

Revista mexicana de investigación educativa

COMIE

comie@servidor.unam.mx

ISSN (Versión impresa): 1405-6666

MÉXICO

2004

Eduardo L. de la Garza Vizcaya

LA EVALUACIÓN EDUCATIVA

Revista mexicana de investigación educativa, octubre-diciembre,
año/vol. IX, número 023

COMIE

Distrito Federal, México

pp. 807-816

LA EVALUACIÓN EDUCATIVA

EDUARDO L. DE LA GARZA VIZCAYA

La naturaleza de la evaluación educativa

La *Revista Mexicana de Investigación Educativa* dedica la sección temática de este número a la evaluación educativa. La evaluación, como conceptointegrante de las políticas públicas, ha estado presente en el contexto educativo mexicano, en los últimos veinte años. El término ha estado vinculado con otro concepto: el de la calidad de la educación. A la evaluación se la concibe como una actividad indispensable y previa a toda acción conducente a elevar el nivel de la calidad de la educación. La evaluación así formulada constituye un momento de la planeación, entendida ésta como una acción racional dotada de propósito. Se identifica a la evaluación con la última etapa del proceso natural del conocimiento que concluye con la emisión de juicios informados, proceso que antecede a las decisiones y a la acción humana.

En su forma más simple, la evaluación conduce a un juicio sobre el valor de algo y se expresa mediante la opinión de que ese algo es significativo. Se llega a este juicio calificando qué tan bien un objeto reúne un conjunto de estándares o criterios. Así, la evaluación es esencialmente comparativa. Supone la adopción de un conjunto de estándares y la especificación del grupo contra el cual el objeto es comparado. El objeto puede ser calificado como “bueno” o “malo”, cuando la referencia es la totalidad de los objetos o el objeto promedio del grupo; o bien como “mejor” o “peor”, si es comparado con un subconjunto particular del grupo. La evaluación, en esencia, supone adoptar un conjunto de estándares, definirlos, especificar el grupo de comparación y deducir el grado en el cual el objeto alcanza los estándares. Una vez realizado lo anterior, el evaluador

Eduardo L. de la Garza Vizcaya es profesor titular de la Universidad Autónoma Metropolitana, Unidad Azcapotzalco: edelag@correo.azc.uam.mx

está en posibilidad de hacer, en un segundo momento, un juicio sobre el valor del objeto evaluado.

Es preciso notar que no todas las evaluaciones son racionales y explícitas; los estudios de caso revelan los estándares, comparaciones y juicios de manera implícita e intuitiva, y algunos buscan entender más que juzgar. El juicio de valor puede ser expresado mediante palabras ordinarias de tipo descriptivo y no necesariamente mediante calificativos de “bueno” o “malo”. Los ejemplos y anécdotas en una descripción pueden servir para este propósito, ya que el juicio de valor reside en la forma en que las palabras son usadas.

El proceso de evaluación real es complejo y no se ajusta a una mera deducción a partir de ciertos estándares; más bien se trata de un proceso dialéctico entre principios abstractos y ejemplos concretos, en donde quien evalúa verifica sus juicios contra aquellas situaciones particulares en las que se siente confiado de conocer sus resultados. Aún más, los estándares pueden estar en conflicto unos con otros, y será necesario balancearlos así como establecer precedencias relativas entre ellos, algunas veces de manera intuitiva. Cuando el evaluador es distinto al que toma las decisiones, hay otro factor que interviene: la credibilidad. La preocupación del tomador de decisiones no es sólo atender al grupo de comparación y a los estándares, y cómo éstos son pesados relativamente entre sí, sino también si los hechos son reales y si están presentados por el evaluador en forma correcta.

Normalmente, la evaluación educativa supone la existencia de acuerdos comunes: juicios previos consensuados, sobre lo que es bueno o conveniente para todos como propósito de las acciones educativas. Este es el espacio que permite el disenso y la controversia sobre el sentido, el propósito y el uso de la evaluación en la educación. Es, asimismo, el que permite adjudicarle un carácter ideológico. La evaluación no es meramente un proceso técnico, neutral e inocuo; por el contrario, como toda acción social posee una dimensión política y, en su carácter instrumental, obedece a una lógica de control. Estos aspectos constitutivos de la evaluación están presentes, lo mismo en una actividad tan antigua y concomitante a los procesos de educación como la evaluación en el aula, como en las recientes políticas gubernamentales identificadas con procesos de metagestión, los cuales condicionan y vinculan recursos económicos a los resultados de la evaluación.

La evaluación asumida como un campo particular de la investigación educativa es reciente. Se ha cuestionado incluso si los proyectos de evaluación pueden considerarse investigación educativa. Se arguye que el mismo carácter instrumental de la evaluación contraviene el sentido desinteresado de la obtención de nuevo conocimiento que subyace a la lógica de la ciencia. Lo cierto es que, en un sentido amplio, el propósito de los proyectos de evaluación es la obtención de conocimiento, si bien se busca que éste sea útil para la toma de decisiones. La mayor parte de los estudios de evaluación requiere, asimismo, diseños metodológicos complejos y debe satisfacer normas que justifican la validez y confiabilidad del conocimiento que obtiene. En ello, las evaluaciones no se apartan de la parsimonia inherente a las investigaciones científicas.

Por otra parte, el fenómeno educativo, como el hombre mismo, es complejo y su conocimiento es siempre parcial e idiosincrático, pues está constreñido a la condición temporal de sus actores y posee una dimensión inevitablemente subjetiva. En este sentido, cualquier pretensión “científica” de las investigaciones educativas debe ser matizada, dado el carácter peculiar de su objeto de estudio. En lo que respecta a los estudios de evaluación, en los últimos años, se ha generalizado su aceptación como una rama de la investigación aplicada en educación.

El origen y evolución de la evaluación educativa

La evaluación educativa tiene su origen en los primeros años del siglo XX, en Estados Unidos. Está influida por las ideas de progreso, de la administración científica y por la ideología de la eficiencia social. El modelo de los objetivos de aprendizaje impactó en los años treinta el desarrollo curricular y su evaluación. Para los sesenta era una tecnología poderosa que, junto con los diseños experimentales, constituían la metodología tradicional de la evaluación educativa. Evaluación, entonces, era sinónimo de medición, prueba o examen. Con la expansión educativa en las décadas de los cincuenta y sesenta surgieron las preocupaciones por la rendición de cuentas y la relación costo-beneficio, lo que condujo al empleo de sistemas de información y de procedimientos para monitorear los programas como parte de la administración social. A finales de los sesenta, se hace la crítica del modelo tradicional de evaluación y, en la siguiente década, surgen nuevas formas para evaluar las innovaciones educativas.

Las críticas expresadas por evaluadores connotados como Atkin (1963), Cronbach (1963, 1975), Scriven (1967), Stake (1967), Parlett y Hamilton (1972) o Stenhouse (1975), señalaban que el modelo de objetivos limitaba la reforma curricular y trivializaba la naturaleza del aprendizaje. Algunos de ellos abogaban por una evaluación que permitiera entender el proceso de la innovación y no sólo emitir juicios acerca de su efectividad. Se decía que las evaluaciones no respondían a las preguntas de los educadores. Estos evaluadores enfatizaban la utilidad de la evaluación en el mejoramiento de los cursos y no en la obtención de juicios terminales. Destacaban el carácter multidimensional de cualquier innovación educativa, el cual no podía alcanzarse con los exámenes. Se decía que el modelo impedía juzgar el valor de las metas propuestas, pues separaba los hechos fácticos de los valores implícitos. Alguno señalaba que conducía a una visión equivocada de la educación, ya que su pretensión de determinar las conductas de los estudiantes contravenía el propósito de la verdadera educación: el desarrollo de la autonomía personal, cuyos resultados son, en sí mismos, impredecibles. El diseño experimental, por otra parte, había mostrado que con frecuencia se daba un conflicto entre la confiabilidad y la relevancia de los resultados.

En la evolución subsiguiente de la evaluación educativa fueron relevantes cuatro conferencias realizadas en Cambridge, Inglaterra, a partir del inicio de los setenta, con el objeto de explorar acercamientos no tradicionales a la evaluación educativa y establecer algunas guías para su desarrollo futuro. Se buscaba justificar prácticas metodológicas y funciones políticas alternativas para la evaluación. El cambio de paradigma abrevaba en la tradición de las investigaciones humanistas, fuera de la psicometría, el experimentalismo y las encuestas sociales.

En la primera de ellas, llevada a cabo en 1972 (Hamilton, 1977), se estableció que la evaluación debería: *a)* responder a las necesidades y perspectivas de diferentes audiencias, *b)* iluminar los procesos complejos organizacionales, de enseñanza y de aprendizaje, *c)* ser relevante para las decisiones públicas y profesionales por venir, y *d)* reportarse en un lenguaje accesible a las audiencias. Recomendaba también que fueran usados datos proporcionados por la observación, cuidadosamente validados (algunas veces en sustitución de datos proveídos por preguntas y exámenes); que la evaluación se diseñara de tal forma que fuera lo suficientemente flexible para responder a eventos no anticipados (que su foco fuera obtenido de manera

progresiva, en lugar de con un diseño preordenado), y que la posiciones de valor del evaluador, tanto si fueran resaltadas como constreñidas por el diseño, se hicieran evidentes a los patrocinadores y a las audiencias de las evaluaciones.

Las tres restantes conferencias se llevaron a cabo en 1975 (Simons, 1980), cuyo foco fueron los métodos de la investigación mediante estudios de caso; en 1979, para examinar el uso de los métodos naturalistas, y en 1987 para considerar la evaluación del desempeño y diferentes formas de evaluar a maestros y alumnos.

Para 1979, había adquirido importancia política la rendición de cuentas y se hacía un énfasis excesivo en la administración. Para entonces, la evaluación educativa compartía los problemas de la investigación vinculada con las políticas. Entre los problemas detectados en la tercera conferencia se encontraban: la inadecuación de las teorías sobre políticas y los procesos para establecerlas; el énfasis usual en la evaluación de programas en cuanto opuesto a la evaluación de las políticas; el deterioro en la relación entre evaluadores y sus patrocinadores, clientes y otras audiencias, con frecuencia debido a la abierta naturaleza política de la innovación; la dificultad de hacer coincidir las necesidades de información de los tomadores de decisiones y el servicio a audiencias múltiples dentro de la escala de tiempo establecida por los administradores de los programas; una historia de productos rechazados, especialmente por los administradores quienes esperaban reportes que simplificaran sus tareas en lugar de hacerlas más complejas; la dificultad de mantener la independencia; la vulnerabilidad personal e institucional vinculada con las formas de evaluación naturalista las cuales emplean retratos de las personas y descripciones amplias de los climas y contextos institucionales, y la naturaleza conservadora de esta misma investigación, la cual favorecía la descripción sobre el análisis.

A fines de la década setenta era claro que la evaluación había tenido poco impacto en las políticas educativas y que, en donde lo había tenido, había hecho más complejo el proceso de toma de decisiones. No se alcanzaba el propósito de proporcionar información útil y confiable.

El nuevo paradigma de la evaluación educativa

La nueva evaluación que surge de los cuestionamientos anteriores posee las siguientes características: *a)* es pragmática desde el punto de vista metodológico, *b)* responde a situaciones particulares: pone su atención

en las preguntas, preocupaciones, problemas y necesidades de información de los implicados y de los tomadores de decisión, *c)* hace a un lado el paradigma hipotético deductivo y acoge otro de decisiones que enfatizan métodos múltiples, acercamientos alternativos y el hacer coincidir los métodos de la evaluación con las preguntas y las situaciones específicas. Se llega a decir que el propósito más importante de la evaluación es responder a los requerimientos de información de las audiencias, particularmente, en formas que tomen en cuenta las variadas perspectivas de sus miembros. La intención es incrementar la utilidad de los resultados, aun sacrificando la precisión proporcionada por las mediciones.

Investigación *naturalista* ha sido el término genérico que fue usado para describir varios de los acercamientos alternativos a la evaluación que obtuvieron prominencia en los setenta como reacción a las formas tradicionales. Se han caracterizado como acercamientos que atienden la pluralidad de las audiencias y asumen el carácter político de la evaluación. Sus rasgos se expresan de la siguiente manera:

Comparados con los métodos clásicos, tienden a ser más extensivos (no necesariamente centrados en datos numéricos), más naturalistas (basados en la actividad del programa o proceso, más que en su intento o logro), y más adaptables (no constreñidos por diseños experimentales o preordenados). Asimismo, muy probablemente son sensibles a los diferentes valores de los participantes en el programa, suscriben métodos empíricos que incorporan trabajo de campo etnográfico, desarrollan materiales de retroalimentación formulados en el lenguaje natural de los destinatarios, y desplazan el lugar de los juicios formales, del evaluador a los participantes (Hamilton, 1977).

Significan un compromiso de estudiar los programas en su contexto social, el uso de métodos cualitativos de indagación, tales como las entrevistas no estructuradas, la observación directa y la reconstrucción histórico dramática, y formas de reportar que permiten a los lectores generalizar por sí mismos, utilizando una generalización *naturalista* (Simona, 1987).

Dirigidos a audiencias no técnicas resaltando el uso del lenguaje ordinario, el razonamiento informal y cotidiano y el uso de argumentos para tratar de establecer la estructura de la realidad. Estos argumentos son: los ejemplos, las metáforas, las ilustraciones y las analogías (House, 1980).

La evaluación educativa en México en los últimos años

Las acciones y los estudios de evaluación han estado creciendo y se han consolidado en el medio mexicano en los últimos quince años. Las pruebas objetivas en gran escala, aplicadas a poblaciones del nivel básico, estudiantes y en menor medida maestros, recibieron un impulso importante a mediados de los noventa desde la Secretaría de Educación Pública y, recientemente, con la creación en 2002 del Instituto Nacional para la Evaluación de la Educación (INEE). La creación en 1994 del Centro Nacional de Evaluación para la Educación Superior (CENEVAL) generalizó también el empleo de estas pruebas en el ámbito de la educación superior. En este nivel, ha cobrado importancia la evaluación de programas a través de pares a partir de la creación de los Comités Interinstitucionales para la Evaluación de la Educación Superior en 1991 y, desde 1994, de los organismos integrados en el Consejo para la Acreditación de la Educación Superior, en 2000. Últimamente, también en el ámbito del nivel superior, ha destacado el Programa Integral de Fortalecimiento Institucional (PIFI) iniciado en 2001, para impulsar la evaluación y planeación de las instituciones.

En el nivel de la educación básica es pertinente mencionar dos líneas más cultivadas actualmente por el INEE. La primera, es la evaluación cualitativa de las escuelas, la cual había sido iniciada previamente por la SEP en programas como el de Escuelas de Calidad. La segunda línea se refiere a la definición de indicadores educativos, aspecto también atendido previamente en la SEP. Desde su creación, el INEE ha fomentado la elaboración de investigaciones y estudios vinculados con sus tareas de evaluación; recientemente lo ha hecho mediante una convocatoria abierta a la comunidad de investigadores. También el CENEVAL ha sido un factor importante para impulsar la elaboración de estudios, pues ha auspiciado, desde 1995, la celebración de foros nacionales bianuales dedicados a la evaluación de la educación. El recuento anterior no pretende ser exhaustivo, sino sólo mostrar el creciente impulso de las actividades de evaluación en nuestro medio, muchas de ellas derivadas de políticas gubernamentales que hacen eco a tendencias de orden internacional. Tampoco es aquí el lugar para hacer una relación de las investigaciones o estudios de corte evaluativo realizados en el país, lo cual rebasaría el propósito de esta introducción. Baste decir que constituyen un campo de investigación creciente y en proceso de maduración.

Las investigaciones en este número

A pesar de la afirmación anterior, la repuesta de la comunidad a la convocatoria para este tema fue reducida. Se presentaron sólo cinco trabajos, todos del medio mexicano. Cuatro de ellos fueron aprobados para integrar la sección que ahora introducimos. En él, el lector encontrará una muestra de las investigaciones educativas de corte evaluativo que se realizan actualmente en el medio mexicano. Tres de ellas son de naturaleza empírica. En éstas, se ha pedido expresamente a los autores reflexionar sobre la metodología utilizada. Consideramos, como debe haber quedado claro con las ideas anteriores, que las investigaciones de corte evaluativo se benefician al explicitar y justificar su método, en relación con su propósito.

El primer trabajo, “¿Aprobar o reprobar? El sentido de la evaluación en educación básica”, es una investigación documental que cuestiona, con base en experiencias y en la evidencia de estudios empíricos realizados en contextos distintos al mexicano, la práctica de la reprobación en ese nivel educativo. En la medida en que se considera al éxito un indicador de la calidad educativa, esta discusión se hace pertinente. Los lectores familiarizados con la educación en los niveles superiores encontrarán también resonancia en algunas conclusiones que vinculan el abandono con la reprobación.

El segundo texto, “Evaluación del desempeño en lectura y escritura. Aportes empíricos a la noción de componentes lingüísticos en el cuarto grado de primaria”, se detiene en el debate sobre la evaluación del aprendizaje, su sentido y el uso de sus resultados. La investigación postula la necesidad de explicitar “una teoría que sustente tanto a los medios de medición utilizados como a las características y propiedades a evaluar” y aporta hallazgos significativos sobre la pertinencia de organizar en bloques el currículo de español en primaria.

El tercer artículo, “Evaluación de la docencia universitaria: una propuesta alternativa que considera la participación de los profesores”, es una investigación que privilegia el carácter formativo de la evaluación, es decir su utilidad para mejorar la calidad del mismo proceso educativo. El trabajo pretende contribuir a la construcción de un modelo de evaluación desde la práctica, “a partir del análisis de la vida cotidiana en las aulas”. El evaluador en este modelo, “acompaña” al docente como consultor y orientador, ayudándole en la reflexión sobre su propia práctica.

El cuarto trabajo, “Evaluación de la docencia y el contexto disciplinario: la opinión de los profesores en el caso de ingeniería y tecnología”, es un estudio que explora las dimensiones de la docencia que, a sugerencia de los propios docentes, deben incorporarse a un cuestionario de evaluación. La investigación identifica componentes relevantes de la enseñanza en la disciplina de estudio. Como en los dos trabajos anteriores, los autores reflexionan sobre el estado del arte de la investigación y las metodologías empleadas, en este caso para la evaluación de la docencia, y hacen aportaciones importantes sobre el uso de los cuestionarios de opinión de alumnos como recurso generalizado.

La entrega del número temático se enriquece con una colaboración en la sección de Aporte de discusión de dos investigadores, quienes –desde la experiencia de la evaluación en el contexto español– abogan por la necesidad de iniciar procesos de “metaevaluación”. Presentan una serie de criterios o parámetros que deben ser satisfechos por las evaluaciones. En su propuesta, el propósito de la metaevaluación es “avivar la reflexión para generar comprensión rica y relevante sobre los programas evaluados” y, al mismo tiempo, garantizar la calidad de la evaluación. Los autores destacan seis contextos “básicos” de toda evaluación que permiten explorar los elementos a ser considerados en la metaevaluación. Más adelante, se detienen en los aspectos de la validez y confiabilidad de las evaluaciones, y discuten su sentido en los distintos modelos utilizados. Se puede estar de acuerdo o no con las posiciones de los autores y el tratamiento que hacen de los tópicos, pero resulta interesante el propósito, aunque no del todo exitoso, de sistematizar en una perspectiva única paradigmas cuantitativos y cualitativos que, hasta hace poco, se ubicaban en parcelas irreconciliables.

La colaboración termina discutiendo tópicos relevantes en los estudios de evaluación: su propósito, la generalización del conocimiento obtenido, la ética y el aprendizaje.

Bibliografía

- Atkin, J. Myron (1963). “Some evaluation problems in a course content improvement project”, *Journal of Research in Teaching* 1, pp.129-32.
- Cronbach, Lee J. (1963) “Course improvement through evaluation”, *Teachers College Record* 64, may, pp. 672-86.
- Cronbach, Lee J. (1975) “Beyond the two disciplines of scientific psychology” *American Psychologist* 30, 2, pp. 116-27.

- Hamilton, David *et alii* (eds.) (1977) *Beyond the numbers game. A reader in educational evaluation*, Londres: Macmillan Education.
- House, Ernest (1980) *Evaluating with validity*, Beverly Hills: Sage Publications.
- Lincoln, Yvonna & Guba, Egon (1985) *Naturalistic inquiry*, Beverly Hills: Sage Publications.
- Martínez R., Felipe (2003). *Elementos para la definición de una política nacional de evaluación educativa* México, D F: Instituto Nacional para la Evaluación de la Educación.
- Norris, Nigel (1993). *Understanding educational evaluation*, Londres: Kogan Page, CARE-UEA.
- Parlett, Malcolm y Hamilton, David (1972). *Evaluation as illumination: a new approach in the study of innovatory programmes*, Occasional Paper 9, Center for Research in Educational Sciences, University of Edinburgh. Reimpreso en Glass, G. (ed.) (1976) *Evaluation Studies Review Annual*, I, Beverly Hills: Sage Publications, pp. 140-57.
- Scriven, Michael (1967) "The methodology of evaluation" en Tyler, R. W., Gagne, R. M. y Scriven, M. *Perspectives of curriculum evaluation* AERA Monograph Series on Curriculum núm. 1, Chicago: Rand McNally.
- Simons, Helen (1980). *Towards a science of the singular* Norwich, G. B.: CARE-UEA.
- Simons, Helen (1987). *Getting to know schools in a democracy. The politics and process of evaluation*, Londres: The Falmer Press.
- Stake, Robert (1967). "The countenance of educational evaluation", *Teachers College Record* 68, abril, pp. 523-40.
- Stenhouse, Lawrence (1975). *An introduction to curriculum research and development*, Londres: Heinemann.