciente número que comentamos de la RMIE es el de Elisa Fernández y Edna Luna. Las autoras buscaron indagar cuáles son los componentes de la docencia que los profesores de los programas del área de ingeniería y tecnología consideran relevantes de integrar al cuestionario de evaluación de la docencia

por los alumnos. Si sólo nos atuviéramos al objetivo del trabajo se podría concluir, erróneamente, que la investigación es simple y que no vale mucho la pena invertir tiempo y recursos en llevar a cabo una investigación de tal naturaleza. Sin embargo, el tema es valioso por sí mismo.

A pesar de la popularidad creciente de los cuestionarios de evaluación, y según lo informado en diversas investigaciones producidas por el grupo mencionado (Arbesú, 2004; Gilio, 2000; Loredó y Grijalva, 2000), el docente no ha sido, en la mayoría de los casos, un partícipe activo en el proceso de evaluación de su desempeño. Para llevar a cabo el trabajo las autoras siguieron una metodología original y válida, pues ya había sido probada en otros estudios de una de las autoras del artículo (Luna, 2002).

Los resultados de su investigación confirman los hallazgos internacionales (Cashin, 1990) y nacionales (García Garduño, 2003): las disciplinas le otorgan un peso diferente a ciertos aspectos de la evaluación docente. Asimismo, que los profesores universitarios aceptan y consideran justo y necesario ser parte activa en este proceso de evaluación. Por otro lado, el estudio señala la importancia de los profesores en el mejoramiento de este tipo de instrumentos. La retroalimentación que dieron nos muestra que en la enseñanza de las ingenierías es importante la solución de problemas y que la enseñanza en áreas que integran el plan de estudios puede ser diferente. Además, que los aspectos a evaluar sobre el desempeño docente varían de acuerdo con el tipo de curso y el nivel en que se encuentre situado dentro del plan de estudios.

El estudio de las autoras puede ser útil para iniciar una discusión sistemática sobre la posibilidad de que los CEDA incluyan aspectos generales aplicados a toda la población universitaria y una sección propia para cada disciplina o área de conocimiento. ¿Cuáles serían las limitaciones y bondades de tener un cuestionario que incluyera aspectos generales para evaluar la docencia de toda la institución y que además incluyera otros reactivos dirigidos a satisfacer las necesidades de evaluación específicas de las diferentes disciplinas o carreras? Esta estrategia es conocida como el método de *cafetería*, creada en la Universidad de Indiana en Estados Unidos. El nombre es una analogía de la manera en que el comensal elige los alimentos en una cafetería de autoservicio; de la misma forma se escogen los reactivos del cuestionario para evaluar al docente.

Podemos apreciar que en México la investigación sobre la evaluación de la docencia ha hecho progresos notables que señalan un desarrollo im-

portante y propio. Los programas de estímulo salarial y el Grupo Inter-institucional de Evaluación de la Docencia han sido las piedras de toque para tal desarrollo.

Finalmente, puede concluirse que el desarrollo del campo de la evaluación en México ha sido, de alguna manera, ahistórico. Poco ha servido lo que se hizo antes de la década de los noventa para la nueva etapa que comenzó en esos mismos años. Para favorecer su desarrollo y continuidad convendría pensar en la creación de una gran asociación mexicana de evaluación. Esta asociación podría crear una revista de mayor alcance que las actuales, que difundiera los avances de la disciplina y los hallazgos producidos. Asimismo el INEE y el CENEVAL bien podrían auspiciar la formación de una asociación profesional para superar el aislamiento y promover el desarrollo de la disciplina en México, como está ocurriendo en otros países de la región.

Referencias

- Arbesú, Ma. I. (2004). "Evaluación de la docencia: Una propuesta alternativa que considera la participación de los profesores", *Revista Mexicana de Investigación Educativa*, 9, núm. 23, pp. 863-890.
- Cashin, W. E. (1990). "Students do Rate Different Academic Fields Differently", en Theall y Franklin (eds.), *Student Ratings of Instruction, New Directions for Teaching and Learning*, núm. 43, San Francisco: Jossey-Bass, pp. 113-121.
- Chianca, Th. y Youker B. (2004). "La evaluación en América Latina y el caribe: Visión general de los desarrollos recientes", publicado originalmente en inglés en noviembre de 2004 en el primer número del *Journal of Multidisciplinary Evaluation*, y disponible en web:<http://evaluation.wmich.edu/jmde/>.
- Centra, J. A. (1993). *Reflective faculty evaluation*, San Francisco: Jossey-Bass.
- Cordero, G. y García Garduño, J. M. (2004). "The Tylerian curriculum model and the reconceptualists. Interview with Ralph W. Tyler (1902-1994)", *Revista Electrónica de Investigación Educativa*, 6 (2), disponible en <http://redie.uabc.mx/vol6no2/contenido-cordero.html>.
- De la Garza, E. (2004a). "La evaluación educativa", *Revista Mexicana de Investigación Educativa*, 9, núm. 23, pp. 807-816.
- De la Garza, E. (2004b). "Evaluación de la docencia en la educación superior", reseña, *Revista Mexicana de Investigación Educativa*, 9, núm. 23, pp. 933-946.
- Fernández, E. y Luna, E. (2004). "Evaluación de la docencia en el contexto disciplinario: La opinión de los profesores en el caso de la ingeniería y tecnología", *Revista Mexicana de Investigación Educativa*, 9, núm. 23, pp. 891-911.
- García Garduño, J. M. (2000). "Las dimensiones de la efectividad docente, validez y confiabilidad de los cuestionarios de evaluación de la docencia: síntesis de investi-

- gación internacional”, en Rueda y Díaz-Barriga (compiladores), *Evaluación de la docencia*, México: Paidós, pp. 41-62.
- García Garduño, J. M. (2003). “Profesores universitarios y su efectividad docente. Un estudio comparativo entre México y Estados Unidos”, en *Perfiles Educativos*, 25(100), pp. 42-55.
- Gilio, Ma. C (2000). “La evaluación docente desde una perspectiva participativa: el caso de la licenciatura en Química de la Universidad Autónoma de Querétaro”, en Rueda y Díaz Barriga (coords.) *Evaluación de la docencia*, México: Paidós, pp. 85-101.
- Joint Committee on Standards for Educational Evaluation (1994). *The program Evaluation Standards* (2a ed.). Thousand Oaks, CA: Sage. Existe traducción al español: Comité Conjunto de Estándares para la Evaluación Educativa (1998). *Estándares para la evaluación de programas*, Bilbao: Ediciones Mensajero.
- Loredo, J. y Grijalva, O. (2000). “Propuesta de un instrumento de evaluación de la docencia para estudios de posgrado. Querétaro”, en Rueda y Díaz Barriga, *Evaluación de la docencia*, México: Paidós, pp. 103-131.
- Luna, E. (2002). *La participación de los docentes y estudiantes en la evaluación de la docencia*, México: Universidad Autónoma de Baja California/Plaza y Valdés.
- Ridings, J. (1981). “An interview with Ralph Tyler”, *Occasional papers*, núm. 13, noviembre, disponible en [http://www.wmich.edu/evalctr/pubs/ ops/](http://www.wmich.edu/evalctr/pubs/ops/)
- Martínez Rizo, F. (2001). “Evaluación educativa y pruebas estandarizadas. Elementos para enriquecer el debate”, *Revista de la Educación Superior*, 30 (120), pp. 71-85, disponible en: www.anuies.mx/principal/servicios/publicaciones/revsup/rres120/menú5.htm.
- Seldin, P. (1993). “The Use and Abuse of Student Ratings of Professors”, *The Chronicle of Higher Education*, vol. 39, julio 21, p. 40.
- Stufflebam, D. y Shinkfield, A. (1987). *Evaluación sistemática*, Barcelona: Paidós.
- Weiss, C. (1975). *Investigación evaluativa*, México: Trillas.