

nº3 enero - junio 2010

G^ap_p

Gestión y Análisis de Políticas Públicas

SUMARIO

7_PRESENTACIÓN DEL NÚMERO - Educación, política y políticas de evaluación educativa - *Saville Kushner y Ester García Sánchez* / 17_ESTUDIOS - El desarrollo de los centros de evaluación y la formación de los evaluadores - *Robert E. Stake* / 35_En búsqueda de la explicación de los resultados educativos: posibilidades y limitaciones de los estudios internacionales de evaluación - *Alejandro Tiana Ferrer* / 67_La apropiación de la evaluación por las administraciones públicas - *Saville Kushner* / 81_EXPERIENCIAS Y CASOS - La evaluación de políticas públicas en el nivel estratégico: sistema general de becas, abandono escolar prematuro y mejora del capital humano - *M.ª Luisa Merino Cuesta* / 113_Autoevaluación y políticas públicas: una experiencia en escuelas primarias argentinas - *Elena Duro y Olga Nirenberg* / 139_Metaevaluación del sistema de evaluación de las titulaciones universitarias en Andalucía - *Baltasar Fernández-Ramírez, Enrique Reboloso Pacheco y Pilar Cantón Andrés* / 167_ABSTRACTS / 173_AUTORES

INAP



GOBIERNO
DE ESPAÑA

MINISTERIO
DE LA PRESIDENCIA

NUEVA ÉPOCA

nº 3 enero-junio 2010



Gestión y Análisis de Políticas Públicas

INAP

INSTITUTO
NACIONAL
DE ADMINISTRACIÓN
PÚBLICA



nº 3 enero-junio 2010

DIRECTORA DEL INAP:
Pilar Arranz Notario

CONSEJO ASESOR:

Inés Alberdi Alonso, Luis F. Aguilar Villanueva, Mariano Baena del Alcázar, María Luisa Carcedo Rocas, Antonio Fernández Poyato, José Enrique Martín Arahuetes, Ives Mény, Marga Pröhl, Paul A. Sabatier, Julio Segura Sánchez, Nicoletta Stame, Josep María Vallès Casadevall, Rosa Virós i Galtier

CONSEJO DE REDACCIÓN:

Director de la revista: Francisco J. Vanaclocha Bellver
Vocales: Rafael Bañón Martínez, José Manuel Canales Aliende, Cecilia Castaño Collado, Rafael Catalá Polo, Laura Chaqués Bonafont, Concepción Toquero Plaza, Manuel Villoria Mendieta, Eduardo Zapico Goñi
Secretaria del Consejo de Redacción: María Isabel Wences Simon
Vicesecretario del Consejo de Redacción: Alfonso Fernández Burgos

EDITA

INSTITUTO NACIONAL DE ADMINISTRACIÓN PÚBLICA
NIPO: 010-10-005-0 (formato papel); 010-10-011-0 (formato electrónico)
ISSN: 1134-6035 (formato papel); 1989-8991 (formato electrónico)
DEPÓSITO LEGAL: M-39.657-1994

Preimpresión: Lerko Print, S.A.
Impresión: Publidisa

Con la colaboración del Instituto de Política y Gobernanza "Fermín Caballero" de la Universidad Carlos III de Madrid (IPOLGOB)

SUSCRIPCIONES Y VENTA

Secretaría de la revista Gestión y Análisis de Políticas Públicas

CENTRO DE PUBLICACIONES

INSTITUTO NACIONAL DE ADMINISTRACIÓN PÚBLICA

Atocha, 106, 28012 Madrid

Tfno.: +34 - 91 273 91 19

Fax.: +34 - 91 273 92 87

www.inap.map.es

www.inap.es/gapp

e-mail: revistagapp@inap.map.es

Catálogo general de publicaciones oficiales: <http://www.060.es>

PRECIOS

Número suelto 15,00 €

Número doble 30,00 €

Suscripción anual 27,03 €

Incluidos IVA y gastos de envío

Índice

| | <u>Página</u> |
|---|---------------|
| PRESENTACIÓN DEL NÚMERO | 7 |
| <hr/> | |
| ESTUDIOS | 17 |
| El desarrollo de los centros de evaluación y la formación de los evaluadores <i>Robert E. Stake</i> | 17 |
| En búsqueda de la explicación de los resultados educativos: posibilidades y limitaciones de los estudios internacionales de evaluación <i>Alejandro Tiana Ferrer</i> | 35 |
| La <i>apropiación</i> de la evaluación por las administraciones públicas <i>Saville Kushner</i> | 67 |
| <hr/> | |
| EXPERIENCIAS Y CASOS | 81 |
| La evaluación de políticas públicas en el nivel estratégico: sistema general de becas, abandono escolar prematuro y mejora del capital humano <i>M.ª Luisa Merino Cuesta</i> | 81 |
| Autoevaluación y políticas públicas: una experiencia en escuelas primarias argentinas <i>Elena Duro y Olga Nirenberg</i> | 113 |
| Metaevaluación del sistema de evaluación de las titulaciones universitarias en Andalucía <i>Baltasar Fernández-Ramírez, Enrique Reboloso Pacheco y Pilar Cantón Andrés</i> | 139 |
| <hr/> | |
| ABSTRACTS | 167 |
| <hr/> | |
| AUTORES | 173 |
| <hr/> | |

Educación, política y políticas de evaluación educativa

Saville Kushner
Ester García Sánchez

“Cuanto más seguro me siento de que estoy en lo cierto, tanto más corro el riesgo de dogmatizar mi postura, de *congelarme* en ella, de encerrarme sectariamente en el círculo de mi verdad.”

Paulo Freire, *Política y educación*

El monográfico que el lector tiene ahora en sus manos responde a un doble propósito. En primer lugar, el de ofrecer un conjunto de trabajos de notable interés y muy representativos del actual estado de la cuestión en la evaluación en el ámbito educativo. En segundo lugar, reflexionar y debatir acerca de los puntos de encuentro entre dos ámbitos que, teniendo como tienen algunos intereses comunes (las políticas y los programas educativos), curiosamente han seguido desarrollos siempre paralelos, sin que haya habido apenas entre ellos un breve “intercambio de miradas”. De un lado, la denominada *evaluación educativa* –en la que el predominio de la pedagogía y la psicología es muy notorio– y, de otro, el análisis (y la evaluación de políticas), en tanto que subdisciplina de la Ciencia Política². No es casual, por tanto, que el monográfico vea la luz en una revista como *Gestión y Análisis de Políticas Públicas*, como tampoco lo es que quienes coordinamos el monográfico procedamos, precisamente, de esos dos mundos.

1 La reflexión que hacemos en estas páginas se refiere esencialmente al caso español, aunque parece probable que esta distancia se dé también en otros países.

2 Es cierto que el análisis de políticas públicas y la evaluación de políticas no han ido ni van siempre de la mano, ni en los estudios teóricos ni en los de carácter aplicado. Sin embargo, y salvo que indiquemos expresamente lo contrario, para esta presentación nos permitiremos aludir a ellas como si fueran dos campos plenamente integrados. El objetivo es “contraoponer” aquella actividad de evaluación de políticas y programas educativos más próxima al enfoque de políticas públicas (y, consiguientemente, a la Ciencia Política) y aquellas otras que se llevan a cabo desde disciplinas como la Pedagogía o la Psicología.

La evaluación de políticas públicas se ha conformado como un campo multidisciplinar con una identidad propia (...), en el que desarrollan su actividad especialistas procedentes de distintos campos y perspectivas y en los que, con el tiempo, se han ido perfilando una serie de tipos y métodos de evaluación, en función de sus respectivas tradiciones e intereses (Bañón 2003; Vanaclócha, García Sánchez y Viñas, 2005). Siendo esto cierto, el simple hecho de que los dos campos de investigación a los que hacíamos referencia al principio compartan un mismo objeto de estudio debiera haber servido para propiciar la reflexión y el aprendizaje mutuo, aunque las perspectivas y los métodos de análisis de cada uno de ellos sean distintos. El saber parece cada día estar más “compartimentado” no ya en disciplinas (lo cual resulta lógico) sino en áreas de conocimiento (en el sentido más “administrativo” del término) que rara vez conversan –aun a sabiendas de que el tema de conversación es el mismo y de que lo que cambia, esencialmente, es el lenguaje (léase, el andamiaje conceptual y metodológico) con el que lo abordamos. Precisamente, uno de los objetivos de este monográfico es contribuir modestamente a tender puentes entre los dos ámbitos mencionados.

¿Cuáles podrían ser los espacios secantes de esos mundos? ¿De qué manera podrían contribuir el análisis y la evaluación de políticas públicas al enriquecimiento de la evaluación educativa? ¿Tendrían algo que aprender de esta última el *policy analysis* y la evaluación de políticas?

1. Hay al menos dos motivos por los que la evaluación educativa merece la atención del *policy analysis* y la Ciencia Política. En primer lugar, la propia riqueza de ese campo de especialidad y la gran difusión de la que goza el término *evaluación educativa*. Bajo esta rúbrica, encontramos muy distintos tipos de evaluación, desde la que se centra en las políticas o en los programas educativos, hasta la que se ocupa de los centros escolares, pasando por la que se dirige al profesorado, las enseñanzas o las instituciones universitarias. El término está sólidamente acuñado y es perfectamente reconocible para cualquier evaluador que se interese por la realidad educativa, a pesar de que la definición precisa del término plantee no pocos problemas (Kellaghan, Stufflebeam y Wingate, 2003; Nevo, 2006; Schwandt, 2009). Más allá de las razones de índole práctico que han podido llevar a integrar tipos tan distintos bajo una denominación tan genérica, la amplitud del término (o su concreción, según se mire) tiene su razón de ser en la existencia de una serie de cuestiones y elementos teóricos y metodológicos comunes a esos diferentes tipos de evaluación³.

En las páginas que siguen, el lector encontrará numerosas alusiones a la idea de *evaluación educativa*. Algunos de los trabajos que aquí se incluyen no abordan directamente la evaluación de políticas o de programas educativos sino que se ocupan de la evaluación de centros, de titulaciones universitarias o de los estudios internacionales de evaluación del rendimiento o bien son reflexiones teóricas de carácter general.

³ Ello no debe impedir al evaluador ser consciente de que la naturaleza del objeto evaluado condiciona la propia actividad evaluadora y de que entre unos tipos de evaluación y otros pueden existir importantes diferencias.

Quede claro, antes de proseguir, que no nos interesan aquí los modelos o los sistemas de evaluación de centros, alumnos o enseñanzas en sí mismos, sino solamente en la medida en que de su análisis podamos extraer lecciones y aprendizajes o en el caso de que constituyan el eje vertebrador de determinadas políticas de evaluación.

Pero la importancia de la evaluación educativa va mucho más allá del amplio reconocimiento con el que cuenta dentro y fuera del mundo académico. Y es que la evaluación de programas es, en gran medida, deudora de la evaluación educativa. Buena parte de los enfoques, modelos, incluso estándares, de los que se sirven hoy los evaluadores (de cualquier parte del mundo y de cualquier campo de actividad) se gestaron en el seno de la evaluación educativa⁴. Nadie puede negar la contribución al desarrollo y consolidación de la evaluación de autores como Ralph Tyler, Lee Cronbach, Egon Guba, Robert Stake, Carol Weiss, Ernest House, Barry McDonald, Yvonna Lincoln, Daniel Stufflebeam, Marvin Alkin, Worthen y Sanders o, más recientemente, William Shadish, Thomas Schwandt, Jennifer Greene o David Fetterman⁵, todos ellos eminentes psicólogos o pedagogos. Tampoco pueden obviarse, por citar otro ejemplo, los trabajos del *Joint Committee for Standards in Educational Evaluation*⁶, una referencia absolutamente imprescindible para todo aquel que desee adentrarse en el tema de los criterios y los estándares de evaluación.

No obstante, y pese a la evidente trascendencia de la evaluación educativa, el término no ha despertado demasiada atención entre quienes se dedican a la evaluación de políticas y programas educativos desde las perspectivas de la Ciencia Política y del enfoque de políticas públicas. Desde ellas, el término *evaluación educativa* se suele asociar exclusivamente –y no sin cierto recelo– a la evaluación del aprendizaje de los alumnos. Es lógico que, siendo así, ni la Ciencia Política ni el análisis de políticas públicas –cuyo foco de atención son las políticas y los programas públicos– muestren o hayan mostrado interés por la evaluación educativa. Pero, detrás del desinterés por el término se encuentran otros factores.

El primero tiene que ver con los problemas que plantea, en nuestro idioma, la misma palabra “evaluación”. En castellano, disponemos de un único significante para expresar tres significados distintos, algo que no sucede en los países de habla inglesa, en los que cuentan con tres vocablos diferentes: *evaluation*, *assessment* y *appraisal*⁷.

4 Son numerosísimos los trabajos que han dado cuenta de la contribución de la evaluación educativa a la génesis y desarrollo de la evaluación. Pueden consultarse, entre otros, Guba y Lincoln (1989) y Nevo (2009).

5 El lector encontrará en el apartado de “bibliografía” las referencias exactas de las principales obras de estos autores, que hemos omitido aquí con objeto de no alargar excesivamente la exposición en el cuerpo del texto.

6 El *Joint Committee* se creó en 1975, fruto de la iniciativa de distintas asociaciones por promover la calidad de las evaluaciones. Este organismo ha publicado tres catálogos de estándares: *The Personnel Evaluation Standards*, en 1988, *The Program Evaluation Standards*, cuya segunda edición apareció en 1994 y *The Student Evaluation Standards* en 2003.

7 Según las definiciones propuestas por el *Joint Committee on Standards for Educational Evaluation* (1994).

El término *assessment* hace referencia a aquellos estudios que sirven para determinar el nivel o el valor de un objeto en una determinada variable aunque, más específicamente, se emplea para hablar de la evaluación de aprendizajes y/o de rendimiento del alumnado. Por su parte, el término *appraisal*, incorpora también como el anterior el matiz de medición pero, en esta ocasión, refiriéndolo al rendimiento del personal. Si esto es bien conocido para quien esté familiarizado con el campo de la evaluación educativa, no lo es tanto para quienes proceden de otras parcelas de la evaluación y mucho menos para quienes se acercan a la evaluación desde el análisis de políticas públicas y la Ciencia Política. Este hecho ha contribuido, en muchas ocasiones y de manera obviamente errónea, a reducir el término *evaluación educativa* a la mera “evaluación” del alumno, vaciándolo así parcialmente de contenido.

El segundo de los factores está relacionado con la preponderancia del enfoque clásico y *cientifista* de evaluación –que en sus inicios propusiera Ralph Tyler (1950; 1967)– en la evaluación en general y en la educativa muy en particular. Según este enfoque, evaluar consiste en comprobar el grado de consecución de los objetivos previamente definidos por las instituciones por parte de alumnos, profesores, centros y servicios y hacerlo, además, sobre la base del uso de indicadores y estándares integrados en tests y otras pruebas de carácter cuantitativo.

Pese a su preeminencia (o quizás precisamente por ello), el enfoque positivista de evaluación por objetivos, ha sido objeto de numerosas críticas. Cuatro son las principales objeciones que se han planteado a este enfoque⁸:

- (i) Su visión reduccionista, que lleva a someter a evaluación sólo a los elementos terminales del sistema; resulta obvio, como apunta Santos Guerra (1999) que, en caso de error o de incumplimiento serán aquéllos quienes tendrán que cambiar para acomodarse a los objetivos y estándares previamente establecidos, pero nada tendrá que modificar la institución.
- (ii) Su consideración acrítica de los objetivos de la política o el programa, tal cual son formulados por las autoridades educativas, como referentes para juzgar su calidad.
- (iii) El análisis del programa sin tener en cuenta el contexto en el que éste se inserta.
- (iv) La debilidad de las técnicas y métodos utilizados (fundamentalmente tests) para captar la complejidad de los fenómenos educativos.

El predominio de esta visión neopositivista ha sido tal que se ha llegado a producir una suerte de metonimia en la que se confunde una parte (la medición) con el

⁸ Las críticas han procedido, fundamentalmente, de aquellos autores que han desarrollado propuestas alternativas a este enfoque. A este respecto, pueden consultarse las obras citadas en la nota a pie núm. 10.

todo (la evaluación). Además, y de forma pretendidamente neutral y objetiva, se evalúan únicamente los eslabones finales de la cadena: alumnos, profesores, enseñanzas, centros y servicios (Angulo Rasco, 1995; 1996). La afirmación de Alejandro Tiana en este mismo número a propósito de los estudios internacionales de evaluación resume perfectamente la debilidad de los sistemas de medición tan generalizados hoy en el ámbito educativo: “su potencia explicativa se ve limitada sobre todo por la perspectiva macroscópica que adoptan, que les hace perder detalles importantes, por la mirada poco contextualizada que necesariamente deben arrojar (...) y por la estandarización que imponen sus métodos y técnicas de análisis”⁹.

Por todo ello, los intentos de evaluar políticas o programas educativos desde la evaluación educativa han tendido a verse, desde fuera, con cierto recelo. La explicación no es otra que la inadecuada conceptualización de los términos “política” y “programa” de la que parte el enfoque *tyleriano* y que implica considerar que la calidad de las políticas y los programas educativos se determina esencialmente a partir del logro, por el alumno, de los objetivos académicos marcados por los poderes públicos. Ni que decir tiene que esto no deja de ser sólo una suposición. No toda evaluación educativa es (ni tiene por qué ser) evaluación del rendimiento. De hecho, la mayoría de las principales alternativas al enfoque clásico han procedido, precisamente, del ámbito de la evaluación educativa¹⁰.

En las últimas décadas, el enfoque clásico se ha ido renovando, integrado en discursos en favor de la gestión y la evaluación por resultados o de la implantación de sistemas de calidad y de evaluación estandarizados (lo que equivale a decir descontextualizados, como sugiere en esta misma obra Saville Kushner). Así, lejos de debilitarse, ha cobrado nueva fuerza, por ejemplo en el ámbito educativo universitario, al hilo del llamado *proceso de Bolonia*, imponiendo al final una cierta lógica *administrativista* o *certificacionista* en los sistemas de evaluación (Fernández *et al.* en este mismo número). Hasta el momento, las críticas vertidas hacia la implantación de sistemas de indicadores como únicas herramientas para “evaluar” enseñanzas, profesores y centros (Rodríguez Sabiote y Gutiérrez Pérez 2003; Fueyo 2004; Duro y Niremberg en este número), no parecen haber logrado frenar esa tendencia. Cabría pensar que ello se debe a que, en el caso de la educación superior, la evaluación no ha alcanzado aún el estadio más avanzado de desarrollo, no ha llegado a la “cuarta generación”, por emplear la conocidísima expresión de Guba y Lincoln (1989). Sin embargo, a juicio de algunos autores como Ryan y Feller (2009), Santos Guerra (1999), Kushner y Norris (2007) o Schwandt (2007), parece que este hecho está también ligado a la supremacía de una visión *gerencialista* y neoliberal acerca de cómo debe funcionar lo público.

Ahora bien, no queremos decir, con todo ello, que el uso de indicadores y estándares deba descartarse definitivamente o que la medición no pueda aportar informa-

9 Véase página 62.

10 Entre otras, las de MacDonald (1977), House, (1980; 1993; 2006), House y Howe (1999), Stake (1983; 2004), Guba y Lincoln (1989), Greene (2001); o Kushner (2000; 2007).

ción útil a ciudadanos, decisores y gestores. Decimos, simplemente, que la complejidad de los objetos que se evalúan hace imprescindible complementar la perspectiva que aportan los indicadores con otra más atenta a las particularidades del objeto y al contexto en el que se enmarca, como muy acertadamente ha puesto de relieve el profesor Stake (1983, 2004 y también su aportación en este número).

2. La Ciencia Política y, más específicamente el análisis y la evaluación de políticas públicas, pueden hacer notables aportaciones a la evaluación educativa:

(i) De entrada, permite una más precisa y rigurosa “identificación” (o “construcción”) de las políticas o programas objeto de evaluación. Por razones obvias, ninguna otra disciplina (y subdisciplina) está en condiciones de ofrecer el instrumental teórico y metodológico necesario para analizar y evaluar esos objetos tan genuinamente políticos que son los programas y las políticas públicas¹¹. Ya hemos señalado antes la importancia de adecuar la evaluación (sea en el plano teórico, sea en su vertiente aplicada) al objeto que se evalúa. El evaluador ha de ser en todo momento consciente de lo que de común tienen, por ejemplo, la evaluación de centros y la de programas pero también de los muchos elementos que las separan. Incluso entre dos tipos próximos como la evaluación de políticas y la evaluación de programas, las diferencias son considerables, como muy oportunamente destaca en esta misma obra M.^a Luisa Merino. Ciencia Política y *policy análisis* pueden contribuir enormemente a la hora de identificar y definir los contornos y los contenidos de la política pública que se somete a evaluación, toda vez que aquella, como tantas veces se ha dicho, no es un fenómeno autodefinido sino una categoría de análisis a la que recurrimos para abordar el estudio de los resultados de la acción pública (Hecló, 1972).

(ii) Nos sirve también para advertir la relevancia que adquieren las ideas en las políticas educativas. Pocos ámbitos hay tan sensibles desde el punto de vista ideológico y político como el educativo: ni la educación ni la práctica educativa son neutras. Los escenarios de las políticas educativas están presididos por la diversidad de opiniones, valores y creencias; el debate, la argumentación y la confrontación entre actores se tornan en habituales, precisamente porque no cabe (sólo) hallar soluciones científicamente verdaderas. Por ello, resultan de especial utilidad para el análisis de las políticas educativas aquellos enfoques que subrayan el papel de las ideas como variables explicativas del cambio en las políticas públicas¹². El potencial analítico de dichos enfoques va, incluso, más allá al pretender elaborar una teoría global del proceso de políticas públicas desde la Ciencia Política. Esas teorías, aplicadas en nuestro caso al estudio de las políticas educativas, nos permitirían comprender “cómo los actores políticos interesados interactúan dentro de las instituciones políticas para producir, implementar, evaluar y revisar las políticas públicas” (Schlager y Blomquist, 1996: 653).

11 En la medida en que, como recuerda Gomà (2002), en ellos se dirime la regulación del conflicto social.

12 Sobre ellos, consúltese Martínón Quintero (2006).

(iii) Proporciona el instrumental analítico adecuado para comprender las evaluaciones en tanto que procesos eminentemente políticos. La Ciencia Política y, en particular, los estudios sobre actores, procesos de toma de decisiones o políticas públicas resultan ser, para ese propósito, una referencia inexcusable. La evaluación, como apuntara Thomas Schwandt, “es una práctica tanto política como científica” (2005: 5). La política no es una amenaza para la evaluación, el *lado de la subjetividad* frente a la dimensión técnico-científica en la que reinan los hechos objetivos y las explicaciones contrastadas, por más que en ciertos momentos así se haya considerado. Y ello es así desde el momento en que en ella “se ven involucrados actores con intereses contrapuestos, políticamente relevantes, que luchan por obtener una mayor cuota de poder e influencia en el proceso evaluativo, con el propósito de modularlo de acuerdo a sus preferencias” (García Sánchez, 2009). Los estudios sobre evaluación y los estudios de metaevaluación no pueden dejar de reconocer y considerar este hecho: la política es un elemento consustancial a la evaluación y, de forma más clara aún, como hemos visto, a la evaluación educativa, “contagiada” de la enorme carga política asociada al mundo educativo.

(iv) Constituye una excelente guía a la hora de abordar el análisis de la política de evaluación de los poderes públicos, en parte por las razones antes apuntadas. Y es que, como nos recuerda Santos Guerra (1999), la cuestión de cómo concibamos la evaluación y de cómo evaluemos no es ni mucho menos baladí. Nuestra idea sobre cómo tiene que ser la evaluación y a qué propósitos ha de orientarse, dice mucho sobre cuál es nuestra visión de la educación y de cuáles creemos que deben ser los fines últimos a los que debe servir. El análisis de cuestiones tales como qué objetos se someten a evaluación, qué actores participan en el proceso evaluativo, a qué enfoques y técnicas se recurre, de quién es la propiedad de la información producida o de qué forma se difunden y/o publican los informes de evaluación, nos permite saber si los poderes públicos intentan ejercer algún tipo de control sobre los procesos de evaluación o si, por el contrario, lo conciben e incorporan como un mecanismo más para la rendición de cuentas a la sociedad¹³. Son cada día más las voces que reclaman la necesidad de convertir a la evaluación en un instrumento para la rendición de cuentas y el fortalecimiento de la democracia, para la determinación de la eficacia social además de para calibrar la eficacia operativa de las políticas (House, 2006; Stake, 2006; MacDonald y Kushner, 2005; Greene, 2001, 2003).

Puede que el lector encuentre otros espacios de convergencia, además de los que aquí se han mencionado, o considere que los puntos de encuentro propuestos no parecen posibles. Si ello es así, nos habremos aproximado a uno de los objetivos que nos marcábamos al comienzo de esta presentación: abrir el debate en torno a las posibilidades de acercamiento entre la evaluación educativa, de un lado, y la Ciencia Política y el análisis y evaluación de políticas públicas, de otro.

¹³ Desde los estudios sobre rendición de cuentas también se ha destacado el papel que puede cumplir, en este sentido, la evaluación. Véanse Borins (1995) e Isunza Vera y Olvera (2006).

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Angulo Rasco, J.F. (1995), “Evaluación del sistema educativo. Algunas respuestas críticas al porqué y al cómo” en P. Manzano, coord., *Volver a pensar la educación*, vol. II. La Coruña / Madrid: Fundación Paideia / Morata.
- Angulo Rasco, J.F. (1996), “Ningún paso adelante; dos pasos atrás. La desaparición de la evaluación institucional”, *Investigación en la escuela*, 30: 63-73
- Bañón i Martínez, R. (2003), comp., *La evaluación de la acción y de las políticas públicas*. Madrid: Díaz de Santos.
- Borins, S. (1995), “The New Public Management Is Here to Stay”, en *Canadian Public Administration*, 38(1).
- Fueyo Gutiérrez, A. (2004), “Evaluación de titulaciones, centros y profesorado en el proceso de Convergencia Europea: ¿de qué calidad y de qué evaluación hablamos?”, *Revista Interuniversitaria de Formación del Profesorado*, 18(3): 207-219.
- García Sánchez, E. (2009), “Metaevaluación de políticas públicas: una visión desde la Ciencia Política”, *Revista del CLAD. Reforma y Democracia*, 43:129-154.
- Gomà Carmona, R. (2002), “Prólogo”, en M. Grau y A. Mateos, eds., *Análisis de políticas públicas en España: enfoques y casos*. Valencia: Tirant lo Blanch.
- Greene, J.C. (2001), “Dialogue in Evaluation: A Relational Perspective”, *Evaluation*, 7(2): 181-187.
- (2003), “War and Peace... and Evaluation”, *Studies in Educational Policy and Educational Philosophy*, 2, Universidad de Uppsala. En línea: <http://www.upi.artisan.se/docs/Doc186.pdf> (consulta: 1 septiembre 2009)
- Guba, E.G. y Lincoln, Y.S. (1989), *Fourth Generation Evaluation*. Newbury Park, Ca: Sage.
- Hecló, H. (1972), “Policy Analysis”, *British Journal of Political Science*, 2: 83-108.
- House, E.R. (1980), *Evaluating with Validity*. Beverly Hills, Ca: Sage (traducción al castellano en Morata con el título *Evaluación: ética y poder*, Madrid, 2002).
- (1993), *Professional Evaluation. Social Impact and Political Consequences*. Newbury Park, Ca: Sage.
- (2006), “Democracy and Evaluation”, *Evaluation*, 12(1): 119-127.
- y Howe, K.R. (1999), *Values in Evaluation and Social Research*. Thousand Oaks: Sage.
- Isunza Vera, E. y Olvera, A., coords. (2006), *Democratización, rendición de cuentas y sociedad civil: participación ciudadana y control social*. México D.F.: Porrúa, CIE-SAS, Universidad Veracruzana y Cámara de Diputados.
- Joint Committee on Standards for Educational Evaluation (1988), *The Personnel Evaluation Standards*. Thousand Oaks, Ca: Corwin Press.
- (1994), *The Program Evaluation Standards*. Thousand Oaks: Sage.
- (2003), *The Student Evaluation Standards*. Thousand Oaks: Sage.
- Kellaghan, T., Stufflebeam, D.L. y Wingate, L.A. (2003), “Introduction”, en T. Kellaghan y D.L. Stufflebeam, eds., *International Handbook of Educational Evaluation*. Dordrecht, Holanda: Kluwer.
- Kushner, S. (2000), *Personalising Evaluation*. Londres: Sage.
- Kushner, S. y Norris, N. (2007), “The New Public Management and Evaluation”, en S. Kushner y N. Norris, eds., *Dilemmas of Engagement. Evaluation and the New Public Management*. Oxford: Elsevier.

- MacDonald, B. (1977), “A political classification of evaluation studies”, en D. Hamilton, ed., *Beyond the Numbers Game*. Londres: MacMillan.
- MacDonald, B., y Kushner, S. (2005), “Democratic Evaluation” en S. Mathison, ed., *Encyclopedia of Evaluation*. Thousand Oaks, Ca: Sage.
- Martinón Quintero, R. (2006), *Las políticas de drogas en España a través del marco de las coaliciones promotoras. Un estudio del cambio en las políticas públicas*. Madrid: Universidad Carlos III, tesis doctoral.
- Nevo, D. (2006), “Evaluation in education”, en I.F. Shaw, J.C. Greene y M.M. Mark, eds., *Handbook of Evaluation*. Londres: Sage.
- Patton, M. Q. (1997), *Utilization-focused evaluation: The new century text*. Thousand Oaks, Ca: Sage (3.a edición)
- Rodríguez Sabiote, C. y Gutiérrez Pérez, J. (2003), “Debilidades de la evaluación de la calidad en la universidad española. Causas, consecuencias y propuestas de mejora”, *Revista Electrónica de Investigación Educativa*, 5(1): 1-26.
- Ryan, K.E. y Feller, I. (2009), “Evaluation, Accountability and Performance Measurement in National Education Systems: Trends. Methods and Issues”, en K.E. Ryan y J.B. Cousins, ed., *The Sage International Handbook of Educational Evaluation*. Thousand Oaks, Ca: Sage.
- Santos Guerra, M.A. (1999), “Sentido y finalidad de la evaluación en la universidad”, *Revista Interuniversitaria de Formación del Profesorado*, 34, enero-abril: 39-59.
- Schlager, E. y Blomquist, W. (1996), “A comparison of three emerging theories of the policy process”, *Political Research Quarterly* 49(3): 651-672.
- Schwandt, T.A. (2005), “Politics of Evaluation”, en S. Mathison, ed., *Encyclopedia of Evaluation*. Thousand Oaks, Ca: Sage.
- (2009), “Globalizing Influences on the Western Evaluation Imaginary”, en K.E. Ryan y J.B. Cousins, ed., *The Sage International Handbook of Educational Evaluation*. Thousand Oaks, Ca: Sage.
- Scriven, M. (1972), “Prons and cons about goal-free evaluation. Evaluation Comment”, *The Journal of Educational Evaluation*, 3(4): 1-7.
- Stake, R.E. (1983), “Program evaluation, particularly responsive evaluation”, en G.F. Madaus, M. Scriven y D.L. Stufflebeam, eds., *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer Nijhoff.
- (2004), *Standards based and responsive evaluation*. Thousand Oaks, Ca: Sage.
- Tyler, R.W. (1950), *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.
- Tyler, R.W. et al. (1967), “Perspectives of Curriculum Evaluation”, *AERA Monograph Series on Curriculum Evaluation*, 1, Chicago: Rand McNally.
- Vanaclocha, F.J., García Sánchez, E. y Viñas, V. (2005), “Algunas referencias sobre metodología de la evaluación de las políticas y los servicios públicos, en J.A. Garde, coord., *La Agencia Estatal de Evaluación de la Calidad de los Servicios y de las Políticas Públicas. Reflexiones y propuestas de creación. Informe de la Comisión de Expertos*. Madrid: Ministerio de Administraciones Publicas / Instituto Nacional de Administración Pública.
- Worthen, B.R. y Sanders, J.R., comps. (1973), *Educational Evaluation: Theory and Practice*. Worthington, Ohio, Ca: Jones Pub. Co.

G^a_pP

ESTUDIOS

El desarrollo de centros de evaluación y la formación de evaluadores¹

Robert E. Stake

Resumen: *La práctica de la evaluación, en tanto que actividad formal, se ha caracterizado y se caracteriza aún por tres elementos clave: (i) el predominio del razonamiento basado en criterios y estándares, (ii) la apuesta por la utilización de tests y otras pruebas cuantitativas y (iii) la consideración de que la calidad es un atributo intrínseco de los programas. En la misma línea, la formación que se imparte desde los centros de evaluación se ha centrado principalmente, en la enseñanza de modelos y enfoques metodológicos. Esta perspectiva presenta importantes limitaciones a la hora de determinar la eficacia de un sistema educativo. Se hace, pues, necesario recuperar otra clase de razonamiento que, con frecuencia ha sido denostado: el conocimiento episódico, que insiste en que la calidad de un programa está íntimamente vinculada a las percepciones y a las experiencias del evaluador acerca del mismo y que sitúa al programa en su contexto. Ambos tipos de razonamiento son necesarios; su combinación e integración como si de las dos caras de una moneda se tratase, la “visión binocular”, vendría a convertirse en la mejor apuesta para el evaluador y para la evaluación.*

Palabras clave: *Evaluación basada en estándares, evaluación comprensiva, centros de evaluación, formación en evaluación, visión binocular.*

1. LOS CENTROS DE EVALUACIÓN

El primer tema que abordaré es el desarrollo de los centros de evaluación. A continuación, la formación de los evaluadores y, finalmente, hablaré sobre el concepto de calidad en la escuela. Al principio, creí que se trataba de un tema meramente técnico pero, tras reflexionar, me di cuenta de que pensar en el desarrollo de los

¹ Texto de la conferencia pronunciada por el autor en el *Simposio Internacional de Evaluación* que tuvo lugar en la Universidad Nacional Autónoma de México (México D.F.) el 25 de octubre de 2007. El texto ha sido revisado por evaluadores anónimos a los que se informó de que se trataba de una conferencia.

Revisión científica de la traducción a cargo de Ester García Sánchez.

centros nos hace pensar en todo el ámbito de la evaluación educativa. Dicho ámbito incluye:

- La búsqueda de la calidad y las diferencias en el modo de entenderla.
- La valoración de resultados, en especial los resultados del rendimiento de los alumnos.
- La naturaleza política de la evaluación.
- El clima de competitividad de las instituciones, que se esfuerzan por desarrollar una imagen de calidad.
- El carácter global de la educación, incluido el trabajo a través de internet.

Antes de adentrarme en ese vasto territorio, debo añadir unas palabras sobre mi persona. Procedo del sector cualitativo de la evaluación. Fui director adjunto del *Illinois State Testing Program* pero, decepcionado con la evaluación de programas basada en tests, decidí buscar otras cosas. Llegué a la conclusión de que los evaluadores necesitan prestar atención a las experiencias de alumnos, profesores y de otras personas para comprender la dinámica, la complejidad y la calidad de la educación. Así que cuando enseño y cuando recojo información, doy preferencia a los diálogos, a las situaciones y a los contextos. Esto es un ejemplo de ello:

Madre de gemelos: Ha enseñado a Pepe a usar la cámara, pero no a Pepita.

Profesor: Pero, Sra. Martín, Pepita tenía que reforzar las matemáticas.

Madre: Eso no es justo.

Profesor: Aprecio a los dos, no quiero ser injusto.

Madre: Pepita odia las matemáticas. ¿Es que no puede hacer que a Pepita le guste el cole tanto como a Pepe?

Profesor: Pero son niños muy distintos.

Madre: Su obligación es lograr lo mismo con los dos.

Al escuchar esta breve conversación, nos paramos a pensar tanto en la filosofía educativa como en la práctica docente. ¿Se puede entender mejor la igualdad de oportunidades en educación a través de una serie de conversaciones como ésta? ¿No es obligación de los evaluadores educativos examinar cuestiones tan complejas como la igualdad de oportunidades?

Entiendo la evaluación como la tarea de búsqueda de la calidad (lo que también implica detectar la falta de calidad, claro). Al final de este trabajo hablaré de la cali-

dad en la escuela. Pero resida donde resida, me gustaría examinar algunas de las dimensiones de la calidad, a veces para representarlas recurriendo a mediciones, pero también para hablar de los episodios que revelan calidad. Dimensiones y situaciones, cuantitativo y cualitativo, son métodos que se mezclan, sí, pero, en mi opinión, los métodos cualitativos son especialmente valiosos para ayudarnos a entender la calidad educativa. Esta manía mía será una idea recurrente en todo lo que hoy exponga.

1.1. El centro de evaluación típico

Consideremos los centros de evaluación universitarios. Éstas son cinco posibles características:

- El personal del centro contará con uno o dos profesores universitarios y la ayuda de estudiantes de doctorado.
- Es más una tienda de artesanía que un centro de informática.
- El director ha de ser una persona dispuesta a trabajar muchísimo.
- Rara vez la financiación es suficiente como para que la calidad de la investigación esté a la altura de la que se realiza en otros centros de la universidad.
- El centro se ajustará a las peculiaridades de su entorno, de las instalaciones, el mercado, la entidad financiadora y las estructuras educativas de la zona.

Algunos centros no presentarán estas características, por supuesto.

Pensemos en una de las formas habituales de concebir el trabajo de un centro de evaluación. Todos conocemos el principio universitario que considera tres dimensiones en el trabajo de los profesores universitarios: docencia, investigación y gestión. Gestión entendida como colaboración con su propio departamento, con la profesión y con la comunidad. Veamos lo que hace un centro de evaluación:

- Los centros participan en la docencia de los estudiantes de su universidad, tanto en materia de destrezas evaluadoras como sobre la profesión de la evaluación y, a veces, también ofrecen formación al público en general.
- Los centros investigan, en ocasiones a través de contratos, tanto sobre la teoría como sobre la práctica de la evaluación educativa.
- Los centros ofrecen servicios de evaluación, algunas veces de forma gratuita, a través de consultas, asistencia en proyectos, mejora y desarrollo de programas.

En definitiva, los centros tienen más trabajo del que pueden asumir. Cada centro es diferente y elige entre las oportunidades que se le presentan; su tarea es encontrar su señal de identidad.

1.2. Todos los organismos evalúan

Según explicó magistralmente Michael Scriven (1996), la evaluación es un proceso humano fundamental. La mayor parte de la evaluación es informal y utiliza la intuición y el sentido común, pero muchas veces es formal. No existe una línea divisoria clara entre formal e informal. La evaluación formal es reflexiva, se basa más en la recogida de datos, es más disciplinada.

En el entorno universitario y en todos los organismos se dan muchas situaciones en las que hay evaluación. La mayor parte del trabajo evaluativo no se denomina *evaluación*, así que la mayoría de la gente no aprovecha la experiencia de los centros que evalúan. Independientemente de cómo lo denomine, el evaluador/a busca la calidad incorporando, de un modo u otro, definiciones tanto particulares como universales. Consideremos algunos entornos de trabajo:

- En facultades e instituciones u organismos educativos, se evalúa el éxito de alumnos y profesores.
- En las facultades de medicina y en los hospitales, se evalúan la salud y los tratamientos.
- En las facultades de derecho, los alumnos aprenden a evaluar lo que sucede en función de si es legal o ilegal.
- En las escuelas y bases militares, los profesionales de todos los niveles evalúan las estrategias y las tácticas.
- En las agencias de inteligencia, se evalúan las amenazas a la seguridad y los riesgos.
- En las agencias reguladoras, se evalúan los riesgos, las medidas de protección y el cumplimiento de normas y reglamentos.
- En las escuelas de negocios y en las empresas, se evalúan los procesos, la gestión, las ventas y la publicidad.
- En las facultades de periodismo, se enseña periodismo de investigación, una tarea con un componente evaluativo muy elevado (aunque los medios rara vez los contratan).

Hay muchos entornos laborales, todos ellos impregnados de evaluación: la ingeniería, los ministerios, las artes y la música; entornos de todo tipo.

La mayoría da por hecho que el término *evaluación* se refiere a algo técnico o relacionado con la gestión. De algún modo, la mayoría de la gente que trabaja en las universidades ha llegado a la conclusión de que una formación general en evaluación carece de importancia. Sin embargo, es necesario trabajar en todos estos ámbitos, si los que forman los centros de evaluación desean ampliar su campo de actuación.

Muchos ámbitos necesitan de la evaluación. Los centros evaluadores compiten con otras empresas de servicios a la hora de conseguir contratos. En ocasiones, la competencia es feroz. Algunos competidores ofrecen precios muy bajos a cambio de muchas promesas, incluso de la promesa de que las evaluaciones² serán más rigurosas que las que han llevado a cabo anteriormente. A veces, he perdido contratos contra otras empresas que luego han tenido que recurrir a mí para averiguar cómo hacer el trabajo.

Hace cuarenta años, en Estados Unidos la mayoría de los programas educativos que contaban con financiación externa eran evaluados por centros pertenecientes a universidades. El personal de estos centros era gente amable y competente. Sin embargo, al aumentar los presupuestos destinados a evaluación y al cobrar mayor importancia las consecuencias la competencia aumentó y se perdieron las formas.

Veamos quiénes ofrecen, en la actualidad, servicios de evaluación en muchos países. Figuran aquí seis ámbitos en los que los evaluadores hacen evaluación formal:

- Grandes empresas de evaluación tales como American Institutes for Research, Abt Associates o Rand Corporation.
- Departamentos o unidades que llevan a cabo programas de evaluación interna en organismos gubernamentales y grandes empresas, tales como el General Accounting Office o los departamentos de evaluación del Banco Mundial.
- Pequeñas empresas de evaluación, a menudo en manos de alguien que antes trabajaba para una de las grandes.
- Pequeños centros de evaluación. Uno de los pequeños se hizo grande, el de Dan Stufflebeam en la Western Michigan University.
- Muchas personas, con poca o ninguna formación en evaluación, pero que son referentes en su profesión y están dispuestas a ser evaluadores en su campo.

² En el original *assessments*.

- Y, por último, otro grupo de personas de cierto prestigio y dispuestas a participar en paneles de expertos, grupos de crisis o comités especiales de investigación, gente que no ha oído en su vida el término “objeto de evaluación” y para los que es un honor que se les pida ayuda para estudiar un determinado problema y evaluar la situación.

Así que, resumiendo, todo el mundo evalúa, y muchos a cambio de dinero. Supuestamente, los centros de evaluación deberían estar vinculados con muchos de ellos y ofrecer formación y asesoramiento a muchos otros.

Actualmente, en los campus técnicamente más avanzados de los países ricos existen, a mi modo de ver, posibilidades de establecer centros de evaluación. Sin embargo, no tengo noticia de que haya mucha demanda o de que se advierta su necesidad.

Generalmente, se supone que estos centros universitarios tendrán valores similares a los de las disciplinas académicas, es decir, las ciencias sociales y las físicas; facultades que buscan la verdad. No se espera que sean especialmente comprensivas con la gente corriente que tiene problemas en su vida cotidiana.

La gran mayoría de los responsables de los programas, cuando encargan a un centro la realización de una evaluación, buscan más un aval a lo que están haciendo que comprenderlo. Buscan el reconocimiento más que el conocimiento.

Y la mayor parte de nosotros, como evaluadores, advertimos que necesitan un refrendo y ofrecemos nuestra ayuda, sobre todo si nos pagan por ello. Y hacemos lo posible por encontrar aspectos positivos; pero también queremos señalar las deficiencias o los problemas que hallamos. La cuestión es, ¿tendrá el centro de evaluación el coraje necesario para señalar los problemas? ¿evaluaremos correctamente la falta de calidad? Creo que la gran mayoría de los centros de evaluación de las universidades se sienten menos seguros que los evaluadores independientes y, por ello, son reacios a señalar defectos.

Parte del negocio de los centros educativos está en impartir cursos, formar evaluadores y ayudar a las diferentes disciplinas académicas, a los profesionales y al público en general a comprender que la evaluación formal es una disciplina en sí misma. La evaluación tiene teorías, métodos, prácticas y un código ético. La evaluación informal es una actividad peculiar. En cambio, la evaluación formal es una disciplina.

2. LA FORMACIÓN DE EVALUADORES

Del mismo modo que los centros evaluadores son diferentes y se adaptan a sus circunstancias, la formación de evaluadores se ha amoldado y debe seguir amoldándose a las necesidades de los estudiantes, de los profesores y de las instituciones. Desde mi punto de vista, no sería bueno para nuestra profesión estandarizar más la

práctica de la evaluación. Podemos imaginarnos otro mundo en el que la evaluación sea tan variada, opaca y adaptada a cada contexto que confunda a nuestros clientes y al público; pero no creo que sea el caso en la actualidad. Así que la formación también debe diversificarse.

El *mentoring* y la formación en proyectos deberían servir para diversificar la formación³ de los evaluadores, al igual que la enseñanza en el aula. Se debe dar oportunidad a quien se está formando en evaluación para que adapte los conceptos y la metodología a su ámbito de trabajo.

Muchos cursos de evaluación hacen hincapié en los distintos modelos o enfoques metodológicos. Creo que pensar en estos modelos distrae la atención de quienes se están iniciando en la evaluación, que debiera dirigirse a las grandes cuestiones de la evaluación. Tal y como yo lo entiendo, estas grandes cuestiones son:

- ¿Se hacen siempre en evaluación comparaciones con estándares?
- ¿Debería la evaluación limitarse a las cuestiones planteadas por el cliente?
- ¿Qué es la calidad?
- ¿Se puede vincular estrechamente la evaluación con la mejora de un programa?
- ¿Cuál es la influencia de la política sobre la evaluación?
- ¿Son siempre útiles los estándares profesionales de evaluación?

La formación de evaluadores siempre debería incluir tales cuestiones. Sin embargo, cuando escribí un libro sobre evaluación (*Standards-Based and Responsive Evaluation*) que se ocupaba de estas cuestiones, no tuvo ningún éxito. No entiendo por qué.

En el libro argumentaba también que explicitar los criterios no es un requisito imprescindible de una buena evaluación ya que, a menudo, hemos de dejar que sean las descripciones de los comportamientos y de las actividades, más que las palabras mismas, las que proporcionen las definiciones. En mi opinión, parte del desarrollo de la disciplina consiste en no abusar tanto de las explicaciones y las nomenclaturas.

2.1. La sistematización de la evaluación informal

Una de mis estrategias formativas es recalcar la importancia de la sistematización continua de la evaluación informal más que su sustitución. Si la evaluación infor-

3 *Education* en el original.

mal pudiese hablar, nos diría que es mejor que la evaluación formal porque, gracias a ella, la humanidad ha sobrevivido miles de años. Hasta hace poco, la humanidad sobrevivió sin la evaluación formal. Probablemente intervino la suerte, pero lo que es seguro es que la supervivencia fue posible porque se tomaron decisiones correctas, y acertar en las decisiones es una cuestión evaluativa. Dado que la evaluación formal tal y como la conocemos hoy no existió prácticamente hasta 1950, y que hasta el sistema judicial, los hospitales, las comunidades religiosas y las ciencias tienen unos cuantos siglos de antigüedad, podemos llegar a la conclusión de que ha sido la evaluación informal la que nos ha traído hasta aquí, no la formal.

Para que la evaluación formal sea mejor que la informal, tenemos que estar convencidos de que nos ayuda a reconocer mejor la “bondad” y, desde luego, sin cambiar nuestra idea acerca de qué es lo bueno y sin incurrir en otros efectos colaterales.

Sin embargo, aunque nos convenciéramos de que, habitualmente, sería mejor evaluar nuestros programas de un modo informal, no formal, la gran mayoría de la gente en la actualidad piensa que la evaluación formal es más fiable, que es más probable que sea ésta la que nos proporcione las evidencias en las que basarnos acerca de la calidad y las deficiencias de los programas. Es aconsejable reflexionar sobre dónde es mejor poner el énfasis, si en la evaluación informal o en la formal. Sin embargo, si en un contrato se establece que es necesario evaluar un programa, se está aludiendo a la evaluación formal. Aun así, dentro de la evaluación formal hay cabida para los datos informales, los juicios y las interpretaciones.

Sigo hablando, para mis adentros, de la creación de centros de evaluación y de la formación de evaluadores.

2.2. Funciones y estilos de evaluación

La evaluación pretende obtener conocimiento sobre el valor de una cosa. Buena parte de ese conocimiento proviene de la experiencia personal, a menudo de nuestra propia conciencia. Sé que el Allopurinol es un medicamento bueno para mí porque si no lo tomo se me forman piedras en el riñón. Lo que me dice el médico es importante y puedo leerme la *Enciclopedia de la Salud* pero, para mí, mi propia experiencia es fundamental. No puedo fiarme siempre de la experiencia pero la utilizo constantemente; la de las personas que conozco y la experiencia indirecta, la reflexión y el análisis de personas a las que no conozco. Reúno todas esas experiencias y conocimientos y, combinando intuición y precisión, hago una valoración de conjunto. También me ayuda beber mucha agua...

La evaluación trata siempre de determinar el mérito y las debilidades. A veces la evaluación hace mucho más, pero su función esencial consiste en determinar el mérito. Ése es su objetivo primordial, su definición, su razón de ser.

En la sala de espera escuché a una señora contarle a un conocido que era “evaluadora”:

—¿A qué te dedicas?

—Ayudo a la gente a perder peso.

—¿Por qué te haces llamar evaluadora?

—Porque la gente te presta más atención si eres evaluadora en vez de dietista.

Los dietistas también son evaluadores, pero si esta señora se dedica a ayudar a las personas a adelgazar, está confundiendo a la gente al hacerse pasar por evaluadora. ¿Puede cada uno llamarse lo que quiera? Claro que sí. Pero nos vamos a entender mejor si consideramos “evaluador/a” a quien, dentro de un campo determinado de actividad, averigua e informa sobre los méritos y carencias del objeto de evaluación. El valor que se le asigne sirve para muchos propósitos, tales como la mejora de un proceso, la concesión de un premio, la aplicación de medidas especiales de refuerzo educativo o comprobar el cumplimiento de un contrato. Estos son algunos de los muchos modos en que se emplea la evaluación (Scriven, 1967).

2.3. Razonamiento basado en criterios y razonamiento episódico

El diccionario no lo recoge, pero los evaluadores normalmente entendemos por *criterio* un descriptor del objeto de evaluación, como su eficacia, su duración, su coste o su valor estético. Y por *estándar* entendemos la “cantidad” de ese criterio con la que clasificamos al objeto de evaluación en un determinado nivel de calidad. Más jerga, pero útil.

Hace 48 años yo era estudiante de doctorado en un Departamento de Psicología. Un día, sentado a mi mesa de trabajo, me di cuenta de que existen las Ciencias (Sociales) de la Educación porque los psicólogos y sociólogos de la educación lograron reestructurar los fenómenos educativos en forma de variables. Habían inventado los constructos, los “ladrillos” para el análisis sistemático de la educación, y llamaron a esos constructos *variables*. A veces, llamaban a estos mismos constructos descriptivos *atributos, propiedades, cualidades, características, facetas o dimensiones*. En muchas ocasiones los llamaban *criterios*. Al reducir a variables los fenómenos complejos -que se dan en el aula, en una sala de juntas, en la historia o en la comunidad- se podían manejar las cosas. Traducir los hechos a variables se denomina *razonamiento basado en criterios*.

Una variable es un atributo que varía. Puede hacerlo de diferentes maneras, pero los científicos sociales decidieron poner el énfasis en la variación en la cantidad. Se considera que ésta (la cantidad) varía a lo largo de una escala. Así que una vez

identificado el constructo, es decir, la escala, lo importante era medir la cantidad. Podemos utilizar estas cantidades para describir, distribuir, comparar, incluso para que parezca que averiguamos las causas, e interpretar dichas causas como base para el control, la mejora o la reforma. Hace 48 años, me parecía que había descubierto la pólvora. Con criterios, con el razonamiento basado en criterios, podíamos medir, y con escalas de medida, podíamos mover montañas (Stake, 2004).

Gracias al muestreo y al razonamiento basado en criterios, el estudio de la educación podía ser algo preciso, permitir que se obtuvieran generalizaciones y ser útil. Todas mis dudas se disiparon. Me apunté a la ciencia de los tests. Me dediqué en cuerpo y alma a convertirme en un “hombre que mide”, y aún lo soy. Soy un “hombre que mide”, trabajo en evaluación de programas. Intento medir la cantidad y la calidad de la educación o de la formación, de los servicios sociales, como también los puntos fuertes y los fallos, los criterios vagos y difusos de la docencia y el aprendizaje.

La psicología me dice que la alternativa al razonamiento basado en criterios es el razonamiento episódico. Los fenómenos educativos se conocen a través de episodios, sucesos, prototipos, actividades, acontecimientos. Estos episodios tienen una base temporal y contextual. Los habitan personas con su propia personalidad, su historia, sus aspiraciones y sus debilidades. A veces hablamos acerca de la personalidad y de las debilidades, de los contextos y de los episodios, en términos de variables. Casi cualquier cosa puede convertirse en una variable; sin embargo, el discurso basado en variables a menudo simplifica en exceso. Conseguimos las herramientas pero perdemos las situaciones.

Para planificar la formación en evaluación y para establecer los principios éticos de un centro deberíamos ser conscientes de que ambos tipos de razonamiento, el basado en criterios y el episódico, coexisten tanto en nuestra cultura como en nuestro pensamiento. Con una especie de visión binocular, a veces podemos reducir el *gap* y, sin darnos cuenta, conseguir una imagen a la que no se llega únicamente con uno de los dos criterios: logramos profundizar combinando ambos tipos de análisis. Un ejemplo de esta combinación de análisis se encuentra en la evaluación de la calidad escolar.

Dediquemos unos minutos al concepto de *calidad*. Se utiliza habitualmente en dos sentidos. Puede referirse a las características de algo, como por ejemplo de una obra musical o de la forma de enseñar de una determinada persona. Hablar de cualidades es una cuestión más descriptiva que valorativa. En música, se referiría a variables como el lirismo, la sonoridad, el clasicismo o el casticismo. En educación, recurrimos a términos como creativa, convencional, centrada en el niño, dispersa. Para John Dewey, “el propio interés y la simpatía tienen c(u)alidades contrarias”. Estos usos de la palabra *calidad* se refieren a la naturaleza o a los ingredientes de las cosas, no a si son buenas o malas.

La acepción que interesa al evaluador *sí* tiene que ver con la bondad de las cosas. La calidad de la interpretación de una obra musical es su grado de excelencia. La cali-

dad de un episodio educativo son sus méritos y carencias. Éste es el sentido en el que normalmente utilizamos el término en el trabajo de evaluación. Buscamos la calidad de los programas, en el sentido de su mérito y su valor. Pero eso deja muchas cosas en el tintero. Con frecuencia, es difícil concretar y llegar a un acuerdo sobre qué es la calidad, el mérito o la excelencia. Los estándares que reflejamos en nuestras palabras o en indicadores son a menudo más simples, menos complejos que lo que experimentamos personalmente. Cuando decimos que la calidad de una redacción de un alumno es mediocre, tenemos en cuenta muchos aspectos: su coherencia, el tema, la gramática, la creatividad, la originalidad, el hilo temporal, el uso de las palabras, el cumplimiento de la tarea, incluso algunas características en las que no pensamos de antemano (aunque no siempre son las mismas, ni siempre las valoramos de la misma manera en todas las ocasiones). El razonamiento basado en criterios es importante, pero la interpretación también lo es.

Como evaluador, no encuentro necesario explicitar qué calidad busco. Algunos evaluadores hacen lo posible por ser explícitos. Pero yo estoy cansado de no utilizar más que uno o unos cuantos criterios. Pretendo enseñar a mis alumnos a que se familiaricen a través de la experiencia con una serie de aspectos del programa. Algunos evaluadores prefieren dedicar sus recursos a medir el mejor o los mejores criterios.

Otra cosa que nos diferencia a los evaluadores es nuestra idea respecto a dónde reside la calidad. Cuando piensas en la calidad de un melón francés, puedes pensar en la bondad como una de sus propiedades. O puedes pensar en su calidad, definida a partir de las opiniones de quienes lo comen. ¿La calidad pertenece al melón o a la experiencia de comérselo? Obsérvese que esto último exige prestar una mayor atención a quien se lo está comiendo. Los evaluadores no se ponen de acuerdo en valorar cuánta atención hay que prestar al “destinatario” del objeto evaluado. Cuando se piensa en la calidad en los tests de rendimiento, se puede considerar como una cualidad de quien se somete a la prueba o como el resultado de la interacción entre quien realiza la prueba y los examinadores. Es parte de la formación.

No hay por qué coincidir conmigo, pero me gusta pensar que la calidad tiene su origen en la experiencia humana. El concepto de calidad no sirve para nada si no se tiene en cuenta la experiencia humana como referente clave. La calidad existe porque las personas la experimentan. Y a través de los años, las experiencias positivas que han contado más han sido quizás la comodidad, la felicidad y el bienestar. Y, desde luego, las experiencias negativas de la falta de comodidades, la ira y miedo. Lo que reconocemos enseguida como una enseñanza de calidad o un melón de calidad tiene sus raíces, en mi opinión, en nuestras experiencias previas. A estas alturas, podemos haber desarrollado estándares formales o informales, convenciones y tradiciones, para valorar los melones y la música, pero las raíces de la idea de calidad están en las experiencias que vamos recordando a lo largo del tiempo. Un reloj es de buena calidad en parte porque marca bien las horas y tiene una maquinaria precisa, pero también porque a la gente le parece muy bueno y superior a otros relojes que hayan conocido. Esto hace que los evaluadores, a medida que nos vamos metiendo en el

negocio de evaluar las cosas, nos damos cuenta de que la calidad depende en gran medida de quién la esté experimentando.

2.4. La calidad en la escuela

Los evaluadores educativos no se ponen de acuerdo acerca de lo que es un buen indicador de la calidad en la escuela. Muchos en Estados Unidos se conforman con las puntuaciones obtenidas por los alumnos. Muchos otros prefieren los sistemas de acreditación e inspección educativa. Todos están de acuerdo en afirmar que la escuela es un entorno complejo y que cualquier indicador por sí solo simplificaría en gran medida el concepto de calidad en la escuela. Sin embargo, existe una gran presión a nivel mundial para considerar sólo una de las dimensiones de la calidad.

La medición de la calidad se vuelve extraordinariamente difícil por el hecho de que distintas personas buscan cosas diferentes en las escuelas, incluso algunas contradictorias, como que los estudiantes sean más conformistas y que sean más reivindicativos. Muchas políticas escolares consideran que estas diferencias son errores de comunicación más que mecanismos democráticos. La calidad es muy complicada.

Los profesores saben mucho de calidad en la escuela, normalmente más que los gestores, que los evaluadores externos y que los miembros del gobierno. Pero los juicios del profesorado y los episodios más significativos son infravalorados. Y a menudo a aquéllos se les acusa de servir a sus propios intereses individuales, que a veces sucede. Pero un buen proceso de evaluación recurre al razonamiento basado en criterios y al razonamiento episódico de los profesores.

¡Un momento! Estábamos hablando de formación. La formación de los evaluadores debería incorporar una reflexión sobre la necesidad de conocer la calidad general de la escuela. El pensamiento moderno y la ética de la gestión dicen que ello es necesario sencillamente para poder comparar las escuelas. Sin embargo, como evaluador profesional, he llegado al convencimiento de que no se suelen emplear bien los indicadores de calidad institucional. Evaluamos la calidad escolar por cuestiones de control, competitividad, castigo y orgullo. Rara vez nos cuestionamos su necesidad.

Es importante para todos en la escuela reconocer la calidad, ver los problemas y buscar la mejora, pero la mejora viene del estudio de problemas concretos, no haciendo un seguimiento de los resultados de la escuela a lo largo del tiempo. La mejora se consigue a partir del trabajo individual y la investigación-acción y, a veces, del estudio de casos o de la evaluación de los participantes. Quien se esté formando en evaluación debe aprender que no existe una única visión de la calidad en la escuela. La escuela puede muy bien estar formada por personas que tienen ideas muy distintas sobre la calidad escolar. La calidad en la escuela y los logros de los alumnos se pueden entender mejor a través de prolongadas observaciones de la actividad escolar. Consideremos este episodio.

La clase de Ciencias comenzó con la observación del crecimiento de una semilla. Cada estudiante coge un par de envases de cartón de leche, cada uno los suyos, del poyete de la ventana. Hoy tienen que registrar el crecimiento vertical de uno o más brotes. Habían plantado nomeolvides, cosmos, y algo llamado mezcla de plantas perennes. Cada alumno lo registra en un diario que ya han comenzado con anterioridad. El diario está hecho con folios blancos doblados por la mitad para hacer un librito.

Un alumno pregunta: *¿Las plantas crecen más rápido si les hablas?*

Profesora: *Buena pregunta, esto lo han estudiado los científicos. Quizás han considerado que, para que crezcan, es mejor la música que hablarles (pausa). Kevin, la estás regando demasiado.*

Un alumno: *¿Cómo hacen las plantas?*

Profesora: *Eso es algo de lo que hablaremos hoy. Primero comprobemos su crecimiento. Levantad la mano si tenéis un tallo de al menos cuatro centímetros o más... ¿y quién tiene seis centímetros o más? ¿nueve centímetros? (se van levantando varias manos, al final sólo una).*

—*Sylvia, tus brotes no han crecido tanto.*

—*Sí que han crecido.*

—*Vale, ¿cuántos habéis notado cambios desde el lunes pasado? ¿Qué cambios?*

—*Dos brotes más.*

—*Uno ha crecido, antes era más pequeño.*

—*El mío estaba saliendo torcido y ahora está derecho.*

—*A lo mejor tenía poca agua, si ahora lo estás regando más...*

La clase continúa. La profesora les hace hablar sobre comparar el crecimiento. Hablan de medidas. La profesora hace que piensen en un experimento. Diseña uno. Hablan de comparar medidas y de medias. La profesora decide enseñarles a calcular la media, y lo hace mal. Cuando el supervisor de la profesora advierte el error, se disgusta mucho. Un evaluador formularía esta pregunta: “¿Lo que les ha enseñado con la realización del experimento queda invalidado por haberles enseñado mal cómo calcular la media?”.

2.5. La aptitud académica

La formación en evaluación debe adaptarse a cada individuo. Pero en educación, la mayoría de quienes se están formando ha de saber que evaluar las escuelas a partir

de tests de rendimiento estandarizados carece de validez. La mayoría de las evaluaciones escolares en Estados Unidos emplean la aptitud académica como indicador de la calidad de la escuela. Esto es, la aptitud académica de sus estudiantes.

La aptitud académica puede definirse como la predisposición a aprender en la escuela, es decir, a beneficiarse de la experiencia. La aptitud es, en parte, una función de la inteligencia “natural”, de las enormes capacidades que provienen de nuestro código genético, cuando tiene ocasión de desarrollarse en su plenitud. Hay muchas cosas en la vida que contribuyen al desarrollo de las aptitudes, sobre todo en los primeros años de vida. Fomentar las aptitudes es, en gran medida, una tarea de madres y padres, pero los hermanos, los compañeros, los parientes, luego los profesores, orientadores y tutores, además de los grupos sociales y de los compañeros de trabajo y todos los demás, ayudan también al desarrollo de las aptitudes académicas. Nuestros futuros evaluadores lo saben.

Desde 1900 hemos utilizado los tests de inteligencia, creados para medir las capacidades, según su desarrollo en nuestra cultura. En parte porque decir que un determinado grupo es más inteligente que otro puede resultar hiriente, se modificó el nombre de test de inteligencia a test de aptitud académica. Los ítems se modificaron mínimamente, sólo se cambió la etiqueta. La inteligencia es algo real. La aptitud es real. Los logros son reales. Pero las aptitudes y los logros son dos cosas bien distintas.

La presión política y los contratos de evaluación forzaron a los evaluadores a poner el énfasis en la medición de la calidad en la escuela. Estas presiones crean dudas en la opinión pública en cuanto a los logros de las escuelas: las escuelas son caras, no cumplen todo lo que prometen y no satisfacen nuestras expectativas. Se les presiona para que “rindan cuenta de sus actos”. Para medir esto, los evaluadores en los Estados Unidos tomaron los ítems de los tests de aptitudes académicas (antes tests de inteligencia) y empezaron a llamarlos ítems de rendimiento del alumno. Lo que en su día medía las aptitudes ahora mide el rendimiento.

En Estados Unidos hemos empleado tests con millones de estudiantes. Era obligatorio en todos los estados. Se ignoraba que los tests de rendimiento tienen poco valor diagnóstico y poquísima validez externa. A veces, las puntuaciones medias se analizan en algunos debates sobre política educativa. En ocasiones, se sanciona a los que obtienen los peores resultados. Es un mundo de fantasía. Sam Wineburg lo explicaba muy bien en su trabajo *Crazy for history* (2004). También lo hizo Neil Postman en su libro *The End of Education* (1995). Los tests no sirven para determinar la eficacia educativa y, sin embargo, la política federal estadounidense en materia de educación y las de la mayoría de los 50 estados están basadas en gran medida en los resultados (puntuaciones) de las pruebas.

Estas mediciones aluden a los logros, pero no los miden. Incluso así, quizás sirvan para motivar a los niños. No sabemos cuánto bien o cuánto daño les están cau-

sando a los niños pero nosotros, como evaluadores, no podemos hablar como si de ese modo estuviéramos midiendo la calidad escolar o los logros de los alumnos. No lo estamos haciendo.

3. A MODO DE CONCLUSION

La formación de los evaluadores no tiene que ver únicamente con determinar qué rendimiento satisface ciertos estándares de calidad, sino de “experimentar” situaciones de éxito y de fracaso. Más que centrarse en buenos ejemplos de evaluación, los cursos de evaluación de programas, creo, deberían prestar mucha atención a los errores del evaluador, las dificultades y la resolución de problemas.

Los buenos centros de evaluación también deben localizar los problemas, los errores y el comportamiento imitativo. Necesitan desarrollar vías de comunicación para hablar de los problemas, y de la mala calidad, con un público que no quiere hablar de esos temas. Para los que creamos centros de evaluación y formamos evaluadores el reto consiste en hacer que nuestra comunicación esté contextualizada, sea sensible con los problemas y se base en la experiencia.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Margin, S. (1992), *The Wellness Encyclopedia of Food and Nutrition*. Berkeley: University of California.
- Postman, N. (1995), *The End of Education*. Nueva York: Vintage.
- Scriven, M. (1967), “The methodology of evaluation”, en R.E. Stake, ed., *Perspectives of curriculum evaluation*. Chicago: Rand McNally.
- (1996), “The skeleton in the disciplinary closet”, Millercom Lecture, Universidad de Illinois, 23 de abril.
- Stake, R. E. (2004), *Standards-Based and Responsive Evaluation*. Thousand Oaks, Ca: Sage.
- Wineburg, S. (2004), “Crazy for history”, *The Journal of American History*, 90(4): 1401-1414.

En búsqueda de la explicación de los resultados educativos: posibilidades y limitaciones de los estudios internacionales de evaluación

Alejandro Tiana Ferrer

Resumen: *Los estudios internacionales de evaluación del rendimiento educativo, que se comenzaron a desarrollar a mediados del siglo XX y han experimentado un auge considerable en las dos últimas décadas, han perseguido un doble propósito. Por una parte, sustentar la comparación rigurosa de los resultados obtenidos por los estudiantes en distintos sistemas educativos, ofreciendo como resultado un valioso conjunto de datos que han contribuido a mejorar nuestro conocimiento acerca del rendimiento educativo. Por otra, aportar elementos para explicar los resultados obtenidos, en función de las variables que inciden en ellos.*

En este trabajo se analiza el segundo de dichos propósitos, estudiando las posibilidades que dichos estudios abren y sus aportaciones a tal explicación, así como las dificultades encontradas y las limitaciones que ofrecen para explicar el efecto de los diferentes factores identificados. Para ello se distinguen dos grandes tipos de variables, unas denominadas extrínsecas, relacionadas con factores externos al sistema educativo, y otras intrínsecas, relativas a su organización y funcionamiento. Dada la perspectiva macroscópica adoptada en tales estudios, se concluye que la información que aportan a dicha explicación resulta forzosamente limitada, debiendo complementarse por otro tipo de estudios de investigación y evaluación capaces de permitir análisis más detallados.

Palabras clave: *Educación comparada, evaluación del rendimiento, evaluación a gran escala, índice socioeconómico y cultural (ISEC), eficacia escolar.*

INTRODUCCIÓN

Desde la última década del siglo XX asistimos a una preocupación creciente por la mejora de la calidad de la educación, tanto en los países desarrollados como en los que están en vías de serlo. Factores tales como la presión económica internacional, la expansión del proceso de globalización, la introducción masiva de las nuevas tecno-

logías de la información y la comunicación o la necesidad acuciante de introducir cambios en los modelos productivos en la actual época de crisis económica han vuelto a atraer la atención hacia la educación, considerada nuevamente como un factor decisivo para asegurar el desarrollo de los pueblos.

La celebración en 1990 de una relevante reunión de los ministros de Educación de los países miembros de la *Organisation for Economic Co-operation and Development* (OECD, en castellano OCDE), con el lema de *Una educación y una formación de calidad para todos* (OECD, 1992), supuso la puesta en marcha de numerosas iniciativas con ese propósito, que desde entonces no han cesado de extenderse. Aunque no se puede ignorar que el término *calidad* es polisémico e impreciso, se ha convertido en uno de los argumentos clave (incluso en un *fetiché*, podría decirse) de las políticas educativas contemporáneas (Tiana, 2006).

En este contexto de renovada preocupación por la educación y su calidad, se ha difundido durante las dos últimas décadas el interés por evaluar y comparar la situación y los resultados de los sistemas educativos y han aparecido nuevas iniciativas en ese sentido, entre las que destaca la construcción de sistemas de indicadores de la educación y la realización de ambiciosos estudios internacionales de evaluación del rendimiento educativo. Una de las manifestaciones más patentes del avance registrado se encuentra en el impacto producido por la publicación de los resultados de los sucesivos ciclos del proyecto *Program for International Student Assessment* (PISA), que ha suscitado importantes debates públicos en muchos y muy diversos países (Kotthoff y Pereyra, 2009).

1. COMPARAR Y EXPLICAR LOS RESULTADOS DE LA EDUCACIÓN MEDIANTE ESTUDIOS INTERNACIONALES

En realidad, el desarrollo de esos estudios representa la consolidación de una tendencia anterior, que no se puede considerar estrictamente novedosa. En efecto, hay que recordar que sus antecedentes se remontan al final de los años cincuenta, cuando un grupo de reconocidos investigadores en educación decidió crear una asociación internacional para evaluar los resultados de los sistemas educativos. Así es como se constituyó la International Association for the Evaluation of Educational Achievement (IEA), tras una primera reunión celebrada en 1958 en el Instituto de Educación de la UNESCO en Hamburgo (Tiana, 2001).

La propuesta de aquellos investigadores consistía en planificar y llevar a cabo estudios internacionales del rendimiento educativo adoptando una perspectiva comparada, lo que resultaba realmente novedoso para su tiempo. Desde su punto de vista, los estudios comparativos de los sistemas educativos debían tomar necesariamente en consideración los resultados que conseguían los estudiantes, cosa que no era habitual hasta entonces. Aunque no menospreciaban los análisis basados en variables tales como los recursos disponibles, los procesos de escolarización o las con-

diciones en que se desarrolla la enseñanza, consideraban que la comparación de los sistemas educativos debía completarse con la emisión de juicios rigurosos sobre los resultados alcanzados por los alumnos.

No obstante, aquel grupo de investigadores que constituyó la IEA no quería limitarse a medir y comparar los resultados educativos conseguidos por los estudiantes, sino que pretendía explicarlos de manera convincente. Además de calcular y valorar los resultados, pretendían identificar los principales factores capaces de explicar las diferencias de rendimiento encontradas y analizar de qué modo interactuaban entre sí. Esa explicación constituía, en última instancia, la base para diseñar cualquier acción de mejora de la educación, que era lo que realmente les preocupaba.

En ese contexto surgió la idea de considerar el mundo como un *laboratorio educativo* (Purves, 1993), concepción ciertamente sencilla, pero cuya originalidad hay que reconocer. Su fundamento se encuentra en la explotación de la *variación natural* que existe entre los sistemas educativos. Frente al caso de otros campos científicos, que permiten la aplicación de procedimientos experimentales para estudiar el efecto de determinadas intervenciones, en el ámbito educativo es muy complicado adoptar modelos semejantes e incluso para muchos resulta éticamente discutible. Los investigadores suelen recurrir a otro tipo de métodos, en muchos casos empíricos pero pocas veces de carácter estrictamente experimental. Para salvar esa limitación, los fundadores de la IEA consideraron que podría explorarse con fines comparativos el efecto de las diferencias *naturales* existentes entre los sistemas educativos (aunque, en realidad, puede argumentarse que dichas diferencias derivan en buena medida de su historia y su contexto).

De acuerdo con esa concepción, los sistemas educativos deben ser considerados como conglomerados complejos, compuestos por un número elevado de escuelas, clases, profesores, estudiantes y comunidades. Cada sistema comparte rasgos comunes con otros, pero también presenta algunas características propias. Por lo tanto, sería teóricamente posible comparar los resultados conseguidos por los estudiantes de distintos sistemas educativos, controlando tales variables diferenciales que inciden en el rendimiento. Ese tipo de comparación permitiría analizar y valorar la influencia de dichos factores, contribuyendo a explicar sus efectos respectivos. Así, por ejemplo, podrían valorarse las consecuencias de los diversos modos de diseño y desarrollo del currículo, la integración o segregación de los alumnos en vías académicas alternativas, los modelos de formación del profesorado o la interacción didáctica en el interior del aula, por no citar sino algunos factores frecuentemente estudiados. Sin necesidad de introducir cambios en un sistema educativo determinado, lo que no siempre resulta posible ni conveniente, la existencia de esa variación natural permitiría analizar y valorar el efecto de los principales factores asociados con el rendimiento que no se presentan de igual modo en todos los sistemas. Esa es, en síntesis, la concepción del *mundo como laboratorio educativo* que la IEA difundió entre los investigadores de la educación.

Esa idea directriz sirvió de esquema básico para diseñar y poner en marcha los primeros estudios emprendidos por la IEA, entre los que destacan el primer Estudio de Viabilidad (*Feasibility Study*), de carácter piloto y desarrollado entre 1959 y 1962 en doce sistemas educativos¹; el Primer Estudio Internacional de Matemáticas (*First International Mathematics Study – FIMS*), que se llevó a cabo entre 1959 y 1967; el Estudio de Seis Materias² (*Six-subject Survey*), iniciado en 1966 y finalizado a comienzos de los setenta; y el Segundo Estudio Internacional de Matemáticas (*Second International Mathematics Study – SIMS*), realizado entre 1964 y 1967. Posteriormente, durante los años setenta y ochenta, la IEA desarrolló el Estudio del Entorno Escolar (*Classroom Environment Study*, 1978-1981), el Estudio Internacional de Composición Escrita (*International Study of Achievement in Written Composition*, 1980-1986), y el Segundo Estudio Internacional en Ciencias (*Second International Science Study – SISS*, 1980-1987). Ya en los noventa desarrollaría el estudio sobre Educación Preescolar (*Pre-Primary Education*) e iniciaría los ciclos conocidos respectivamente como Tercer Estudio Internacional de Matemáticas y Ciencias (*Third International Mathematics and Science Study – TIMSS*), Comprensión Lectora (*Reading Literacy Study* y *Progress in Reading Literacy Study – PIRLS*), Tecnologías de la Información en Educación (*Computers in Education* y *Second Information Technologies in Education Study – SITES*), y Educación Cívica (*Civic Education*), que con distintas características y vicisitudes continúan desarrollándose en la actualidad.

Este conjunto de trabajos, que ha dado lugar a una amplia serie de informes y publicaciones (Degenhart, 1990; Papanastasiou, 2004), ha proporcionado una valiosa información sobre el nivel comparativo de rendimiento de los sistemas educativos, por una parte, y sobre las variables asociadas a los resultados, por otra. Para permitir tales logros, la IEA ha aplicado una metodología generalmente cuantitativa, basada, por una parte, en la elaboración de marcos conceptuales rigurosos sustentados en análisis curriculares de las etapas estudiadas y, por otra, en la aplicación de pruebas estandarizadas de rendimiento y cuestionarios dirigidos a profesores, directores escolares y alumnos, construidos cooperativamente con el propósito de conseguir que estén libres de sesgos culturales, idiomáticos o de otro tipo. En ocasiones se han aplicado otros procedimientos, tales como realización de tareas prácticas (*performance tasks*) o técnicas cualitativas (estudios de caso, por ejemplo), pero insistiendo siempre en su carácter estandarizado. La recogida de datos por medio de dichos instrumentos permite aplicar técnicas de análisis rigurosas, respetuosas de las condiciones de una comparación fiable y válida. Con el paso de los años, los procedimientos concretos de recogida y análisis de datos han ido variando, esforzándose siempre por mantener su carácter innovador.

1 La IEA es una asociación voluntaria e independiente, no gubernamental, compuesta por instituciones de investigación que representan a sistemas educativos diferentes. Esa definición hace que países como Bélgica o el Reino Unido tengan una doble representación (Bélgica francesa y flamenca o Inglaterra-Gales y Escocia), aunque la mayoría de sus miembros correspondan a países con un único sistema educativo (incluso en el caso de países federales como Alemania, Estados Unidos, Australia o Canadá). Para más información sobre la IEA y los estudios que desarrolla en la actualidad, se puede consultar su página web (www.iea.nl).

2 Comprendía las materias de Ciencias, Comprensión lectora, Literatura, Francés e Inglés como lenguas extranjeras y Educación Cívica.

Con posterioridad, otros organismos internacionales comenzaron a producir sus propios estudios de rendimiento educativo, aplicando este tipo de procedimientos y logrando avances importantes, tanto conceptuales como metodológicos. Entre ellos hay que destacar especialmente los estudios promovidos por la (OECD). Aunque su interés por los asuntos educativos viene de antiguo (Papadopoulos, 1994), en la década de los noventa puso en marcha un ambicioso proyecto de construcción de indicadores internacionales de la educación, que ha dado importantes frutos (CERI, 1994). Se trata del denominado Proyecto INES, que ha producido la serie de informes que viene publicándose desde 1992 con el título genérico de *Education at a Glance / Regards sur l'éducation*.

Los indicadores elaborados por la OECD se refieren tanto al contexto de la educación como a los recursos humanos y económicos utilizados, los procesos escolares y los resultados logrados. No obstante, esta última dimensión planteó dificultades, dado que los únicos datos existentes en los años noventa eran los procedentes de los estudios de la IEA, que no siempre se ajustaban a las demandas formuladas, ni por el número de países participantes, ni por las áreas evaluadas, ni por su periodicidad. Para salvar esa carencia de indicadores de resultados educativos, la OECD diseñó y puso en marcha a finales de esa década el proyecto denominado PISA, que pretende valorar el grado de formación de los jóvenes de quince años, en tres áreas básicas: matemáticas, lectura y ciencias. La primera recogida de datos se realizó en el año 2000 (OECD, 2001b), repitiéndose a partir de entonces cada tres años, aunque haciendo énfasis sucesivamente en cada una de las áreas mencionadas (OECD, 2004 y 2007).

La OECD también puso en marcha otros estudios orientados a la evaluación de las capacidades de la población adulta (*International Adult Literacy Survey, International Life Skills Survey*), que han producido algunos resultados de interés (OECD & Statistics Canada, 1995 y 2000). Su objetivo consiste en aproximarse a una estimación del capital humano de los países, a través del grado de formación de su población en edad laboral (de 16 a 65 años).

Por su parte, la UNESCO ha emprendido algún proyecto de la misma naturaleza, entre los que destaca especialmente el denominado Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), cuyos trabajos se han desarrollado también en la última década (OREALC, 1998 y 2008).

Ahora bien, toda esta serie de estudios, aun siendo semejantes, tienen sin embargo características distintas, que conviene destacar. Ante todo, hay que distinguir los proyectos directamente orientados a evaluar los resultados conseguidos por los estudiantes (como son el TIMSS, PIRLS o PISA) de aquellos otros que tienen unos objetivos más amplios, como es el caso de la construcción de sistemas de indicadores (Proyecto INES). Si bien es cierto que estos últimos suelen incluir algunos indicadores de resultados, no siempre constituyen su núcleo central, ni consiguen en todos los casos aportar datos originales. Puede decirse que los sistemas de indicadores de la educación elaborados hasta el momento son tributarios del desarrollo experimentado por la estadística educativa, más que de los avances registrados en el campo de la evaluación del rendimiento.

Descendiendo al análisis de los estudios propiamente dichos de evaluación del rendimiento educativo, conviene distinguir al menos cuatro aspectos diferenciales, cada uno de los cuales puede presentarse en forma dicotómica, como un eje lineal con dos extremos, aunque puedan identificarse posiciones intermedias.

- a) Una primera distinción es la que puede establecerse entre los estudios encaminados a la *evaluación del rendimiento escolar* y los orientados a la *estimación del nivel formativo* de una población. Los del primer tipo pretenden evaluar los resultados que logran los alumnos que alcanzan un determinado grado o curso escolar. Es el caso de estudios como TIMSS, centrado en los aprendizajes adquiridos en matemáticas y ciencias por los alumnos de los cursos cuarto y octavo (equivalentes a los diez y catorce años de edad modal) o PIRLS, centrado en el aprendizaje lector logrado al alcanzar el cuarto curso de escolaridad. En ambos estudios, se evalúa a los alumnos matriculados en dicho curso, con independencia de su edad, aunque esa variable se controle y se tenga en cuenta en los análisis. Los estudios del segundo tipo pretenden estimar el nivel formativo alcanzado por las personas que tienen una determinada edad, independientemente de su ubicación escolar. Es el caso de PISA, que evalúa las competencias adquiridas por los jóvenes de quince años de edad en relación con las matemáticas, la lectura y las ciencias, o de IALS o su sucesor PIAAC (Programa para la Evaluación Internacional de las Competencias de los Adultos), cuya población de estudio es la comprendida entre 16 y 65 años. En estos últimos casos, los sujetos evaluados son elegidos al azar entre quienes tienen dicha edad o edades, con independencia del nivel escolar que hayan alcanzado. Mientras que los estudios del primer tipo están más directamente relacionados con la eficacia interna del sistema educativo, los del segundo se orientan a valorar sus resultados finales.

- b) Una segunda distinción es la existente entre los estudios centrados en las *áreas curriculares básicas* y los dedicados a la evaluación de *capacidades o competencias transversales*. Los del primer tipo pretenden medir el grado de dominio de diversas habilidades ligadas a áreas tales como las matemáticas, la lengua materna o extranjera o las ciencias, esto es, a áreas instrumentales básicas del currículo escolar. Es el caso de TIMSS, PIRLS o PISA, antes mencionados. Los estudios del segundo tipo pretenden medir el grado de desarrollo de diversas capacidades no vinculadas a áreas curriculares concretas, sino relacionadas con otras dimensiones formativas, como pueden ser la educación cívica o la utilización de las nuevas tecnologías de la información. Es el caso de los estudios CIVICS o ICCS de la IEA, que evalúan los conocimientos y las actitudes cívicas de los jóvenes y adolescentes, y también del SITES o el componente adicional del PISA, que valoran la familiaridad con la tecnología informática y el uso que los escolares hacen de la misma. Mientras que los estudios del primer tipo están muy relacionados con el currículo escolar, los segundos tienen que ver con un tipo de competencias más amplias y vinculadas con distintos tipos de aprendizajes.

- c) Una tercera distinción es la que contrapone los estudios centrados en la evaluación de las *capacidades o conocimientos adquiridos por los estudiantes individualmente considerados* con los que tienen como objetivo estimar el *rendimiento educativo general de la población escolar*. Los del primer tipo se basan en la aplicación de pruebas idénticas para todos los alumnos evaluados, lo que proporciona la posibilidad de establecer comparaciones o de realizar análisis a partir de los resultados individuales. Su contrapartida consiste en que obligan a limitar la cobertura curricular de las pruebas, para evitar una extensión excesiva, lo que implica reducir la profundidad de los análisis subsiguientes. Los estudios del segundo tipo se basan en la aplicación de pruebas rotadas, respondiendo los alumnos a combinaciones diferentes de cuadernillos parciales, con lo que no existe un modelo único de prueba. A cambio, puede ampliarse notablemente la cobertura curricular y se pueden hacer análisis más profundos, aunque sin poder establecer conclusiones de nivel individual, sino únicamente grupal. Cada tipo de estudios permite llegar a conclusiones de distinta naturaleza sobre la situación del sistema educativo, de los centros docentes y de los estudiantes.
- d) Una cuarta distinción puede establecerse entre los estudios que evalúan los *resultados educativos en un momento concreto* y los que pretenden medir el *progreso educativo a lo largo del tiempo*. Los del primer tipo proporcionan una información valiosa acerca del estado de una parcela determinada del sistema educativo en un momento dado, en una circunstancia temporal concreta. Los del segundo tipo se proponen valorar el progreso producido en un lapso de tiempo determinado. Los estudios más sofisticados de los que se desarrollan en la actualidad combinan ambos propósitos. Así, por ejemplo, TIMSS, PIRLS y PISA están concebidos como ciclos continuados, lo que permite, por una parte, conocer cuál es la situación educativa en un año concreto, y por otra, valorar el progreso producido cada varios años. De ese modo, los países participantes no sólo pueden compararse con otros en un momento dado, sino que pueden analizar cuál es su evolución.

Como puede apreciarse, los estudios internacionales tienen características diferentes, que permiten establecer distinciones entre ellos, tanto por el enfoque que adoptan como por la población elegida y el tipo de comparación predominante. Por lo tanto, se puede concluir que para poder interpretar o valorar correctamente un estudio determinado hay que analizar y tener en cuenta sus características específicas.

Los estudios mencionados tienen dos objetivos fundamentales, que no deben entenderse necesariamente como contrapuestos. Por una parte, se proponen presentar una información confiable sobre la situación comparativa de los diversos sistemas educativos, o al menos sobre algunas parcelas concretas de los mismos. Por otra

parte, pretenden ofrecer algunas claves para explicar e interpretar los resultados logrados. Aunque ambos objetivos no son alternativos, pues muchos estudios combinan ambos, cabe distinguirlos a efectos de análisis.

Esa voluntad de explicación que subyace en muchos estudios está ligada a un compromiso con la mejora de la educación, pues para poder emprender acciones de mejora es necesario conocer cuáles son las claves fundamentales del rendimiento educativo. Sin embargo, hablando en términos rigurosos, existe la convicción generalizada de que no es sencillo elaborar modelos causales del rendimiento, pese a los intentos que se han realizado en ese sentido (Álvaro et al., 1990). Por ese motivo, y consciente de esa dificultad, me he limitado aquí a presentar y analizar los factores que se toman generalmente en consideración a la hora de explicar los resultados, sin pretender llegar a conclusiones definitivas.

Desde esa perspectiva, las variables que pueden ayudarnos a explicar los resultados de la educación son muy diversas. No obstante, en ese amplio conjunto se pueden distinguir dos grandes categorías:

- Un primer grupo está compuesto por variables que se refieren a factores que no están directamente ligados al funcionamiento del sistema educativo o de los centros escolares. Guardan relación con las condiciones sociales y económicas en que se desenvuelve la tarea educativa, por lo cual no resulta sencillo actuar sobre ellas desde el marco estrictamente escolar. Aquí se denominan *variables extrínsecas*.
- Un segundo grupo está formado por variables que se refieren a factores directamente relacionados con el funcionamiento del sistema educativo o de los centros escolares. Tienen que ver con el modo de organización del sistema educativo, así como con el funcionamiento de las escuelas y el desarrollo de los procesos de enseñanza y aprendizaje, por lo que puede actuar sobre ellas desde el marco educativo y escolar. Aquí se denominan *variables intrínsecas*.

Entre la diversidad de variables de ambos tipos que pueden identificarse, los investigadores y evaluadores se han centrado especialmente en algunas de ellas, que he sintetizado en el cuadro siguiente y que servirán de eje de análisis en los apartados sucesivos.

| Variables extrínsecas | Variables intrínsecas |
|--|--|
| <ul style="list-style-type: none"> • Nivel socioeconómico y cultural de los estudiantes • Nivel de desarrollo educativo y cultural de la población • Recursos destinados a la educación | <ul style="list-style-type: none"> • Organización del sistema educativo • Procesos de centro escolar • Procesos de aula |

2. BUSCANDO EXPLICACIONES: VARIABLES EXTRÍNECAS DEL RENDIMIENTO

Entre las variables extrínsecas, hay tres que tienen especial interés, además de contar con tradición en la investigación educativa. Se trata del nivel socioeconómico y cultural de los estudiantes y de sus familias, el nivel de desarrollo educativo y cultural del conjunto de la población y los recursos que se destinan a la educación. Aunque esa relación no agota el conjunto de variables extrínsecas que podrían considerarse, tiene al menos en cuenta las más relevantes.

— Nivel socioeconómico y cultural de los estudiantes

A la hora de explicar los resultados logrados por los estudiantes, tomar en consideración el nivel socioeconómico y cultural de la población evaluada constituye una práctica habitual de la investigación educativa (Tiana, 2002). Hoy se acepta plenamente que los estudiantes están condicionados por una serie de factores de carácter restringido o *microsocial*, relacionados con las características socioeconómicas y culturales concretas de su grupo familiar y de su entorno cercano, y por otros de carácter más amplio o *macrosocial*, relacionados con aspectos tales como su posición social o las características de su entorno más distante. En consecuencia, hay que tomar en consideración ambos tipos de variables cuando se analizan los resultados logrados por los alumnos.

Puesto que en el párrafo anterior se habla del nivel socioeconómico y cultural, resulta necesario señalar que los primeros estudios internacionales de evaluación del rendimiento que se llevaron a cabo hace ya medio siglo optaron generalmente por controlar el nivel estrictamente socioeconómico de la población analizada, adoptando un enfoque por entonces habitual en los estudios empíricos de sociología de la educación. Sin embargo, posteriores análisis más detallados pusieron de manifiesto que los resultados educativos tienen mayor relación con los indicadores de posición social cuando éstos incluyen algunas variables relativas al grado de riqueza cultural del entorno.

Ese fenómeno ha sido bien analizado en algunos estudios de evaluación de ámbito nacional. Así, por ejemplo, en la evaluación de la entonces nueva Educación Primaria que se llevó a cabo en España en 1995³ se aprovechó la ocasión para construir dos índices complementarios y parcialmente coincidentes, a partir de algunas variables extrínsecas como las que aquí se mencionan. El primero fue el *índice socioeconómico*, compuesto por variables relativas a la profesión de los padres y la posesión familiar de bienes materiales. El segundo fue el *índice sociocultural*, compuesto

3 Ese estudio fue el primero íntegramente diseñado y desarrollado por el Instituto Nacional de Calidad y Evaluación (INCE). En consecuencia, se utilizó para hacer varios ensayos pilotos, como algunos que se comentan en estas páginas.

por variables relativas a los estudios y la profesión de los padres, más otras referidas a las expectativas educativas para los hijos y el acceso a algunos bienes culturales básicos (como libros, diccionarios u ordenador). Los análisis estadísticos realizados demostraron que la relación del segundo índice con los resultados logrados por los alumnos era claramente mayor que la del primero, lo que avala la importancia que poseen y el carácter predictivo que tienen las variables de tipo educativo y cultural por encima de las estrictamente económicas (INCE, 1996). Dicho de otro modo, el grado de riqueza material de la familia parece influir menos en los resultados académicos de los hijos que la riqueza del entorno educativo y cultural, lo que además concuerda con nuestra experiencia cotidiana.

En aplicación de esas conclusiones, los cuestionarios para las familias o los alumnos que acompañan a los recientes estudios españoles de evaluación (como los estudios periódicos de evaluación de la Educación Primaria o los relativos a la Educación Secundaria Obligatoria), al igual que los del proyecto PISA, incluyen preguntas acerca de aspectos tales como el número de libros disponibles en el hogar, la posesión de ordenador o diccionarios, el nivel educativo de los padres o las expectativas académicas hacia los hijos. La combinación de estas variables culturales con otras más clásicas como los bienes materiales en el hogar o la situación laboral de los progenitores eleva considerablemente el porcentaje explicado de la varianza, por lo que su uso se ha convertido en una práctica habitual.

En consecuencia, los estudios recientes de evaluación han incluido la elaboración de índices socioeconómicos y culturales con el propósito de contribuir a explicar las diferencias de resultados educativos. Uno de los modelos más completos actualmente en vigor es el Índice Socioeconómico y Cultural (ISEC) calculado por PISA, que permite llevar a cabo interesantes análisis acerca de la equidad de la educación, ya que permite estimar la correlación que existe entre el origen y la posición social de los estudiantes y sus resultados escolares (OECD, 2001b, 2004 y 2007).

Esos índices son utilizados para ponderar los resultados obtenidos, eliminando el efecto de las diferencias sociales y estimando con mayor precisión el de las estrictamente educativas. Además, la ponderación del nivel socioeconómico y cultural no sólo se efectúa cuando se comparan estudiantes, individualmente considerados, sino también cuando se evalúan los resultados logrados por colectivos más amplios, como pueden ser centros docentes o incluso sistemas educativos. Su propósito consiste en última instancia en asegurar la comparación justa entre las unidades evaluadas, evitando atribuir una eficacia mayor a aquellas unidades (centros educativos o distritos escolares) que simplemente se benefician de un alumnado más favorecido social y culturalmente. La realización de análisis que tienen en cuenta el denominado *valor añadido* no es sino uno de los ejemplos más claros y sofisticados de ese tipo de aproximación (Thomas, 1998).

Aunque en la actualidad se utilizan distintos modelos para calcular el valor añadido por los centros docentes, todos ellos combinan dos elementos fundamentales. El

primero consiste en la posición y el origen social de los estudiantes, que permite comparar los resultados obtenidos por un centro concreto en relación con los que cabría esperar estadísticamente en función de las características del alumnado que a él concurre. Es el modelo utilizado para calcular la eficacia de los liceos franceses que se utilizó en los años noventa (DEP, 1994). El segundo consiste en la comparación del resultado obtenido en dos momentos diferentes, lo que permite valorar el progreso realizado en una institución y tiempo concretos. Como puede apreciarse, los resultados más interesantes derivados de la aplicación de estas técnicas se consiguen acercando la escala de análisis al tamaño del centro docente, lo que no resulta factible en muchos estudios internacionales.

Los estudios internacionales del rendimiento educativo han tomado generalmente en consideración esta variable externa, aunque aplicando diversos procedimientos para analizar su efecto. La primera aproximación que se utilizó, todavía muy poco sofisticada, consistió en comparar los resultados logrados por los diversos países con algún indicador global de su grado de desarrollo económico, tal como el Producto Interior Bruto *per capita*. En este caso, los análisis se limitaron generalmente a yuxtaponer las tablas de los resultados obtenidos y del PIB nacional, o todo lo más a ajustar las puntuaciones de acuerdo con el PIB, como se aprecia en el ejemplo incluido en la Tabla 1.

TABLA 1

Resultados de los estudiantes, Producto Interior Bruto y puntuaciones ajustadas al PIB per capita

| País | Puntuación PISA (2006) | Producto Interior Bruto (US\$) (2005) | Puntuaciones ajustadas al PIB per capita |
|-------------|------------------------|---------------------------------------|--|
| Finlandia | 563 | 30.923 | 558 |
| Corea | 522 | 22.277 | 534 |
| Alemania | 516 | 30.826 | 511 |
| Reino Unido | 515 | 32.890 | 506 |
| Austria | 511 | 34.409 | 499 |
| Suecia | 516 | 23.000 | 496 |
| Dinamarca | 496 | 34.090 | 485 |
| Francia | 495 | 30.352 | 491 |
| EEUU | 489 | 41.674 | 464 |
| España | 488 | 27.507 | 490 |
| Italia | 475 | 28.168 | 476 |
| Portugal | 474 | 20.043 | 490 |
| Grecia | 473 | 29.564 | 471 |
| México | 410 | 10.767 | 443 |

Fuente: OECD (2007, vol. 2: 45).

Otro procedimiento que también se ha utilizado habitualmente consiste en el cálculo de líneas de regresión entre los resultados y el PIB, derivadas de tablas como la mencionada, si bien conducen generalmente a análisis poco sofisticados. Así, por ejemplo, PISA 2006 introduce ese tipo de análisis, llegando a la conclusión de que el 28% de la varianza entre países se explica por el PIB *per capita* (OECD, 2007, vol. 1: 59).

Recientemente se ha realizado una aproximación más precisa al efecto del nivel socioeconómico y cultural, consistente en utilizar los datos individualizados de los alumnos evaluados, referidos a variables tales como el nivel de estudios y la ocupación profesional de los padres, los bienes materiales de que dispone la familia, el nivel de renta familiar, las expectativas académicas hacia los hijos o los recursos culturales disponibles en el hogar. La complejidad que supone obtener datos fiables de este tipo (por la pérdida de respuestas que suele producirse cuando se pregunta a las familias y por la imprecisión que implica muchas veces preguntar a los propios alumnos) ha sido el motivo de que hasta ahora se haya utilizado de forma limitada esta aproximación en los estudios internacionales⁴. El estudio PISA, que se dirige a jóvenes de 15 años, se ha animado a aplicar este tipo de instrumento, pese a las eventuales imprecisiones que pueda implicar, permitiendo análisis interesantes⁵.

TABLA 2

Relación entre los resultados de los estudiantes y su nivel socio-económico y cultural (PISA, medias brutas y ponderadas)

| País | PISA 2000 | | PISA 2003 | | PISA 2006 | |
|-------------------|------------|------------|------------|------------|------------|------------|
| | MB | MP | MB | MP | MB | MP |
| Corea | 525 | 534 | 542 | 549 | 522 | 522 |
| España | 493 | 504 | 485 | 504 | 488 | 499 |
| EEUU | 504 | 497 | 483 | 463 | 489 | 483 |
| Francia | 505 | 512 | 511 | 516 | 495 | 502 |
| Italia | 487 | 487 | 466 | 473 | 475 | 478 |
| Portugal | 470 | 488 | 466 | 508 | 474 | 492 |
| Suecia | 516 | 504 | 509 | 492 | 503 | 496 |
| <i>Media OCDE</i> | <i>500</i> | <i>505</i> | <i>500</i> | <i>500</i> | <i>500</i> | <i>500</i> |

Fuente: OECD (2001b: 308; 2004: 358 y 2007, vol. 1: 184).

4 Como se señalaba en la nota anterior, el estudio de Evaluación de Educación Primaria realizado por el INCE en 1995 sirvió para estudiar la aplicabilidad de los cuestionarios a los alumnos y a sus familias acerca de sus características socioeconómicas y culturales, arrojando unos resultados poco alentadores.

5 En este caso, la edad de los estudiantes evaluados ha resultado más conveniente para aceptar la fiabilidad de sus respuestas.

La Tabla 2 proporciona una información sintética de los resultados obtenidos en el Proyecto PISA, cuando se tiene en cuenta el nivel socioeconómico y cultural de los estudiantes evaluados, para los tres ciclos ya realizados y publicados. En cada una de las columnas de la izquierda correspondientes a cada uno de los tres años contemplados se presentan las medias brutas (MB) obtenidas por una selección de los países participantes, mientras que en las columnas de la derecha aparece su media ponderada (MP), obtenida mediante la aplicación del índice socioeconómico y cultural calculado a partir de los datos del alumnado.

Como puede apreciarse fácilmente, hay algunos países (España, Francia y Portugal) cuya media ponderada es en todos los casos superior a la bruta, lo que indica que sus resultados hubiesen sido superiores si los alumnos evaluados tuviesen una situación socioeconómica y cultural equivalente a la media de la OECD. Ese efecto es independiente de que sus resultados ponderados deban considerarse satisfactorios o no. Así, Francia se sitúa por encima de la media ponderada, mientras que España se coloca casi en ella y Portugal sigue quedando por debajo, aunque con una menor distancia que con respecto a la media bruta. El caso de Italia es interesante, pues su puntuación casi no cambia en 2000 al ponderarla, aunque queda algo más alejada de la media de la OECD al aumentar ligeramente esta última, mientras que mejora en 2003 y 2006. Una tendencia inversa presenta Corea, que mejora claramente su media al ponderarla en 2000 y 2003, pero ya no en 2006. Hay otros países, como Estados Unidos o Suecia, cuya puntuación ponderada es inferior a la bruta, como consecuencia de tener una población con un índice socioeconómico y cultural superior a la media de la OECD. En ambos casos la caída de posición es notable, pues Suecia pasa a situarse al nivel de España o algo por debajo, mientras que Estados Unidos cae claramente por debajo de la media ponderada de la OECD.

La lectura de esa tabla, aunque sea un simple ejemplo que no deba extrapolarse a cualquier otro ámbito, permite extraer varias conclusiones. Ante todo, resulta indudable que el nivel socioeconómico y cultural familiar influye en los resultados alcanzados por los alumnos. Pero también hay que reconocer que el efecto que produce la ponderación de esa variable es diferente para los diversos países, como no podría ser de otro modo. Así, para los países cuya población disfruta de unas condiciones socioeconómicas y culturales mejores que la media internacional, la ponderación produce un descenso de la puntuación, mientras que el efecto es el contrario para los países con una población más desfavorecida desde el punto de vista socioeconómico y cultural.

Por otra parte, no hay que creer que la ponderación de la situación socioeconómica y cultural de la población evaluada produzca necesariamente un efecto de igualación en las puntuaciones. El caso de Corea ilustra adecuadamente cómo en determinadas ocasiones la ponderación realizada eleva aún más la distancia con la media (que en este caso aumenta de 25 a 29 puntos por encima de la media de la OECD en 2000 y de 42 a 49 puntos en 2003, si bien permanece estable en 2006). El efecto concreto producido por este tipo de ponderación depende del nivel socioeconómico y

cultural de la población nacional, pero también del nivel de rendimiento logrado por el país. Por lo tanto, hay que concluir que los mejores resultados obtenidos por algunos países no se explican exclusivamente en función de esta variable extrínseca, debiendo incluir algunas otras en el análisis.

A la vista de estos análisis, cabe concluir que el nivel socioeconómico y cultural de los estudiantes y sus familias tiene un efecto importante en los resultados que aquellos logran, pero que en modo alguno es un efecto unívoco. Por lo tanto, hay que tenerlo en cuenta a la hora de explicar los resultados y de realizar interpretaciones de los mismos, pero no constituye una variable determinante, puesto que existen casos de países que, por arriba o por abajo, escapan a la lógica puramente socioeconómica.

— *Nivel de desarrollo educativo y cultural de la población*

Como se apuntaba más arriba, en los estudios nacionales que intentan valorar y explicar el rendimiento de los sistemas educativos se ha detectado un fenómeno interesante, que consiste en la relativa disociación del efecto de las condiciones socioeconómicas propiamente dichas de los alumnos, de un lado, y la riqueza de su entorno educativo y cultural, de otro (INCE, 1996). Dicho de otro modo, cuando se intenta explicar los resultados logrados por los estudiantes se aprecia que guardan una cierta relación con sus condiciones socioeconómicas, esto es, con los bienes materiales y la renta disponible de la familia, si bien la relación encontrada es bastante mayor si a esas variables se les agregan algunas otras sobre el acceso a bienes culturales, las aspiraciones y expectativas educativas familiares y las prácticas culturales, o sea, los elementos fundamentales de lo que Bourdieu y Passeron (1970) denominaron el *capital cultural* de la familia y los individuos.

Cabe suponer que ese fenómeno se debería también apreciar en los estudios internacionales, si bien hasta ahora no ha sido objeto de análisis detenido. En los trabajos promovidos por la IEA no se ha llevado a cabo dicho análisis hasta el momento actual, debido a las dificultades que se plantean para obtener la información necesaria⁶. El proyecto PISA, al obtener datos singularizados del nivel socioeconómico y cultural de los alumnos participantes, debería permitir hacer algún estudio de ese tipo, aunque todavía no se haya llevado a cabo de manera completa y rigurosa. Los únicos análisis disponibles combinan ambas dimensiones en un mismo índice, lo que impide apreciar su efecto diferencial.

En todo caso, parece lógico preguntarse por el efecto que pueda producir la existencia de condiciones educativas y culturales muy diferentes entre los países par-

⁶ Hay que tener en cuenta que muchos de los estudios promovidos por la IEA se dirigen a escolares de nueve o diez años de edad, lo que hace muy poco confiable la información de tipo socioeconómico que puedan proporcionar. Por otra parte, la distribución de cuestionarios a las familias plantea problemas importante, tanto logísticos como de tasa de respuesta, por lo que no se han utilizado habitualmente en estos estudios internacionales.

tipicantes en los estudios internacionales. Y además conviene preguntarse por el diferente efecto que pueda producir la riqueza socioeconómica y cultural de la familia propia, de un lado, y del entorno más amplio, de otro. Esto es, si el anterior conjunto de variables tenía una dimensión predominantemente individual y familiar, en este segundo grupo se abordan otras variables de dimensión colectiva y social. El presupuesto que subyace en tal distinción se basa en el fenómeno bien conocido del efecto que ejerce sobre el individuo el carácter más o menos estimulante del ambiente en que se desenvuelve.

En la Tabla 3 se han resumido algunos indicadores relativos al nivel educativo y cultural de la población de los países de la Unión Europea con quince miembros⁷. Un primer indicador seleccionado consiste en el porcentaje de la población comprendida entre los 15 y los 64 años de edad que ha finalizado al menos la educación secundaria superior (aunque bajo esa denominación común quepan realidades muy diversas). Otro indicador complementario consiste en el número de lectores habituales de periódicos por cada mil habitantes. Como tercer elemento de comparación, se ha incluido la puntuación bruta alcanzada por los países de la Unión Europea en el estudio PISA del año 2000, centrado en la comprensión lectora y que ha servido como criterio de ordenación⁸. Aunque los años de referencia de los tres indicadores no sean idénticos, no constituye un obstáculo insalvable para un análisis general como el que aquí se hace.

Lo primero que llama la atención es que los países europeos que obtienen en el proyecto PISA puntuaciones por encima de la media de la OCDE (500 puntos) tienen, en general, mayor porcentaje de población adulta con estudios superiores que los situados por debajo de esa media. Sin embargo, hay varias excepciones a esta regla general. Por una parte, Irlanda está incluida entre los países con menor proporción de población con estudios superiores (51%) y sin embargo obtiene una alta puntuación en PISA. Por otra parte, Alemania y Dinamarca son los países que tienen mayor proporción de población con estudios superiores (81% y 80%, respectivamente) y sin embargo quedan por debajo de la media. Es evidente que esa sencilla comparación no avala la existencia de una relación perfecta entre ambas variables. Pero si nos fijamos en los países del sur de Europa, comprobaremos que tanto Italia como España, Portugal y Grecia se encuentran entre los que menos población tienen con estudios superiores y también entre los que obtienen puntuaciones por debajo de la media, lo que nos obliga a no desechar rápidamente la existencia de algún tipo de relación.

7 Se ha preferido utilizar aquí los datos procedentes del estudio PISA, ya que en él han participado más países de la Unión Europea que en PIRLS. No obstante, se podrían y deberían realizar análisis comparativos de los resultados logrados en los distintos estudios internacionales, dadas las diferencias de enfoque, materia y población que implican, a las que se ha hecho mención con anterioridad.

8 Conviene hacer notar que, siendo PISA un estudio realizado por la OECD, las medias calculadas corresponden a los países miembros de tal organización. En consecuencia, los informes realizados no incluyen la media de los países de la Unión Europea, por lo que no se puede hacer ninguna comparación con la misma.

TABLA 3

Resultados de los estudiantes en comprensión lectora, población con estudios superiores y lectores de periódicos en los países de la Unión Europea-15

| País | Puntuación PISA (2000) | Porcentaje de población 25-64 años con estudios superiores (1999) | Lectores de periódicos por 1000 habitantes (2000) |
|----------------------|------------------------|---|---|
| Finlandia | 546 | 72 | 445 |
| Irlanda | 527 | 51 | 150 |
| Reino Unido | 523 | 62 | 319 |
| Suecia | 516 | 77 | 417 |
| Austria | 507 | 74 | 308 |
| Bélgica | 507 | 57 | 153 |
| Francia | 505 | 62 | 148 |
| Dinamarca | 497 | 80 | 277 |
| España | 493 | 35 | 106 |
| Italia | 487 | 42 | 104 |
| Alemania | 484 | 81 | 291 |
| Grecia | 474 | 50 | 64 |
| Portugal | 470 | 21 | 68 |
| Luxemburgo | 441 | 56 | 348 |
| Holanda ⁹ | - | - | 278 |

Fuentes: OECD (2001^a: 45 y 2001b: 308), *El País* (18 abril 2002).

Por otra parte, al comparar los resultados de PISA con la lectura de periódicos vuelve a apreciarse un fenómeno similar. Ahora son tres los países incluidos en el grupo de menor proporción de lectores que quedan por encima de la media en PISA (Irlanda, Bélgica y Francia), mientras que dos de los países con mayor proporción de población lectora de periódicos quedan por debajo de la media (Luxemburgo y Alemania). Los países del sur de Europa vuelven a quedar por debajo de la media en ambos indicadores.

Aunque no es propósito de este trabajo descender a un análisis detenido y riguroso de esta posible relación, estudio que sin duda reclama atención y que deberá emprenderse en algún momento, una sencilla reflexión sobre los datos incluidos en la Tabla 3 apunta a la existencia de algún tipo de conexión entre el grado de desarrollo educativo y cultural de un país (medido aquí por dos indicadores simples) y los

⁹ No se incluyen datos de Holanda, dado que no aparecen en las dos publicaciones de la OECD que se han utilizado como fuente.

resultados logrados por sus estudiantes. Al igual que se observaba en el caso de la variable anterior, no podemos hablar de una relación perfecta, ni mucho menos de una determinación causal, pero hay indicios que obligan a analizar con más atención la conexión que pueda existir. En concreto, el análisis de la situación educativa de los países del sur de Europa, que han sufrido un retraso educativo histórico y que han evolucionado muy rápidamente en las últimas décadas, pero sin poder salvar fácilmente esa brecha centenaria que les separa de otros países europeos, merece un estudio más detenido, puesto que parece estar en relación con los resultados que alcanzan sus estudiantes. Desde ese punto de vista, valdría la pena hacer un estudio comparativo más cuidadoso con el caso de Irlanda, que podría arrojar información interesante.

— *Recursos destinados a la educación*

La tercera variable extrínseca aquí analizada se refiere a los recursos que cada país destina a la educación. Dichos recursos son de naturaleza muy variada, debiendo distinguirse al menos entre los humanos (profesorado, personal de administración y servicios educativos) y los económicos y materiales (infraestructura, equipamiento, presupuestos). Además, hay que tener en cuenta que esos recursos pueden proceder de muy diversas fuentes, tanto públicas como privadas, y de los distintos niveles de las administraciones públicas (estatal, provincial, regional, local). Puede incluso discutirse si esta variable debe considerarse extrínseca, puesto que entra dentro de la esfera propia de decisión de los gobiernos. El motivo de incluirla en este grupo consiste en que la determinación del presupuesto educativo es una tarea que no compete solamente a las autoridades educativas, sino que se inserta en un cuadro más complejo de asignación de prioridades y fondos presupuestarios para el conjunto de las políticas nacionales. Por lo tanto, no se trata de una decisión interna al propio sistema educativo, sino externa a él.

Las dificultades que plantea la comparación de los recursos realmente utilizados en las tareas educativas ha implicado que la mayor parte de los análisis hasta ahora realizados estén basados en el gasto que se realiza en educación (incluyendo a veces la inversión propiamente dicha), medido en términos monetarios, siendo mucho menos frecuente otro tipo de comparación. Además, hay que señalar que los estudios internacionales de evaluación del rendimiento raramente incluyen análisis precisos acerca de la relación entre los resultados educativos y los recursos monetarios utilizados. Lo más que hacen, como en el caso de TIMSS o del proyecto PISA, es yuxtaponer ambas informaciones, pero sin analizar detenidamente su relación. Por lo tanto, al igual que sucedía en el apartado anterior, no es fácil llegar a conclusiones definitivas sobre la conexión que pueda existir entre ambas variables.

A pesar de esa dificultad, también en este caso he intentado analizar el efecto que ejerce la variable que aquí se considera. En la Tabla 4 he recogido algunos datos

correspondientes a varios de los países participantes en el proyecto PISA¹⁰. La selección de países se ha efectuado de acuerdo con la relevancia que dichos casos tienen en el contexto de este trabajo y con su carácter ejemplificador. La primera columna incluye los resultados brutos obtenidos en PISA 2000 que, al igual que en otras tablas precedentes, han servido para ordenar los países. La segunda columna recoge el Producto Interior Bruto *per capita* de dichos países, mientras que la tercera incluye el gasto acumulado en instituciones educativas por estudiante. Estas dos últimas cifras se expresan en dólares USA, utilizando un algoritmo de paridad del poder adquisitivo (PPP), a fin de hacer las cifras comparables. La última columna incluye el cociente resultante de dividir las dos anteriores, representando el esfuerzo económico efectuado por estudiante en relación con el Producto Interior Bruto nacional.

TABLA 4

Resultados de los estudiantes, Producto Interior Bruto y gasto acumulado en instituciones educativas por estudiante

| País | Puntuación PISA (2000) | Producto Interior Bruto (US\$) (1999) | Gasto acumulado por estudiante (US\$) (1998) | Cociente Gasto acumulado / PIB |
|-------------|------------------------|---------------------------------------|--|--------------------------------|
| Finlandia | 546 | 22.800 | 45.363 | 1,99 |
| Corea | 525 | 15.900 | 30.844 | 1,94 |
| Reino Unido | 523 | 22.300 | 42.793 | 1,92 |
| Suecia | 516 | 23.000 | 53.386 | 2,32 |
| Austria | 507 | 24.600 | 71.387 | 2,90 |
| Francia | 505 | 21.900 | 50.481 | 2,31 |
| EEUU | 504 | 33.900 | 67.313 | 1,99 |
| Dinamarca | 497 | 26.300 | 65.794 | 2,50 |
| España | 493 | 18.100 | 36.699 | 2,03 |
| Italia | 487 | 21.800 | 60.824 | 2,79 |
| Alemania | 484 | 23.600 | 41.978 | 1,78 |
| Grecia | 474 | 14.800 | 27.356 | 1,85 |
| Portugal | 470 | 16.500 | 36.521 | 2,21 |
| México | 422 | 8.100 | 11.239 | 1,39 |

Fuente: OECD (2001b: 264).

¹⁰ En este caso he preferido utilizar también los datos de PISA en vez de los de TIMSS, por su mayor cobertura europea (como ocurría en la Tabla 3) y por proporcionar en el mismo informe los datos sobre gasto acumulado por alumno, lo que implica la aplicación de una metodología, si no idéntica, al menos coherente entre ambas series de datos.

Como puede apreciarse en la tabla, los resultados de los países no siguen el mismo orden que el gasto acumulado por estudiante. Así, Corea se sitúa en la parte alta de la tabla, pese a que su gasto por estudiante es de los más bajos de la OCDE, exactamente al contrario que le sucede a Italia. Además, hay países con un nivel similar de rendimiento en PISA 2000, a los que corresponden niveles bastante dispares de gasto. Es lo que sucede al comparar Francia con Estados Unidos, Corea con el Reino Unido o Grecia con Portugal. A la inversa, hay países con niveles comparables de gasto acumulado por estudiante pero con niveles dispares de resultados, como el caso del Reino Unido y Alemania o España y Portugal.

Tampoco el cociente que se obtiene al dividir el gasto acumulado por estudiante entre el PIB sigue un orden igual al de los resultados. Aunque México ocupa la última posición en ambos casos, Italia está muy por debajo de lo que le correspondería en el caso de existir una correlación perfecta, y también estarían fuera de su lugar países tan dispares como Corea, el Reino Unido y Finlandia (todos ellos por encima de su posición teórica), así como Portugal y Dinamarca (en este caso por debajo).

Por lo tanto, también en el caso de esta tercera variable extrínseca pueden hacerse algunas consideraciones similares a las que se formulaban con las dos anteriores. Nos encontramos en una situación en la que podemos afirmar que todo apunta a la existencia de una cierta relación entre los recursos dedicados a la educación y el rendimiento educativo, pero la relación dista mucho de ser lineal ni perfecta. Sin duda, es una de las variables que deben ser tenidas en cuenta a la hora de valorar y explicar los resultados, pero no es la única. Como se ha afirmado muchas veces, no se trata solamente de dedicar más recursos a la educación, sino de emplearlos bien. Y en esta última parte de la afirmación intervienen otras variables, muchas de ellas de carácter intrínseco, que han de tenerse en cuenta.

Antes de entrar en el análisis de las variables intrínsecas, habría que señalar que las reflexiones realizadas en este apartado, concordantes por otra parte con una orientación ya tradicional en la investigación educativa, apuntan a la existencia de relaciones múltiples y cruzadas entre todas estas variables extrínsecas. Por lo tanto, aun cuando pueda interesar en ocasiones el estudio singular de algunas de ellas, no es posible explicar los resultados logrados en los estudios internacionales por medio de una sola variable. Los análisis deben respetar la complejidad de los fenómenos estudiados e integrar distintos factores en cada explicación.

3. EN BUSCA DE MÁS EXPLICACIONES: VARIABLES INTRÍNSECAS DEL RENDIMIENTO

Las variables extrínsecas ayudan a explicar los resultados conseguidos por los estudiantes en las pruebas de rendimiento escolar, pero no dan cuenta total

de los mismos. Por ese motivo hay que explorar otro tipo de factores, relativos a las características y al funcionamiento del sistema educativo, que tienen además especial interés cuando se pretende llevar a cabo acciones de mejora, ya que caen plenamente dentro de la esfera de actuación de las autoridades educativas, de los centros escolares y del profesorado. Son las aquí denominadas *variables intrínsecas*.

Las variables intrínsecas que analizan los estudios actuales de evaluación del rendimiento son muy diversas. Muchos de los estudios promovidos por IEA, OECD y UNESCO recogen información acerca de variables tales como la formación y el desempeño de los profesores, el grado de autonomía de la escuela, el tiempo dedicado a la enseñanza y al estudio personal, la organización del currículo, el tipo de refuerzo prestado a los alumnos y otras similares, dimensiones todas ellas de la vida educativa que caen dentro de esta categoría.

Entre este tipo de variables he seleccionado aquí tres grupos. En primer lugar, analizaré los efectos producidos por algunos factores relacionados con la organización general del sistema educativo. En segundo lugar, pasaré revista al efecto de los denominados *factores de centro escolar*, que se refieren a la organización, el funcionamiento y el clima o ambiente escolar, considerados en términos generales. En tercer lugar, analizaré el efecto de los denominados *factores de aula*, que guardan relación con las condiciones en que se desarrollan los procesos de enseñanza y aprendizaje en el interior del aula, en una situación de contacto directo entre el profesor y los alumnos.

— *Organización del sistema educativo*

Las variables intrínsecas de mayor nivel de generalidad son las que se refieren a la organización del sistema educativo. Se trata de variables que afectan al conjunto de los centros y de los agentes escolares, puesto que tienen que ver con las normas fundamentales y las condiciones generales en que aquellos deben desenvolverse. Este tipo de variables son en realidad características relativas a la estructura del propio sistema educativo y a sus reglas básicas de funcionamiento.

De acuerdo con los supuestos de la concepción del *mundo como laboratorio educativo* que se comentaba al comienzo del trabajo, los estudios internacionales deberían contribuir a proporcionar explicaciones ligadas precisamente a este tipo de variables, ya que son las que distinguen más claramente a unos sistemas educativos de otros. Sin embargo, hay que reconocer que los estudios realizados rara vez son concluyentes y casi siempre proporcionan una información que requiere interpretaciones complementarias. Hasta ahora son pocas las conclusiones definitivas que han aportado los estudios internacionales sobre este tipo de variables. Sin embargo, se pueden señalar algunas, aunque haya que tomarlas con cierta precaución.

Entre los ejemplos más relevantes de este tipo de análisis, hay que acudir una vez más a PISA. Uno de los fenómenos que más llamó la atención tras la publicación del informe de resultados correspondiente al año 2000, y que más debate generó, consistió en el pobre rendimiento educativo de los jóvenes alemanes de quince años. Como se ha visto en las tablas insertas en el trabajo, su puntuación está claramente por debajo de la media de la OECD y también por debajo de lo que cabría esperar de un país con un alto nivel de desarrollo. Ese dato provocó alarma y generó una viva discusión (Kotthoff y Pereyra, 2009). Lo que no llamó tanto la atención es que esa circunstancia no era exclusiva de Alemania, sino que también se apreciaba, en mayor o menor grado, en otros países con sistemas educativos semejantes. La característica fundamental de ese tipo de sistemas consiste en una *diferenciación institucional temprana*. Dicho de otro modo, los adolescentes de esos países son separados muy tempranamente (entre los diez y los doce años de edad) en itinerarios formativos muy diferenciados, que se corresponden con otros tantos modelos institucionales escolares. Aunque el sistema prevé la posibilidad de trasvases entre unos y otros itinerarios, ese paso se produce en la práctica en muy pocos casos.

A partir de estos datos, los responsables del estudio PISA llegaron a una triple conclusión que merece la pena destacar (OECD, 2001b). En primer lugar, las diferencias que existen entre los diversos tipos de escuelas en esos países son generalmente mayores que las que se observan en otros lugares. Si en la mayor parte de los países participantes en el estudio la varianza intra-escuelas era mayor que la varianza entre escuelas, en estos países sucedía exactamente lo contrario. En segundo lugar, la variación de resultados que se aprecia entre los estudiantes evaluados también suele ser mayor que en otros países, con muy pocas excepciones. En tercer lugar, ninguno de esos países se encuentra entre los de mayor rendimiento promedio. O sea, el estudio PISA parece concluir que la discriminación institucional temprana (la que tiene lugar antes de los doce años de edad) produce desigualdades importantes en los países que la aplican, sin llegar por ello a obtener unos resultados promedios satisfactorios. Los mejores estudiantes de esos sistemas educativos están entre los mejores del mundo, pero también los peores se encuentran entre los peores, provocando una situación global poco positiva y desequilibrada. Cabría inferir que este modelo de organización del sistema educativo parece poco aconsejable para las circunstancias actuales, tanto por la escasa equidad que logra como por su bajo rendimiento.

Otro ejemplo concreto del análisis de ciertas variables que parecen incidir en los resultados de los estudiantes se encuentra en las conclusiones sobre la autonomía escolar que aporta el estudio PISA. A partir de la información recogida en dicho estudio, se ha elaborado la Tabla 5, que sintetiza la relación existente entre los resultados que logran los alumnos y el porcentaje de directores escolares que afirman tener al menos alguna autonomía en una serie de áreas de actuación educativa.

TABLA 5

*Autonomía escolar y resultados educativos
(relación entre los resultados de los alumnos y el porcentaje de directores
que tienen alguna responsabilidad en los siguientes aspectos de la gestión escolar)*

| | |
|---|------|
| Decisión sobre los cursos ofrecidos | 0,51 |
| Determinación de asignaciones presupuestarias | 0,37 |
| Elección de manuales escolares | 0,30 |
| Determinación de los contenidos de los cursos | 0,25 |
| Establecimiento de la política disciplinaria | 0,21 |
| Establecimiento de la política de evaluación de los alumnos | 0,20 |
| Nombramiento de profesores | 0,16 |
| Despido de profesores | 0,10 |

Fuente: OECD (2001b: 175).

Lo que la tabla manifiesta es que no se puede afirmar taxativamente que una mayor o menor autonomía escolar produzca globalmente mejores o peores resultados de los estudiantes. Como puede apreciarse a partir de esos datos, la autonomía escolar es una realidad que abarca diversas dimensiones y no todas ellas tienen la misma incidencia sobre el rendimiento escolar. En términos generales, podría decirse que la autonomía en los aspectos relativos a la organización curricular, la vida y el ambiente del centro escolar y la gestión de sus recursos presupuestarios parece ejercer un mayor efecto sobre el rendimiento que la autonomía en lo relativo a la gestión del personal docente. Así, cabría concluir que ciertas dimensiones de la autonomía escolar merecen un especial apoyo por parte de las autoridades educativas, mientras que hay otros aspectos en que la acción no es tan prioritaria.

— *Procesos de centro escolar*

Este último tipo de reflexiones conducen directamente al segundo grupo de variables intrínsecas, las relacionadas con la vida de los centros escolares. Son los comúnmente denominados *procesos de centro*. Los procesos de centro y los de aula, que serán objeto de comentario en el apartado siguiente, constituyen en conjunto lo que habitualmente se conoce como *procesos escolares*. Esta dimensión constituye un nivel intermedio entre el contexto socioeconómico y cultural y los resultados educativos. O sea, la acción desarrollada por los profesores y por otros miembros de la comunidad escolar, mediante el desarrollo de diversos procesos y actuaciones, es la que explica que se logren unos determinados resultados, a partir de unas determinadas condiciones de partida. De ahí deriva su importancia y el interés que el tema despierta en los investigadores (Marchesi y Martín, 1998).

La importancia de este tipo de procesos es indudable, aunque haya que situarla en su justo término. Frente a la tendencia inaugurada por la publicación del célebre Informe Coleman (Coleman *et al.*, 1966) y por la sociología crítica que insiste en la *reproducción* (Bourdieu y Passeron, 1970), la investigación educativa reciente ha puesto de manifiesto que no todos los resultados de la educación pueden explicarse exclusivamente en función de las fuerzas reproductoras del contexto, aunque su presencia y su impacto sean innegables. Así, un estudio de hace unos años en que se revisaron veinticuatro investigaciones y se realizaron varios metaanálisis llegaba a la conclusión de que entre un 12% y un 15% de la varianza en los resultados que alcanzan los alumnos es debida a la acción del centro escolar (Teddlie, Reynolds y Sammons, 2000). Se trata de un porcentaje que, sin ser demasiado elevado, no resulta despreciable y que está en línea con lo que otros investigadores han concluido. Podríamos decir que ese es el espacio propio de actuación de los centros educativos y del profesorado y que, por tanto, tiene notable importancia desde el punto de vista pedagógico.

Los estudios de evaluación del rendimiento de ámbito nacional han concedido importancia a los procesos de centro y de aula. Los estudios internacionales también han incluido este tipo de procesos entre sus variables de análisis, aunque con unos enfoques y una formulación ligeramente diferente de uno a otro caso. Además, algunos proyectos concretos de evaluación institucional, como el desarrollado en España por el Instituto IDEA, han prestado notable atención a su estudio, en la convicción de que constituyen el ámbito privilegiado de cualquier plan de actuación para la mejora de la educación (Marchesi y Martín, 2002).

Hablando concretamente de los procesos de centro escolar, he revisado las variables de proceso contempladas por los estudios desarrollados en las últimas décadas por la IEA, OECD y UNESCO antes mencionados, con el fin de identificar las que parecen guardar mayor relación con el rendimiento. Entre todas las variables incluidas en esos estudios, las que parecen gozar de un mayor consenso son las siguientes:

- a) Liderazgo pedagógico, entendido como la existencia de una dirección con autoridad pedagógica reconocida, capaz de promover un buen funcionamiento del centro escolar, facilitadora del logro de un buen clima escolar y eficaz en su funcionamiento cotidiano.
- b) Coordinación pedagógica y curricular entre los profesores y las etapas, con el fin de asegurar una coherencia interna en la enseñanza proporcionada y en los medios puestos en juego para lograr el aprendizaje.
- c) Implicación familiar en la escuela, como señal de la participación de los padres en la educación de sus hijos y de un estilo de actuación capaz de asegurar la coherencia educativa entre la escuela y el hogar.
- d) Buen ambiente o *clima* escolar, entendido como la existencia de unas buenas relaciones entre los distintos miembros de la comunidad escolar (estudiantes, profesores, familias), facilitadoras de la tarea educativa.

- e) Formación y estabilidad del profesorado, ya que existe la certeza de que la actuación de los docentes constituye un elemento capital para el logro de una enseñanza de calidad, y que esa actuación no puede ser positiva si no se dan unas condiciones favorables, tanto en lo que se refiere a la capacitación de los profesores, como a su seguridad laboral, una de cuyas variables fundamentales es la estabilidad (Barber y Mourshed, 2007).
- f) Tamaño adecuado de la escuela, que no siempre resulta fácil de determinar, pero que tiene relación con la existencia de una ambiente diversificado y rico (que exige unas dimensiones mínimas) pero no masificado (que impone unos límites superiores).

El modo en que los distintos estudios internacionales formulan y exploran el efecto de estas variables difiere de unos a otros casos, pero existe una notable coincidencia en la elección de las áreas abarcadas. Aún a falta de realizar un análisis comparado detenido de esas formulaciones y aproximaciones, cuestión que tiene indudable interés pero aún no se ha llevado a cabo, se aprecia la existencia de unas áreas prioritarias de análisis en dichos estudios. No obstante, hay que señalar que la dispersión de aproximaciones utilizadas por las diferentes organizaciones que los promueven, la dificultad que plantea la definición común de variables como las mencionadas para sistemas educativos tan distintos entre sí y la limitada consistencia de los análisis realizados, obligan a seguir trabajando en este campo, como han hecho tanto la IEA como la OECD en sus estudios respectivos¹¹.

— *Procesos de aula*

Las variables referidas a los procesos de centro se complementan estrechamente con las relativas a los denominados *procesos de aula*. Si las variables anteriores configuran un campo complejo y multidimensional, en el que no resulta sencillo encontrar explicaciones unívocas y obliga a combinarlas en análisis multifactoriales, esa complejidad se acentúa incluso más cuando descendemos al plano de las prácticas educativas en el aula. Como afirma Elena Martín, lo que sucede en una clase se caracteriza por la multidimensionalidad, la simultaneidad, la inmediatez, la impredecibilidad, la singularidad y el carácter valorativo de las actividades (Martín, 2002). Por ese motivo, todos los estudios que intentan valorar el efecto de este tipo de variables en el rendimiento de los alumnos tienen que reducir su foco de atención y centrarse en algunas variables concretas.

Si eso sucede en las investigaciones que adoptan el aula como objeto de estudio, aún con mayor motivo se ven obligados a hacerlo los estudios internacionales, dada la

¹¹ Un ejemplo destacable de este empeño es el estudio denominado TIMSS-Video, que desarrolló la IEA en conexión con la aplicación de TIMSS y que estaba basado en el análisis de clases de matemáticas y ciencias grabadas en video en varios países, con la intención de analizar en profundidad los diversos procesos aplicados para la enseñanza de esas materias, sin limitarse a procesar las respuestas que los propios profesores dan a través de cuestionarios (Hiebert *et al.*, 2003). La información obtenida permitió completar y contrastar los datos aportados por los propios profesores, enriqueciendo así el análisis.

distancia que mantienen con las aulas y la diversidad de estilos de configuración y funcionamiento de éstas que existen en el plano internacional. Por esa razón, dichos estudios se centran en un número limitado de variables, seleccionando aquellas que la investigación educativa reconoce como más potentes a la hora de explicar los resultados.

Entre las variables relativas a los procesos de aula que se han tomado en consideración en los estudios internacionales realizados durante las dos últimas décadas y que han arrojado unos resultados más prometedores, he identificado las siguientes:

- a) Tiempo dedicado a la enseñanza, entendido tanto en el sentido del tiempo efectivamente dedicado a las tareas docentes en la escuela (excluyendo, por ejemplo, interrupciones no deseadas, que en algún caso llegan a ser considerables), como el tiempo dedicado a tareas escolares en el hogar.
- b) Estructuración de la enseñanza, entendida como una organización adecuada y cognitivamente estimulante de las tareas de aprendizaje por parte del docente.
- c) Oportunidad de aprender, entendida en el sentido en que la IEA desarrolló este concepto hace tiempo, esto es, como la exposición real a la enseñanza (en el sentido del currículo efectivamente impartido), que proporciona al estudiante la oportunidad de desarrollar su aprendizaje.
- d) Altas expectativas de rendimiento por parte de los profesores, factor que la investigación educativa ha demostrado cumplidamente que va ligado a unos mejores resultados de los estudiantes.
- e) Evaluación continua y control del progreso de los alumnos, en el sentido de ofrecer una realimentación permanente del proceso de enseñanza y aprendizaje y de ayudar a los alumnos a valorar por sí mismos su progreso.
- f) Refuerzo positivo al aprendizaje, entendido de diversos modos, no necesariamente en su concepción conductista, y que tiene que ver con la introducción de estímulos para continuar en la tarea y para esforzarse en conseguir mejores resultados.
- g) Proporción adecuada de alumnos por profesor que, al igual que sucedía con el tamaño de la escuela, no es una función lineal, sino que tiene un mínimo y un máximo convenientes, por debajo o por encima de los cuales la tarea docente y el aprendizaje resultan menos estimulantes o más difíciles.

El problema que plantean todas estas variables, tanto las referidas a los procesos de centro como a los procesos de aula, radica en la dificultad que implica su investigación empírica. Los docentes y los estudiosos de la educación saben que se trata de variables muy influyentes en el desarrollo de las actividades escolares y, en conse-

cuencia, en los resultados logrados por los alumnos. Pero dada su relación multidimensional, no resulta sencillo medir ni valorar su efecto concreto.

Una de las líneas de investigación que ha pretendido estimar el efecto de este tipo de variables sobre el rendimiento académico es la que se conoce como estudio de la *eficacia escolar*. Se trata de una línea de trabajo con una larga tradición, pero que se renovó en los años noventa (Scheerens, 1992). Entre los principales autores que han contribuido a renovarla merece la pena citar al holandés Jaap Scheerens. En un trabajo publicado en 1996, uno de los pocos existentes que han intentado realizar un análisis global de los factores de eficacia escolar, se preguntaba por la aportación que este movimiento puede realizar para la gestión y la mejora de la educación (Scheerens, 1996). Lo cierto es que sus conclusiones no resultaban excesivamente esperanzadoras. Tras realizar diversos análisis empíricos a partir de investigaciones propias y de otros autores, llegaba a la conclusión de que muchas de las variables acerca de cuyo efecto existe un consenso bastante generalizado no disfrutaban de una base empírica suficiente, como recoge la Tabla 6.

TABLA 6

Grado en el cual las variables intrínsecas se han visto confirmadas por la investigación empírica

| Característica | Confirmación múltiple | Base empírica razonable | Confirmación dudosa | Hipotética |
|--|-----------------------|-------------------------|---------------------|------------|
| Enseñanza estructurada | X | | | |
| Tiempo de aprendizaje | X | | | |
| Oportunidad de aprender | | X | | |
| Orientación hacia el logro | | X | | |
| Altas expectativas | | X | | |
| Liderazgo pedagógico | | | X | |
| Habilidad evaluadora | | | X | |
| Clima escolar | | | X | |
| Reclutamiento de personal | | | | X |
| Condiciones estructurales y organizativas | | | X | |
| Condiciones físicas y materiales | | X | | |
| Características del contexto | | | X | |
| Estímulos externos hacia la eficacia escolar | | | | X |
| Implicación familiar | | X | | |

Fuente: Scheerens (1996: 111).

La lectura correcta de esta tabla no implica que dichas variables deban ser despreciadas ni que su efecto sea irrelevante. Sabemos que muchas de ellas desempeñan un papel importante en la actividad escolar. El problema está en que no resulta fácil cuantificarlas ni medirlas. Y si esa dificultad surge en estudios de pequeña escala, aún es mayor en los estudios internacionales, que adoptan una perspectiva muy amplia. Por ese motivo, los resultados que hasta ahora ha proporcionado en este campo la investigación internacional del rendimiento son limitados. Aunque todas las organizaciones que trabajan en este campo están desarrollando modelos finos y sofisticados para valorar el impacto de las variables intrínsecas sobre el rendimiento, todavía no contamos con aproximaciones unánimemente aceptadas.

4. CONCLUSIONES

Como se ha analizado en estas páginas, los estudios internacionales de evaluación del rendimiento educativo surgieron con la intención primordial de permitir una comparación rigurosa de los resultados logrados por los estudiantes de los diversos sistemas educativos. No cabe ninguna duda de que ese objetivo se ha saldado con éxito. En efecto, este tipo de estudios, que ya cuentan con una considerable tradición de casi cincuenta años, permiten comparar la situación de un sistema educativo determinado con otros que pueden servirle de referencia, dando la posibilidad de identificar sus principales puntos fuertes y débiles, así como de estimar el progreso que se produce a lo largo del tiempo. Hoy en día, la información proporcionada por proyectos como TIMSS, PIRLS o PISA es objeto de estudio especializado, de debate público, de análisis académico y de valoración política. La comparación de los sistemas educativos se ha enriquecido notablemente con la información suministrada por los sistemas internacionales de indicadores y por los estudios de evaluación de gran escala.

Pero dichos estudios también pretendieron desde el comienzo contribuir a explicar los resultados conseguidos, por medio del análisis de los denominados *factores asociados*. Es en este aspecto en el que me he detenido en estas páginas y sobre el que conviene plantear algunas reflexiones finales.

Quizás el principal problema que he abordado aquí no se refiera sobre todo a qué explicaciones han proporcionado hasta ahora los estudios internacionales del rendimiento, sino a qué podemos legítimamente esperar de ellos. Hasta ahora, podemos afirmar que tales estudios han arrojado luz sobre algunos de los factores analizados. Así, hoy poseemos bastantes más certezas que hace unos años acerca de la relación entre los resultados educativos y el nivel socioeconómico y cultural de los estudiantes, las familias y el entorno en que se desenvuelven. No es que sea una novedad absoluta, pero los análisis comparativos internacionales han ayudado a clarificar la cuestión. También hemos encontrado alguna confirmación estadística, basada en el manejo de grandes bases de datos, del efecto producido por características estructurales del sistema educativo, tales como la diferenciación institucional temprana o las diversas dimensiones de la autonomía escolar. Y lo mismo podemos decir de

los principales efectos sobre el rendimiento de los factores contemplados por el movimiento de la eficacia escolar. Pero, sin embargo, la impresión que nos queda al final del análisis es de cierta insatisfacción, pues tenemos más preguntas y dudas que respuestas y certezas. Podemos decir que hemos encontrado algunas nuevas explicaciones, pero aún no podemos explicar convincentemente y con rigor el rendimiento educativo.

Desde este punto de vista, se puede decir que la concepción del *mundo como laboratorio educativo* que alumbró la IEA en los años sesenta fue una idea feliz, pero más como inspiración que como resultado. No es justo considerar que se trata de una quimera, pero sí de un planteamiento tendencial, hacia el que nos debemos orientar y que quizás no alcancemos nunca.

No cabe duda de que los estudios internacionales ofrecen algunas posibilidades interesantes, como las arriba señaladas, aunque a condición de reconocer sus limitaciones. En mi opinión, su potencia explicativa se ve limitada sobre todo por la perspectiva macroscópica que adoptan, que les hace perder detalles importantes, por la mirada poco contextualizada que necesariamente deben arrojar sobre los sistemas educativos y por la estandarización que imponen sus métodos y técnicas de análisis. Cuando se trata de estudiar fenómenos singulares, de pequeña escala o que deben estar correctamente contextualizados, la opción de los estudios internacionales puede no ser la más adecuada.

Aunque aquí no se han analizado otras iniciativas de evaluación, como los estudios nacionales de diagnóstico o la evaluación institucional, no cabe duda de que constituyen un complemento muy valioso, incluso indispensable, para comprender y valorar más profundamente las razones y los factores que explican los resultados alcanzados. Por más valiosos e interesantes que sean los estudios internacionales, que sin duda lo son, no bastan para comprender el funcionamiento profundo del sistema educativo ni para adoptar todas las decisiones que redunden en la mejora de la educación. De ahí deriva el interés que tiene la comprensión de la evaluación como una tarea integrada por acciones de diversa naturaleza y con diverso enfoque y perspectiva, idea cada vez más compartida.

De esta serie de reflexiones se deduce la necesidad que tienen los sistemas educativos de adoptar instrumentos adecuados para llevar a cabo la gestión de sus recursos y el seguimiento de sus resultados (Tiana, 2003). Conocer los resultados y su explicación se ha convertido cada vez más en una exigencia para la mejora. No es casualidad que en este contexto se hayan desarrollado nuevos conceptos, como el de *pilotage o steering* que, en castellano, quizás podría traducirse por *conducción*. La denominación fue introducida por Gilbert de Landsheere, para quien podría definirse como “la toma de decisiones al nivel macroscópico (es decir, de un sistema o subsistema educativo) o microscópico (es decir, de un establecimiento o incluso una clase), sobre la base de constataciones tan objetivas como sea posible, relativas al estado, al funcionamiento o a los productos del sistema” (De Landsheere, 1994:8). La evaluación

es un instrumento muy adecuado y valioso para proporcionar ese tipo de información objetiva y confiable (Michel, 1996).

Pero, si se concibe la conducción como una tarea que se desarrolla a diversos niveles, los estudios internacionales no resultan suficientes para realizarla de forma completa. Los estudios internacionales de evaluación del rendimiento no son sino uno de los instrumentos disponibles, pero en modo alguno el único. Los países pueden obtener de ellos información relevante sobre su situación educativa, pero necesitan combinarla con otro tipo de datos y de análisis. Ahí es donde se encuentra el interés y las posibilidades reales que ofrecen los estudios internacionales, y por su contribución a esa tarea habrá que valorarlos.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Álvaro, M. *et al.* (1990), *Hacia un modelo causal del rendimiento académico*. Madrid: Ministerio de Educación y Ciencia – Centro de Investigación, Documentación y Evaluación.
- Barber, M. y Mourshed, M. (2007), *How the world's best-performing school systems come out to the top*. Londres: McKinsey & Co.
- Bourdieu, P. y Passeron, J.C. (1970), *La reproducción: éléments pour une théorie du système d'enseignement*. Paris: Editions de Minuit.
- CERI (1994), *Making Education Count. Developing and Using International Indicators*. Paris: CERI – OECD.
- Coleman, J.S. *et al.* (1966), *Equality of Educational Opportunity*. Washington D.C.: Office of Education.
- De Landsheere, G. (1994), *Le pilotage des systèmes d'éducation*. Bruselas: De Boeck.
- Degenhart, R.E., ed. (1990), *Thirty years of international research. An annotated bibliography of IEA publications (1960-1990)*. The Hague: IEA.
- DEP (1994), *Trois indicateurs de performance des lycées, Baccalauréat 1993, Résultats lycée par lycée*. Paris: Ministère de l'Éducation nationale – Direction de l'Évaluation et de la Prospective (2 tomos).
- Hiebert, J., *et al.* (2003), *Teaching Mathematics in Seven Countries: Results From the TIMSS 1999 Video Study (NCES 2003-013)*. Washington DC: United States Department of Education - National Center for Education Statistics.
- INCE (1996), *Evaluación de la Educación Primaria. Informe preliminar*. Madrid: Instituto Nacional de Calidad y Evaluación (documento policopiado).
- Kotthoff, H-G. y Pereyra, M.A. (2009), “La experiencia del PISA en Alemania: Recepción, reformas recientes y reflexiones sobre un sistema educativo en cambio”, *Profesorado. Revista de currículum y formación del profesorado*, vol. 13, 2: 1-24.
- Marchesi, A. y Martín, E. (1998), *Calidad de la enseñanza en tiempos de cambio*. Madrid: Alianza Editorial.
- Marchesi, A. y Martín, E., comps. (2002), *Evaluación de la educación secundaria. Fotografía de una etapa polémica*. Madrid: Fundación Santa María – SM.

- Martín, E. (2002), “Los procesos de aula y su influencia en la calidad de la enseñanza”, en A. Marchesi y E. Martín, comps., *Evaluación de la educación secundaria. Fotografía de una etapa polémica*. Madrid: Fundación Santa María – SM.
- Michel, A. (1996), “La conducción de un sistema complejo: la Educación Nacional”, *Revista Iberoamericana de Educación*, n.º 10: 13-36.
- OECD (1992), *High-Quality Education and Training for All*. Paris: OECD.
- (2001a), *Education at a Glance. OECD Indicators*. Paris: OECD.
- (2001b), *Knowledge and Skills for Life. First Results from PISA 2000*. Paris: OECD.
- (2004), *Learning for Tomorrow’s World. First results from PISA 2003*. Paris: OECD.
- (2007), *PISA 2006. Science Competencies for Tomorrow’s World*. Paris: OECD, 2 vols.
- OECD & Statistics Canada (1995), *Literacy, Economy and Society. Results of the first International Adult Literacy Survey*. Paris: OECD – Ministry of Industry of Canada.
- (2000), *Literacy in the information age: final report of the international adult literacy survey*. Paris y Ottawa: OECD – Statistics Canada.
- OREALC (1998), *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado*. Santiago de Chile: OREALC – UNESCO.
- (2008), *Los aprendizajes de los estudiantes de América Latina y el Caribe. Primer reporte de los resultados del Segundo Estudio Regional Comparativo y Explicativo*. Santiago de Chile: OREALC.
- Papadopoulos, G.S. (1994), *L’OCDE face à l’éducation, 1960-1990*. Paris: OCDE.
- Papanastasiou, C., ed. (2004), *Proceedings of the International Research Conference 2004*. Nicosia: University of Cyprus – IEA, 3 vols.
- Purves, A. (1993), “The World as an Educational Laboratory”, en W.A. Hayes, ed., *IEA Activities, Institutions and People. IEA Guidebook 1993-1994*. The Hague: IEA.
- Scheerens, J. (1992), *Effective Schooling: Research, Theory and Practice*. Londres: Cassell.
- (1996), “Can the School Effectiveness Knowledge Base Guide School Management?”, en *II International Conference on School Management. Participative Management and School Evaluation*. Bilbao: Universidad de Deusto.
- Teddlie, Ch., Reynolds, D. y Sammons, P. (2000), “The Methodology and Scientific Properties of School Effectiveness Research”, en D. Reynolds y Ch. Teddlie, eds., *The International Handbook of School Effectiveness Research*. Londres: The Falmer Press.
- Thomas, S. (1998), “Value-added measures of school effectiveness in the United Kingdom”, *Prospects*, vol. XXVIII, 1: 91-108.
- Tiana, A. (2001), “Le monde comme laboratoire éducatif”, *Politiques d’éducation et de formation. Analyses et comparaisons internationales*, vol. 2001/3: 47-57.
- (2002), “El contexto sociocultural en la evaluación de los centros educativos”, en A. Marchesi y E. Martín, comps., *Evaluación de la educación secundaria. Fotografía de una etapa polémica*. Madrid: Fundación Santa María – SM.
- (2003), “Pilotage par les résultats et amélioration de l’éducation: deux réalités indissociables?”, *Administration et Éducation*, 98: 19-29.

- (2006), “Assessing Quality in Education: Concepts. Models and Instruments”, en Dobbelstein, P. y Neidhart, T., eds., *Schools for Quality – What data-based Approaches Can Contribute*. Sint-Katelijne-Waver: Consortium of Institutions for Development and Research in Education in Europe (CIDREE).

La *apropiación* de la evaluación por las administraciones¹

Saville Kushner

Resumen: *En los últimos treinta años venimos asistiendo a una progresiva “internalización” de la evaluación de programas y políticas en las administraciones públicas. Esto se ha producido bien porque las administraciones mismas han asumido la función de evaluación o porque éstas han estipulado cómo hacer las evaluaciones y con qué objetivos. Este hecho plantea preguntas importantes para el ciudadano y para la democracia. En este artículo se examinan las consecuencias de esta “apropiación” de la evaluación por parte de la administración a través de dos ejemplos: uno tomado del ámbito de la cooperación internacional y otro extraído del encargo que una agencia gubernamental hizo para la evaluación de uno de sus programas.*

Palabras clave: *Independencia de la evaluación, evaluación de programas, política.*

1. A MODO DE INTRODUCCIÓN

En su trabajo sobre “la inversión del ejecutivo en evaluación”, MacDonald (1981) planteaba la cuestión de qué es lo que gana la administración con la contratación de evaluaciones, si son actividades potencialmente peligrosas para la burocracia política. Los observadores profesionales (los evaluadores), que a menudo dan mucho valor a su independencia, obtienen acceso directo a lo que a menudo son fuentes confidenciales de información y además tienen capacidad de publicarlo. Por supuesto, los evaluadores proporcionan lo que puede ser información útil, pero dicha información no siempre compensa los riesgos políticos. Sin embargo, lo que es más importante, como dice MacDonald, es que sólo el racionalista piensa que un programa social es un medio para lograr fines políticos (lo que convierte la evaluación en algo útil en el sentido democrático y en el sentido formativo). A menudo, la realidad es que el programa suele ser un fin en sí mismo –con fines políticos sin consecuencias más allá

1 Revisión científica de la traducción a cargo de Ester García Sánchez.

de su fecha de finalización-, en cuyo caso la evaluación es sólo “simbólica... un elemento esencial de la decoración de cualquier ministerio que se precie”.

Aquí tenemos, pues, las razones principales por las que, para MacDonald, las administraciones subvencionan las evaluaciones; el reto psicológico, al leer esto, es mantener un sólido equilibrio entre el cinismo y el realismo:

- No hay opción (legislación, órdenes, presión, fondos especiales).
- La única alternativa puede ser un nivel *no* aceptable de trabajo común y consulta con los *stakeholders* que, a menudo, tienen agendas incompatibles.
- Mantiene a los escépticos –críticos potenciales o reales– ocupados en tareas “inofensivas”.
- La presencia de evaluadores especializados evita, o al menos pospone, una evaluación más exhaustiva por parte de evaluadores no especializados.
- Hace ganar tiempo: tiempo para que el gobierno disipe el interés que hizo necesaria la inversión en el programa.
- El gobierno puede requerir, sobre todo en los primeros momentos, un cierto nivel de implantación del programa y la presencia de evaluadores contendrá el comportamiento subversivo del colectivo que ha recibido los fondos.
- Es una señal para los escépticos de que el gobierno quiere conocer la verdad, se preocupa por la eficacia y está abierto a las evidencias.
- Con toda seguridad, resultará inofensiva gracias a la combinación de debilidades epistemológicas, impopularidad social y al propio interés de los evaluadores.
- La mejor razón de todas: alguna otra entidad puede financiar una evaluación de su programa.

Continúa mostrando cómo la administración puede protegerse frente a los riesgos de encargar una evaluación (incluso, por ejemplo, “retrasar la evaluación porque aumenta su necesidad...”). Aquí MacDonald no está argumentando que la evaluación no deba servir a la administración, sino que no debería servir *exclusivamente* a la administración y que mientras esté prestando el servicio, sería bueno que el evaluador conociera las intenciones que hay tras su contratación.

En la época en la que MacDonald escribió esto, la evaluación era una actividad que se estaba extendiendo pero que aún estaba poco desarrollada y se llevaba a cabo sobre todo en las universidades y por parte de académicos independientes. Este au-

tor se refería a la evaluación *externa*. Desde entonces, ha habido una tendencia a nivel mundial para la internalización de la evaluación, para incorporarla a las administraciones como una actividad propia. Según el análisis de MacDonald, esto permite controlar todos los riesgos y centrarse en los beneficios, independientemente de dónde se enmarcan. Hay dos vías principales de internalización: la de los organismos que cuentan con sus propios evaluadores y la de aquellos organismos que encargan evaluaciones externas a través de un contrato en el que se establecen unas determinadas condiciones a partir de sus necesidades de gestión de la información.

Lo que pretendo con este artículo es analizar dos de los efectos de esta internalización: uno, en el campo de la cooperación al desarrollo², en el que veremos cómo la preocupación de los gobiernos por la *gestión basada en resultados* “aplasta” el compromiso democrático y participativo durante la evaluación; otro, más cercano, es un caso de contratación de una evaluación por el gobierno del Reino Unido, en el que vemos cómo la internalización de la evaluación ha conducido a un marco muy rígido de cláusulas que han predeterminado la evaluación y limitado sus objetivos, haciendo que la evaluación fuera menos eficaz y democrática. Claramente, lo que me preocupa es la utilidad de la evaluación para la democracia. Sin embargo, primero necesitamos entender la evaluación en los términos más amplios y democráticos de sus primeros años. Últimamente, a medida que la evaluación se ha ido internalizando como un elemento para la gestión y la reforma en las administraciones públicas, sus distintas formas y enfoques se han ido restringiendo, cada vez más, a la medición *post hoc* de la productividad de los programas. Esta situación oculta el hecho de que los 25 años transcurridos desde la aparición de la evaluación de programas como una práctica y una disciplina diferenciada han visto cómo la evaluación adquiría facetas muy variadas, útiles y definidas. Esto se refleja en la amplia gama de denominaciones³ que se han usado para los distintos enfoques, cada una de ellos (enumerados a continuación) con unos rasgos característicos precisos⁴:

- Evaluación democrática (MacDonald)
- Evaluación libre de objetivos (Scriven)
- Evaluación comprensiva (Stake)
- Evaluación iluminativa (Parlett and Hamilton)
- Evaluador como científico público (Cronbach)
- Evaluación naturalista (Guba)
- *Empowerment evaluation* (Fetterman)
- Modelo CIPP - Contexto, Input, Proceso, Producto (Stufflebeam)
- *Dialogic evaluation* (Greene)
- *Affirmative evaluation* (Mertens)

2 En el original *international development*.

3 En la revisión científica se ha optado por dejar en su original en inglés aquellas denominaciones que no cuentan en castellano con una traducción suficientemente acuñada.

4 Véanse House (1980), para una clasificación de los enfoques de evaluación y un análisis de las tendencias en el surgimiento de la evaluación, y Alkin (2004) sobre el “árbol de la evolución” de la evaluación.

- Evaluación para la utilización (Patton)
- Evaluación personalizada (Kushner)
- Evaluación democrática deliberativa (House)

Cabe destacar, de paso, que la mayoría de estos autores hablaban de *evaluación educativa* y uno de los ejemplos que expondré a continuación pertenece a dicho campo. No es casual. En gran parte, fue en el campo de la educación, concretamente en el de la enseñanza y el aprendizaje⁵, donde se pusieron los cimientos de la evaluación de programas. Esto se debió, en cierta medida, a que éste fue el laboratorio en el que el gobierno hizo los primeros encargos de evaluación (Norris, 1990; House, 1993)

Para apreciar cómo se ha concebido la evaluación fuera de las “limitadas” definiciones dadas por las administraciones, Cronbach *et al.* (1980), en el famoso prefacio de su importantísima obra sobre evaluación de programas, establecieron 95 “tesis” o principios sobre evaluación, unos postulados valiosos y fundamentales, todo un manifiesto. Fueron escritas en la misma época en la que MacDonald se preguntaba por el uso de la evaluación por la burocracia y una inquietud similar subyace en el trabajo de Cronbach. Él hablaba del evaluador como “científico público”. El primero de estos 95 principios dice, simplemente: “La evaluación es el medio por el cual la sociedad aprende sobre sí misma”.

El sentido de esto es que cada evaluación de un programa es un estudio de caso sobre la sociedad, en la medida en que todos los programas son una representación, en miniatura, de la sociedad en la que se desarrollan. Tienen estructuras sociales de poder, una cultura, funciones y rituales, una economía y una teoría de la justicia. Tanto si el evaluador lo reconoce como si no, tanto si lo contempla su contrato como si no, es lo que sucede y ello plantea muchos interrogantes, que después han sido explorados en profundidad, sobre el carácter democrático de los programas y de su evaluación; por ejemplo, ¿quién tiene derecho a disponer de información sobre los programas y sobre su funcionamiento?

Con ésta y otras preguntas similares en mente analizaré a continuación mis dos ejemplos del impacto de la internalización de la evaluación. El primero tiene que ver con la situación de la evaluación en la cooperación al desarrollo y comienza con una historia personal.

2. PRIMER EJEMPLO: “DENTRO DE LA TIENDA DE CAMPAÑA” DE LA EVALUACIÓN DEL DESARROLLO

Cuando me estaban entrevistando para el puesto de responsable regional de control y evaluación de UNICEF que después llegaría a ocupar, me encontré metido

5 En el original *schooling*.

en una teleconferencia con cuatro países: Brasil, Panamá, Nicaragua y Estados Unidos. Por algún motivo parecía importante preguntar con qué país estaba hablando en cada momento. La pregunta más penetrante vino de Nueva York. “Interesante currículum vitae –muy creativo-, qué extenso. Pero durante 25 años ha estado usted trabajando, desde fuera de las administraciones públicas, en el análisis de lo que sucede dentro de ellas; ahora quiere trabajar dentro de la maquinaria burocrática más grande del mundo, para analizar lo que sucede fuera. ¿Cómo logrará ese cambio de mentalidad?”. Simplemente respondí que no tenía la menor idea, pero que era consciente del reto –que se trataba de una cuestión práctica– y que tendría que resolver ese problema cuando me enfrentara a él. Valoraron mi sinceridad.

Sin embargo, no respondí a todas las implicaciones de la sabia pregunta que luego tuve que encarar. La verdadera pregunta era: ¿qué sucede cuando la evaluación entra en la “tienda de campaña”?

Naciones Unidas es una organización muy fragmentada. Tiene muchos organismos que diversifican sus esfuerzos, por ejemplo, entre la ayuda humanitaria, el desarrollo cultural, el SIDA, los derechos del niño, el cambio climático, la reducción de la pobreza, las migraciones y la producción y distribución de alimentos; mientras, los que habitan en el hemisferio sur se ven afectados por todas estas cuestiones a la vez. Cambio climático, problemas de acceso a la escuela, migración, escasez de alimentos, enfermedades; se pueden padecer todos a la vez. No es de extrañar que exista una gran presión dentro de la ONU –y entre sus donantes– para la armonización, la normalización y la integración. Es lógico que el programa estrella de reforma se llame *One UN*.

Dentro de los propios organismos existe también una fragmentación similar. En mi oficina regional había expertos en pediatría, en educación, en políticas públicas, en protección de la infancia y en otros temas -y desarrollaban sus actividades de forma independiente. De hecho, había cierta reticencia a integrarlas. “¿Me han contratado como pediatra!” me contestó, protestando, uno de estos asesores cuando intenté sugerirle que todos ellos, en conjunto, debían gestionar el cambio. Era una respuesta comprensible pero disfuncional, teniendo en cuenta que los problemas sociales están interrelacionados.

Quizás, lo que produce una mayor sensación de fragmentación es la “geografía”; recordemos esa teleconferencia. La sede central de UNICEF en Nueva York, por ejemplo, está a una enorme distancia de sus siete oficinas regionales, a lo que se suma la distancia entre éstas y sus más de cien oficinas en distintos países; y de éstas a las miles de oficinas de campo en los centros de operaciones. La cuestión que se le plantea a la organización es cómo interpretar esas distancias. Puede hacerse de dos modos: como un problema de comprensión (aprender de la diversidad) o como un problema de control (disminuir o atenuar las diferencias). Es la típica cuestión del centro y la periferia. La sede central tiene dos preocupaciones: cómo aprender de las operaciones que llevan a cabo sobre el terreno y cómo asegurarse de que su política

se sigue de una manera coherente, de modo que la voz de UNICEF sea también coherente. Ambas cuestiones son reales e ineludibles y ambas, en mayor o menor medida, se asumen. De hecho, UNICEF es conocida como la agencia más descentralizada dentro de la ONU, en la que los representantes por países gozan de una autonomía mayor que la de, por ejemplo, sus colegas del PNUD –la mayor agencia dentro de la ONU– que gestiona los problemas geográficos de un modo más jerarquizado. La ONU (sí es que se puede hablar de una sola ONU) es, en términos de gestión, un organismo extremadamente heterogéneo y singular.

Sin embargo, en la actualidad se está volviendo al centralismo, auspiciado por un proceso de reforma en la ONU. De hecho, al igual que las demás agencias de la ONU, UNICEF debe cumplir con un nuevo sistema de gestión basado en resultados, una versión del *New Public Management* (resultados previamente especificados dentro de un sistema que exige responsabilidades pero dota a los responsables de escasa capacidad de decisión –*low-trust accountability*– y que asigna los recursos en función del rendimiento). Este sistema, como bien sabemos los que trabajamos en el sector público en el hemisferio norte, es un método para centralizar el poder y la autoridad en las organizaciones. Cuando me marché de UNICEF, la sede central estaba poniendo en marcha una nueva política que cambiaba el papel de los (poderosos) Directores Regionales de asesores e impulsores a asesores y *supervisores*: se esperaba que fueran más exigentes con determinado tipo de resultados y que fueran más sensibles a las comparaciones en cuanto al avance en el logro de los Objetivos de Desarrollo del Milenio. Se mantiene la autonomía, pero supeditada a la preocupación por garantizar una mayor homogeneidad. En mi opinión, la balanza se está inclinando (por las presiones de los países donantes) a favor de una agenda globalizada (igualdad de género, violencia infantil, medidas de supervivencia infantil, gripe aviar) y a costa de la diversidad de las operaciones en los distintos países.

Un síntoma de este cambio es el llamamiento a la “armonización” y a la “estandarización”. Hay tres razones principales para ello, dos de ellas son obvias y otra es más sutil. Una de las obvias es que a menudo hay maneras alternativas de medir las “patologías” sociales y es importante ponerse de acuerdo sobre cuál se utiliza. La desnutrición a menudo se analiza de dos formas: retraso en el crecimiento y atrofia. El retraso en el crecimiento se calcula midiendo la estatura con respecto a la edad; la atrofia se calcula midiendo el peso con respecto a la estatura. Estos indicadores arrojan unos resultados muy distintos. Por supuesto, pueden utilizarse en conjunto (y debería hacerse así porque son indicadores que se modulan mutuamente), pero a veces también se usan como una opción para llegar a las conclusiones políticas más convenientes según las circunstancias. Para evitarlo, se fomenta la “armonización” de los indicadores como vía para controlar la politización de la información. El otro ejemplo más publicitado y notorio es la incidencia del VIH/SIDA en una población determinada. En el Reino Unido, por ejemplo, hace mucho tiempo que dejamos de hacer extrapolaciones de las estadísticas obtenidas de las pruebas realizadas a mujeres en clínicas obstétricas, una muestra “cautiva” muy útil. La falta de fiabilidad se debe a razones obvias (por ejemplo, estábamos contabilizando únicamente a las mujeres

sexualmente activas y a mujeres que se presentaban voluntariamente) y las estadísticas antes estaban muy infladas. En África, estas metodologías y las estimaciones “infladas” sobre la prevalencia de la enfermedad en esos países persistieron durante muchos años; ¿recuerdan las cifras que se manejaban de un tercio de la población infectada? Había resistencia a cambiar la metodología, entre otras cosas porque las elevadas tasas de incidencia de la enfermedad estimulaban un flujo mayor de fondos “excepcionales”⁶.

La otra razón obvia para la armonización es que hay una necesidad impuesta por los burócratas de comparar unos países con otros, al igual que las distintas zonas o regiones del planeta. Esto forma parte del sistema global de *accountability* que ha propiciado la aparición de un fenómeno histórico único: los Objetivos de Desarrollo del Milenio. Independientemente de que existen realidades globales que superan contextos y culturas, el hecho es que la supervisión por parte de la ONU de determinadas responsabilidades a nivel global –en derechos humanos, cambio climático, conflictos violentos, etc., por dar algunos ejemplos– ha supuesto un impulso para el progreso moral y la reducción del caos (Black, 2008).

La razón sutil es la que más me preocupa y tiene que ver con el “control administrativo”. Los que trabajan en las administraciones no pueden permitirse los mismos lujos que nosotros, los evaluadores, de documentarse, teorizar y debatir sobre la complejidad de éste o aquel problema: tienen que trabajar con ese problema y resolverlo; ése fue el principal cambio de mentalidad que experimenté. Y además han de resolverlo en un contexto en el que deben rendir cuentas a los donantes por los impactos (atribuibles y mensurables). Hay que lograr la cuadratura del círculo; el deseo de guardar las “formas” dentro del proceso –*input*, proceso causal, *output*, impacto– es abrumador.

A veces esas “formas” se corresponden con la realidad, sobre todo cuando los problemas se pueden tratar con soluciones técnicas: porcentaje de niños vacunados, número de nacimientos registrados, cobertura con mosquiteras impregnadas de repelente, ingesta de sal yodada, proximidad a agua potable; todos los ejemplos anteriores tienen un impacto más o menos atribuible y mensurable sobre la salud y los derechos de las personas. Hay muchas otras cosas en el ámbito de la cooperación al desarrollo que no lo tienen, aún más desde que empezó a producirse un nuevo cambio que nos ha llevado de las operaciones de campo a la defensa de determinadas políticas en las que los impactos son dudosos y, en el mejor de los ca-

6 Véanse, por ejemplo: (i) M. Cheng, “AIDS cases drop, but bad data to blame”, *The Washington Post*, 20 de noviembre de 2007; (ii) J. Chin (2007), *The AIDS Pandemic: the Collision of Epidemiology with Political Correctness*, Oxford: Radcliffe; (iii) J.R. Glynn, A. Buve, M. Carael, R. Musonda, M. Kahindo, I. Macculey, M. Hawken, F. Tembo, L. Zekeng (2000), “A new method of estimating HIV prevalence in men using antenatal clinic sentinel surveillance”, ponencia presentada en la *Conferencia Internacional sobre el SIDA*, 9-14 de julio de 2000; 13: abstract núm. ThOrC669; (iv) C.A. McGarrigle, A. Cliffe, A.J. Copas, C.H. Mercer, D. DeAngelis, K.A. Fenton, B.G. Evans, A.M. Johnson y O.N. Gill (2006), “Estimating adult HIV prevalence in the UK in 2003: the direct method of estimation”, *British Medical Journal*, 82(3): 78-86.

sos, *inferidos* y contextualizados. En este caso, la evaluación como análisis de contextos es esencial, al igual que lo es utilizar esa evaluación para poner nuevamente en contacto a los políticos con la realidad de cada comunidad. Cuanto más nos alejamos de las soluciones técnicas en los problemas muy específicos (como la vacunación), más nos adentramos en un territorio incierto en cuanto a los resultados, respecto a lo que significan para los distintos *stakeholders*, cuáles son sus causas, lo que representan en términos de calidad y, lo que es más importante, hasta qué punto dependen de los contextos de la actuación. “Al final, la evaluación precisa de la cadena de hechos y datos y de las interpretaciones de quienes tienen experiencia en el campo correspondiente, es insustituible”, afirmaba Deaton (2008) al cuestionarse la enorme difusión de los enfoques experimentales que pueden ser útiles para intervenciones muy concretas, pero que no logran reflejar la complejidad de los programas sociales.

En estos casos, necesitamos la evaluación para lo que con excesiva frecuencia se consideran funciones accesorias: aprendizaje, intercambio de información, descripción, deliberación (debate) o para dejar claro cuál es el contexto. Cuando las causas no están claras o son difíciles de averiguar necesitamos la evaluación para conocer las contingencias; no se trata tanto de saber cómo A da lugar a B sino más bien de saber cómo A y B tienen una relación dinámica entre sí, pero también con X, Y y Z. Picciotto (2007) defiende que la evaluación ha de situar las intervenciones en cooperación al desarrollo dentro de su contexto político y económico, sacarlas del enfoque miope que se centra en la propia intervención y sus objetivos.

La dificultad que se nos plantea es que la apropiación de la evaluación por las administraciones tiende a que se devalúen estas cosas en favor de funciones supelementalmente cruciales y por las que el burócrata es responsable: la medición de resultados. Como me comentó en cierta ocasión un alto funcionario de UNICEF, “el seguimiento nos dice si vamos bien encaminados; la evaluación nos dice si hemos llegado”. Suena bien, pero es un sofisma. Dado que los Objetivos Globales del Milenio definen puntos de destino universales, podemos ver la lógica que subyace al deseo de armonizar y estandarizar la evaluación. A esto se suma el deseo de incrementar y mantener niveles de competencia y calidad en la práctica evaluadora, defender los logros en el desarrollo de capacidades que no han sido fáciles de alcanzar, consagrándolos como estándares o normas e insistiendo en el compromiso de aprender de “los mejores”.

El problema es la tensión entre la estandarización y el contexto: lo primero niega lo segundo. La urgencia para asegurar que se ponen en práctica las políticas y se aseguran los impactos tiende a desplazar las funciones democráticas de la evaluación: estimular la crítica y el debate social; hacer que los programas de intervención sean transparentes; facilitar el intercambio de información; dar voz a la ciudadanía; proporcionar datos públicos como base para el diálogo entre la sociedad y la administración. Estas funciones sólo se pueden realizar en su contexto, *estas* personas en *este* lugar hablando entre ellas de sus preocupaciones en *este* preciso momento. Como he

sugerido en algún otro momento, la evaluación democrática fue concebida a imagen de la política municipal, local (Kushner, 2009). En el nivel municipal las personas se conocen, hablan cara a cara, cuentan con pocos recursos para desarrollar una ciencia muy sofisticada, la rendición de cuentas se basa en una conciencia de las relaciones humanas, la política se basa en el conocimiento del lugar y en las preocupaciones de sus ciudadanos. Todo esto se ve amenazado por la armonización. La estandarización es enemiga del contexto.

No es que esto no se reconozca; desde luego en UNICEF lo saben. Hay un intenso debate sobre la democracia y la evaluación. El problema es que debe verse en paralelo al compromiso ineludible con la gestión basada en resultados -y conste que se hacen meritorios esfuerzos por reconciliarlas, por compaginar ambas exigencias (Franklin, 2008). Ciertamente son casi irreconciliables y simplemente ponen de relieve la dificultad de la tarea de los burócratas, que tienen que hacer juegos malabares con “ortodoxias” rivales. Es injusto para ellos y para los ciudadanos. La democracia no debe ser vista como responsabilidad de las administraciones.

3. SEGUNDO EJEMPLO: “PERSONAS SENSATAS” EN CIRCUNSTANCIAS INSENSATAS

El siguiente ejemplo viene del Reino Unido y tiene que ver con el encargo de una evaluación que hizo un organismo gubernamental. A principios de los años 70 se publicó en Reino Unido el Informe Rothschild (1971), en el que se establecían las condiciones para la contratación de proyectos de investigación por parte del gobierno. El Informe definía la relación entre el gobierno y el investigador como la de un “cliente y un prestador de servicios”. El cliente (gobierno) establece las condiciones, mientras que el prestador del servicio (investigador) cumple con dichas condiciones a tenor de unas cláusulas contractuales. Básicamente, definía el encargo de evaluación como una transacción económica, como “condiciones de contratación” (Pettigrew, 1994). Ésta fue la causa principal de la internalización gradual de la investigación y la evaluación. Promovía el concepto de “cliente informado”, es decir, el burócrata con suficientes conocimientos de investigación y evaluación como para poder especificar exactamente cómo deben llevarse a cabo.

El cambio subsiguiente, y bastante rápido, de la contratación externa de la evaluación a su internalización, quedó ampliamente documentado en su momento por evaluadores preocupados por el control de la investigación académica (Pettigrew, 1994; Norris, 1995; Simons, 1995). Más recientemente, la preocupación por el control de la información pública ha sido absorbida por otras preocupaciones más generales sobre lo que ha dado en llamarse *The New Public Management* -la contratación por resultados y la centralización del control (Kushner y Norris, 2007)-, incluso al nivel de los cambios sociales estructurales (Powers, 1999; O’Neill, 2002). Estos debates rara vez afloran en la evaluación del desarrollo: es una cuestión democrática que se ignora en aras de la supuesta necesidad urgente de acción.

El texto que incluyo a continuación es parte de un memorándum que escribí y, que posteriormente dio lugar a que la *British Educational Research Association* revisara los consejos que daba a sus miembros a la hora de negociar sus contratos con organismos oficiales. Esto muestra el tipo de control sobre el conocimiento que surge cuando la administración asume la responsabilidad de definir las condiciones en las que aquél se produce. Trata de las condiciones y los detalles concretos de una evaluación establecidos por un organismo gubernamental responsable de una cuestión educativa fundamental:

Para los que buscan colaborar en el desarrollo pedagógico y educativo, las oportunidades de trabajar con la [Agencia] son de máxima importancia. Sin embargo, las condiciones contractuales de financiación son tan onerosas como para resultar inaceptables. Dichas condiciones suscitan graves dudas sobre la independencia y la base ética de los estudios encargados por la Agencia, pero lo que es aún más preocupante, amenazan con comprometer la calidad de la información que se traslada a la propia Agencia. Tratemos cada uno de ellos:

(1) Condiciones de financiación: la [Agencia] impone tales exigencias en la elaboración de los informes sobre los que a menudo son proyectos a corto plazo, que comprometen nuestra capacidad de realizar estudios rigurosos y de proporcionar informaciones correctas. En los requisitos de un concurso reciente para un contrato de diez meses de duración se exigían informes de progreso quincenales, un “diálogo constante” y la presentación de seis informes parciales/finales junto con un número indeterminado de reuniones para hablar sobre el contrato y tres reuniones adicionales para hablar sobre los borradores del informe final, además de un número no especificado de reuniones con el denominado “comité de seguimiento”. Al parecer, los “prestadores del servicio” también deben presentar y negociar un “plan de salida” al mes de comenzar el estudio.

(2) Ética e independencia: la Agencia impone unos niveles inusualmente elevados –incluso para las agencias oficiales– de control y revisión de los informes que están sujetos a procedimientos de calidad de la [Agencia], incluso por su “tono”. Resulta sorprendente que insistan en inspeccionar y examinar la correspondencia con terceros (externos) además de supervisar los instrumentos de investigación y los informes. Los informes parciales tienen que contar con el visto bueno de la Agencia antes de convertirse en informes finales. La [Agencia] se reserva el derecho de decidir a quiénes se les envían los informes (‘audiencias’). La [Agencia] no sólo insiste en quedarse con los derechos de propiedad intelectual –en el mejor de los casos, por la confusión entre “comprador” y “entidad financiadora”, y en el peor de los casos porque abusan de sus privilegios – sino que también insiste en la propiedad de datos personales y confidenciales que, por contrato, deben ser enviados a la Agencia. Esto sólo puede provocar que se supriman los comentarios críticos de profesores y profesores-educadores a quienes se les aconsejaría que se abstuviesen de aportar puntos de vista que quizás pudieran estar cuestionando la labor de la [Agencia], por ejemplo, privando así a la Agencia de una valiosa fuente de opinión. Aparte de todo esto, no está claro que la Agencia pueda “hacerse con la propiedad” de los datos personales de ninguna persona.

No está claro sobre qué base se arroga la [Agencia] unos niveles tan altos de control sobre los informes que se pagan con dinero público y que frecuentemente contendrán información de gran interés e importancia para los profesionales de la educación y para los padres con hijos en edad escolar. ¿Por qué le “pertenecen” más a la [Agencia] estos informes que a los profesores y a los padres que forman parte del consejo escolar, por ejemplo? En las reuniones, los trabajadores de la Agencia suelen insistir en que la potestad de controlar sólo se utilizará de forma aislada y justificada. Aunque los trabajadores de la [Agencia] son personas sensatas y honradas, la “sensatez” y la “honradez” no describen a un ente jurídico, ni garantizan un comportamiento sensato. Cuando concedemos a los entes jurídicos (no a sus empleados) un control tan férreo sobre el conocimiento, se abre la posibilidad de procesos propios de “sociedades cerradas”. Si la eliminación gradual del derecho público al conocimiento sobre la enseñanza no se parece a un contrato con la [Agencia], ¿a qué otra cosa se podría parecer?

¿Qué vemos en los párrafos anteriores? Desde el punto de vista de la Agencia hay una preocupación por la calidad de la investigación. No es raro, en los últimos años en Reino Unido, que una agencia de este tipo insista en supervisar todos los instrumentos de investigación –protocolos de entrevistas, encuestas– y la explicación que dan es que el rápido incremento en el número de expertos universitarios y consultoras privadas que compiten para conseguir contratos de evaluación (todos buscando ingresos para pagar sueldos) ha llevado a unos niveles insospechados de calidad metodológica. Agencias con procedimientos como los que aquí describimos se ven obligadas a desempeñar un papel de control de calidad.

A esto se añade la necesidad de quienes trabajan en la Agencia de velar por la reputación del organismo, en especial porque ellos también han de rendir cuentas. El compromiso profesional de los cargos públicos no tiene como prioridad la pluralidad de valores, el rigor de la metodología o el derecho de la ciudadanía a la información: su prioridad son las necesidades inmediatas de sus superiores políticos. La burocracia es, necesariamente, la cultura del cumplimiento del servicio.

4. CONCLUSIÓN

Se pueden extraer dos lecciones principales de estas dos historias sobre la internalización de la evaluación. Una, que al evaluador se le niega la oportunidad de diseñar una evaluación comprometida, esto es, en términos “deliberativos y democráticos”. Los primeros teóricos de la evaluación, en su búsqueda de las funciones más apropiadas de ésta, la definían menos en términos de metodología y más como un *espacio* social y político único, dentro del cual se producían unas interacciones que no tenían cabida en ningún otro espacio social. Para ocupar ese espacio (de hecho, un espacio ético) la evaluación debe ser independiente. La independencia debe entenderse no como un privilegio del evaluador como actor político, sino como garantía al ciudadano de que hay un proceso político en marcha (que da cuentas a la sociedad) y al cual tienen acceso. Internalizar la evaluación equivale a denegar el libre acceso a sus resultados.

La otra lección es que la internalización de la evaluación expone al burócrata al escrutinio público y a la rendición de cuentas hasta un punto que, al final, es irracional e insostenible. El burócrata, por definición, no puede rendir cuentas por las decisiones metodológicas ni por la “inspección” metodológica y política que se deriva de su control contractual sobre la evaluación. Esto no le es de mucha ayuda, supone importantes trabas para el evaluador, como ya hemos visto, y niega al ciudadano la oportunidad de exigir responsabilidades a quienes toman las decisiones.

Si la sociedad ha de “aprender sobre sí misma”, deberíamos prestar atención a la relación entre el ciudadano y lo público, ya que las administraciones en estos momentos tienen prácticamente el monopolio de los medios para lograr ese aprendizaje. No fueron concebidas para asumir esa responsabilidad y, lógicamente, no lo llevan bien. Evidentemente, no está en nuestra mano, ni siquiera en democracia, reajustar esa relación, ya que la raíz del problema está en las disposiciones constitucionales, aun cuando la realidad contemporánea hace tiempo que las ha hecho irrelevantes y anticuadas.

Sin embargo, hay medidas que podemos y debemos adoptar como evaluadores para mejorar la situación. Sin duda, las asociaciones profesionales (de psicólogos, profesores, sociólogos y evaluadores) deben entablar un diálogo con las entidades financiadoras oficiales sobre la información pública y las obligaciones del gobierno a este respecto. Pero también deberíamos guiarnos, más de lo que lo hacemos, por los principios no formalizados sobre la ética de la evaluación. Un protocolo ético que regule el acceso a la información y a la publicación de la misma, que establezca las condiciones de confidencialidad y negociación del control de los datos, puede servir de base para llegar a un consenso sobre la delicada cuestión de cómo hacer pública la información. Por ejemplo, una cláusula que estipule que los individuos son dueños de la información sobre su persona o su trabajo, les daría al menos autoridad moral ante cualquier intento de control de un informe de evaluación. Este tipo de acuerdos se pueden negociar con la entidad financiadora con menor presión y mayor transparencia que durante el proceso de encargo y contratación. Los contratos a menudo también cambian los valores y las preferencias de la entidad financiadora y esta clase de acuerdos pueden satisfacer a todos los implicados en la medida en que logran recoger valores comunes.

Los acuerdos éticos tienen una ventaja adicional. Modifican, en cierta medida, la relación económica entre el ciudadano, el evaluador y la administración/entidad financiadora. Las relaciones económicas –como señaló Pettigrew (1994), “condiciones de contratación”– no son una buena vía para las prácticas éticas y las relaciones que se basan en valores. De hecho, recientemente, hemos asistido al estrepitoso fracaso de las relaciones económicas en las que durante más de una década nos convencieron para que nos considerásemos “clientes”, “consumidores”, “recursos humanos” o *stakeholders*, lo que nos condujo a una profunda crisis económica, de la que la cultura del contrato y los sistemas administrativos y regulatorios fallidos tienen mucha culpa. Debiera haber una autoridad moral que apoye el argumento de que las entidades públicas que financian la evaluación y la regulación desempeñen un papel más humilde y comprometido democráticamente.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Alkin, M., ed. (2004), *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, Ca: Sage Publications.
- Black, M. (2008), *The No-Nonsense Guide to the United Nations*. Oxford: New Internationalist.
- Cronbach, L.J. et al. (1980), *Towards Reform of Program Evaluation*. San Francisco: Jossey-Bass Publishers.
- Deaton, A. (2008), "Randomisation in the tropics, and the search for the elusive key to economic development", *The Keynes Lecture*, Londres: The British Academy, 9 de octubre.
- Franklin, T. (2008), "Reaching the Millennium Development Goals: Equality and justice as well as results", *Development in Practice*, 18(3): 420-423.
- House, E. (1979), "Technology versus craft: a ten year perspective on innovation", *Journal of Curriculum Studies*, 11(1): 1-15.
- (1980), *Evaluating with Validity*. Thousand Oaks, Ca: Sage.
- (1993), *Professional Evaluation*. Thousand Oaks, Ca: Sage.
- Kushner, S. (2009), "Own goals: evaluation and accountability in millennium development planning", en K.E. Ryan y J.B. Cousins, eds., *Sage International Handbook on Educational Evaluation*. Thousand Oaks: Sage.
- Kushner, S. y Norris, N., eds. (2007), *Dilemmas of Engagement: Evaluation and the New Public Management*. Oxford: Elsevier.
- MacDonald, B. (1976), "Evaluation and the control of education", en D. Tawney, ed., *Curriculum Evaluation Today: Trends and Implications*. Londres: Macmillan.
- (1981), "Mandarins and Lemons. The executive investment in program evaluation", ponencia presentada en el Congreso Anual de la *American Educational Research Association*, Los Angeles, 13-17 abril.
- Marsden, D. (2003), "Rights, culture and contested modernities", conferencia inaugural de la INTRAC 5th International Evaluation Conference, Holanda, abril.
- Norris, N. (1990), *Understanding Educational Evaluation*. Londres: Kogan Page.
- (1995), "Contracts, control and evaluation", *Journal of Education Policy*, 10(3): 271-285.
- O'Neill, O. (2002), *A Question of Trust*. Cambridge: CUP.
- Pettigrew, M. (1994), "Coming to terms with research: the contract business", en B. Halpin y B. Troyna, eds., *Researching Education Policy: Ethical and Methodological Issues*, Londres: Falmer.
- Picciotto, R. (2007), "The new environment for development evaluation", *American Journal of Evaluation*, 28(4): 509-521.
- Powers, M. (1999), *The Audit Society: Rituals of Verification*. Oxford: OUP.
- Rothschild Report (1971), Report of *The organization and management of Government R&D*, en *Cabinet Office. A Framework for Government Research and Development*. Londres: HMSO.
- Simons, H. (1995), 'The politics and ethics of educational research in England', *British Educational Research Journal*, 21(4): 435-449.

G^a_pP

EXPERIENCIAS Y CASOS

La evaluación de políticas públicas en el nivel estratégico: sistema general de becas, abandono escolar prematuro y mejora del capital humano

M.^a Luisa Merino Cuesta

Resumen: *Este artículo tiene como objetivo contribuir al conocimiento de la evaluación de las políticas públicas en el nivel macro a través de un encargo de evaluación del campo educativo encomendado a la Agencia de Evaluación de Políticas Públicas y Calidad de los Servicios (AEVAL), cuyo informe está disponible en su página web. El trabajo examina las peculiaridades que tiene la evaluación en el nivel estratégico y en el operativo de gobierno, y expone la dificultad de evaluar grandes conglomerados. Las lecciones aprendidas se exponen a través de la concepción que tiene la Agencia del propio proceso evaluador: análisis del encargo, planificación, implementación y resultados.*

Palabras clave: *Evaluación de políticas públicas en el nivel macro, evaluación estratégica, evaluación de grandes conglomerados.*

1. INTRODUCCIÓN

El trabajo que aquí se expone recoge las lecciones extraídas de un encargo del Consejo de Ministros relativo al Plan de Trabajo de la *Agencia de Evaluación de Políticas Públicas y Calidad de los Servicios* (AEVAL) en 2008. Presenta interés por su esencia de política pública, en el nivel macro y estratégico, y por la interrelación de objetos a evaluar que se encadenan, así como por la particularidad de ligar diferentes tipos de evaluación en un mismo encargo: “una doble evaluación del Sistema General de Becas educativas que deben contribuir a la disminución del abandono escolar prematuro, dentro de las medidas del eje 3 del Programa Nacional de Reformas (PNR) de España para el aumento y mejora del capital humano”¹.

El objetivo del encargo se sitúa en el centro del debate generalizado sobre resultados de la educación que tiene lugar desde hace años, tanto en nuestro país como

¹ Consejo de Ministros de 1 de agosto de 2008, por el que se aprueba el Plan de Trabajo AEVAL, 2008.

en el espacio global y europeo, considerando la educación y el aprendizaje como pieza fundamental en el desarrollo económico y social.

Como antecedentes de este largo debate podemos señalar una fecha y un informe clave. En 1964, el Congreso de los Estados Unidos encargó una macroencuesta nacional a J.S. Coleman para conocer las oportunidades educativas de los alumnos, independientemente de su grupo étnico-racial, religión u origen socio-cultural. La investigación implicó a seiscientos mil alumnos, sesenta mil docentes y cuatro mil centros educativos de Primaria y Secundaria de todo el país. El estudio ha pasado a la historia con el nombre de “Informe Coleman” (Coleman *et al.*, 1966) y ha sido considerado como uno de los más importantes del pasado siglo en las ciencias sociales. Generó desde su publicación un fuerte debate social y científico que en cierta medida sigue abierto. Cambió la investigación y la forma de ver la educación, incluso desde posiciones opuestas y, supuso, a la vez, el espaldarazo para un nuevo campo profesional: la evaluación de políticas públicas.

La mayoría de las cuestiones de las que se ocupaba ese Informe siguen siendo objeto de análisis y evaluación en la actualidad: naturaleza y extensión de las desigualdades en la escuela, papel de los centros educativos y diferencias en el rendimiento académico según los recursos educativos. Además, nuevas consideraciones y componentes se han incorporado a los estudios de educación y a su evaluación. Disciplinas como la economía, la sociología, la ciencia política, la psicología o el derecho han influido con sus análisis en la educación. Aspectos como la motivación de los alumnos, el profesorado y la función directiva, los procesos educativos y su incidencia en los resultados, la gestión de los centros, el papel de la administración al elaborar las normas o al brindar los servicios son elementos que interesan cada vez más en la evaluación educativa.

El valor de la educación en las sociedades actuales se enmarca en el contexto del creciente valor del capital humano como factor de competitividad en la economía global, donde la educación desempeña un papel predominante como transmisora de valores y conocimientos frente a los nuevos cambios sociales, culturales y tecnológicos.

En los países de la Organización para la Cooperación y el Desarrollo Económico (OCDE), los notables avances en materia educativa han supuesto una acumulación destacada de capital humano; si bien, este desarrollo ha venido acompañado, a la vez, de altas tasas de fracaso y abandono escolar prematuro, y de una falta de cualificación para hacer frente a los nuevos retos formativos y profesionales de este siglo.

La Agenda de Lisboa (2000) estableció importantes objetivos de formación del capital humano para el año 2010, poniendo especial énfasis en el problema del abandono escolar prematuro y su reducción al 10% al cumplirse la década. Para desarrollar e impulsar esa estrategia educativa, el Consejo Europeo adoptó el programa de trabajo *Sobre los futuros objetivos precisos de los sistemas de educación y formación* (Consejo de la Unión Europea, 2001).

El aumento de los recursos destinados a la educación plantea una mayor exigencia de la ciudadanía y la necesidad de rendir cuentas por parte de los poderes públicos. Como intervención pública, la educación ha de atender, de una parte, al enfoque educativo y, de otra, a elementos relacionados con el gasto público. Esto ha impulsado estudios sobre el rendimiento privado y social de la educación, la relación entre educación y movilidad social, los efectos redistributivos del gasto público o las relaciones de poder que se ocultan tras el currículo escolar. Las políticas educativas tienen una gran repercusión social y son de las más cuestionadas en el debate público.² Su evaluación es, hoy, ineludible para responder al cumplimiento de los fines que constituyen su razón de ser en función de las necesidades sociales.

La complejidad de los elementos que conforman el sistema educativo y los problemas sociales que van surgiendo en cada momento reclaman sistemas de evaluación cada vez más sólidos e integrales.

La evaluación del sistema de educación y formación en España fue establecida por la LOGSE en 1990³, con la creación del Instituto Nacional de Calidad y Evaluación (INCE); aunque no fue hasta 1993 cuando se regularon sus funciones⁴. La institución (en la actualidad Instituto de Evaluación) elabora planes plurianuales de evaluación general del sistema educativo, coordina la participación del Estado español en las evaluaciones internacionales y elabora el Sistema Estatal de Indicadores de la Educación. Las Comunidades Autónomas (CC.AA.) han creado también sus propias instituciones evaluadoras que realizan estudios sobre educación de diversa naturaleza.⁵ Por lo que respecta a la educación superior⁶, es la Agencia Nacional de Evaluación de Calidad y Acreditación (ANECA), la que tiene como objetivo contribuir a la mejora de la calidad del sistema mediante la evaluación, certificación y acreditación de enseñanzas, profesorado e instituciones de estudios superiores.

Las políticas educativas cuentan, por tanto, en España, desde hace tiempo, con instituciones encargadas de su evaluación. Sin embargo, los elementos que componen esas políticas son excesivamente complejos y requieren cada vez más evaluaciones integrales y de impactos. Las exigencias ciudadanas hacen necesario el desarrollo de estos nuevos enfoques. Además, cimentar una cultura de evaluación requiere de

2 El Informe del Observatorio de la Calidad de los Servicios Públicos (AEVAL, 2009) recoge que la educación es una de las políticas públicas con la que los ciudadanos están más insatisfechos.

3 “La administración educativa debe esforzarse continuamente por satisfacer las necesidades y las expectativas sociales; por ello la evaluación del sistema educativo se orientará a la permanente adecuación del mismo a las demandas sociales” (artículo 62.1 de la Ley Orgánica de Ordenación General del Sistema Educativo, LOGSE). Sobre el proceso de institucionalización anterior ver García Sánchez (2004).

4 Mediante Real Decreto 928/1993, de 18 de junio.

5 Sobre el desarrollo de las instituciones evaluadoras en las CC.AA., véase INCE (1999). Para los distintos tipos de evaluación llevados a cabo puede consultarse *Apuntes del Instituto de Evaluación* (MEPSYD, 2007).

6 Iglesias Ussel *et al.* (2009). Sobre la calidad de las universidades, la revista *Gestión y Análisis de Políticas Públicas* dedicó a este tema, en 1998, con carácter monográfico, el número 11/12.

evaluaciones externas a los responsables del sistema. Por ello, nuevas instituciones vienen a sumarse a la evaluación de los resultados educativos desde una perspectiva macro, externa y más estratégica. Algunos expertos hacen la distinción entre el nivel sistémico o macroeducativo (el conjunto del sistema y en particular las políticas públicas educativas), el nivel institucional o mesoeducativo (instituciones universitarias, centros y programas educativos) o el microeducativo (el profesorado, la función directiva y los alumnos) (Puente Azcutia, 1999).

La Agencia de Evaluación de Políticas Públicas y Calidad de los Servicios (AEVAL)⁷ adscrita recientemente al Ministerio de la Presidencia⁸, recibe encargos de evaluación de cualquiera de las políticas públicas sobre las que el Ejecutivo quiera conocer. Se trata de una institución de reciente creación cuyos encargos se aprueban anualmente por el Consejo de Ministros, centrándose, por el momento, en la evaluación de las principales medidas puestas en marcha para la consecución de los objetivos de Lisboa. Sus encargos en el campo de la evaluación educativa se han centrado en cuestiones críticas que preocupan a la sociedad: reducción del fracaso, abandono escolar prematuro y en el nivel estratégico de la formulación de nuevas intervenciones públicas. Su cometido abarca cuestiones transversales que afectan a varios Ministerios y a la política de convergencia con la UE⁹. Pretende sumar su acción al mayor conocimiento que buscan los gobiernos sobre las políticas –en este caso de la educación– en la nueva sociedad y en los nuevos tiempos. La Agencia responde ante el ejecutivo y actúa como órgano externo e independiente del Ministerio de Educación.

En nuestro país, las transformaciones educativas de los últimos 25 años han supuesto –hay que considerarlo– un gran avance en relación con nuestro desfase histórico. El *Panorama de la educación* (2009)¹⁰ que ofrece la OCDE, aunque es similar al de años anteriores, deja vislumbrar cambios esperanzadores. Sin embargo, siguen persistiendo importantes retos para el sistema. España presenta uno de los peores resultados de la UE en abandono escolar prematuro, con una tasa que ha oscilado en la última década entre el 29,1% y el 31,7%, lo que supone un claro obstáculo para incorporarse a la sociedad del conocimiento y para el logro de una mayor cohesión social. Por ello, España se ha fijado reducir el abandono escolar prematuro hasta el 15% en 2012. Entre las medidas barajadas por los responsables de la educación para alcanzar ese objetivo se plantea la posibilidad de que nuevas modalidades de becas puedan contribuir a la reducción del abandono escolar prematuro.

7 La AEVAL fue creada por la Disposición Primera de la Ley 28/2006 de 18 de julio, de Agencias Estatales para la mejora de los servicios públicos. El Real Decreto 1418/2006 de 1 de diciembre aprobó el Estatuto de la Agencia de Evaluación de Políticas Públicas y Calidad de los Servicios.

8 El Real Decreto 1039/2009 adscribe la Agencia al Ministerio de la Presidencia que en origen estaba adscrita al extinto Ministerio de Administraciones Públicas.

9 Su papel está claramente diferenciado de la función de seguimiento, como se recoge en el documento *Convergencia y Empleo. Programa Nacional de Reformas* (PNR, 2005) que encomienda el seguimiento de la Estrategia de Lisboa a la Oficina Económica del Gobierno y la evaluación de la implementación y logro de las medidas a la AEVAL.

10 Se puede ver *Education at a Glance/Regards sur l'Éducation* (2009). La versión en español, *Panorama de la educación. Indicadores de la OCDE 2009* está a disposición del público en el Ministerio de Educación.

Los estudios sobre la política de becas en los últimos años sostenían que: “España tiene pocas becas y de una cuantía relativamente reducida, lo que dificulta que cumplan su papel” (San Segundo, 2006). De ahí que la política educativa del período 2004-2008¹¹ se propusiese impulsar un buen número de medidas encaminadas a mejorar la política de becas: a) cualitativas: la beca como derecho y b) cuantitativas: aumento de cuantías, umbrales, partidas presupuestarias y número de becarios.

El impulso de la política de becas juega un papel crucial en la búsqueda constante de la igualdad de oportunidades educativas, económicas y sociales, y su utilización al servicio de la reducción del abandono quedó plasmada en el *Programa Nacional de Reformas* en España (PNR) de 2005¹².

En la base de las intervenciones con becas para la resolución del abandono escolar prematuro se encuentra el compromiso político expresado por altos responsables del Ministerio de Educación de que ningún alumno abandone por falta de recursos, así como reforzar la equidad en el mantenimiento y resultados del sistema, no sólo en el acceso a él.

Una buena política de becas y ayudas al estudio potencia la eficiencia al promover el desarrollo de todo el capital humano del país y al fomentar la dedicación intensiva a los estudios (Díaz Malleo y San Segundo, 2000).

No obstante, y a pesar de la voluntad declarada del actual gobierno cuyos compromisos se han visto concretados en los aumentos presupuestarios destinados a becas, el Informe del *Observatorio Social de España* (2007) concluía: “España tiene un gasto en becas, como porcentaje del gasto público en educación por nivel, equivalente a la mitad de la media de la UE-27”; si bien, este dato debe ponerse en relación con el global de la financiación de la educación en España, ya que en nuestro país prima frente a otros la subvención directa a través de unos gastos de matrícula muy inferiores al coste real.

Esta exposición se plantea bajo una doble perspectiva:

- a) Extraer lecciones en relación con lo que la evaluación enseña sobre las políticas evaluadas; en este caso las de educación.
- b) Extraer lecciones del proceso evaluador: lo que la evaluación de políticas educativas enseña sobre la propia evaluación.

¹¹ Véase Tiana (2007, 2008).

¹² Eje 3. Capital Humano. Apartado 6. Medidas: “Se continuará ampliando la política de becas y ayudas al estudio a fin de estimular la continuidad de los jóvenes en los estudios postobligatorios de bachillerato, formación profesional y educación superior. Estas ayudas deben contribuir a disminuir el abandono escolar prematuro y permitirán mejorar las tasas de alumnos graduados en educación secundaria superior. La ampliación afectará en 2006 a cuarenta mil nuevos becarios y el incremento total previsto del gasto destinado a becas superará los 85 millones de euros”.

Se tratará de equilibrar ambos enfoques, resaltando las cuestiones de relevancia del segundo dada la dedicación monográfica de esta publicación al campo de la evaluación. En cualquier caso, el lector tendrá la posibilidad de profundizar más a través de la lectura completa del Informe *Evaluación del sistema general de becas educativas. Primera Evaluación: Diagnóstico de la situación actual y principales alternativas para mejorar su eficacia* (AEVAL, 2009)¹³.

La exposición pondrá especial énfasis en clarificar las diferencias que existen entre evaluar políticas públicas y evaluar programas o dicho de otra manera, evaluar objetos macro y objetos mezo o bien, evaluar objetivos estratégicos u operativos, cuestiones que siguen generando cierta confusión.

Por último, habría que destacar la concepción evaluadora de la Agencia que considera la evaluación como una intervención política, como tal, responde a las tradicionales fases de análisis, planificación, implementación y resultados¹⁴. Es, precisamente, este planteamiento el que servirá de estructura a las lecciones extraídas.

2. PRESENTACIÓN Y PLANTEAMIENTO DEL TRABAJO EVALUATIVO

2.1. El encargo

Se trata de una evaluación doble en el tiempo que entrelaza objetivos estratégicos y operativos, y plantea una modalidad de objeto de investigación¹⁵ bastante peculiar. Se encadenan tres de naturaleza diferente: el sistema general de becas –instrumento financiero de las políticas de compensación de las desigualdades en educación– el problema del abandono escolar prematuro (AEP) –considerado como objetivo educativo prioritario de la Estrategia de Lisboa, reconocido por el Ministerio de Educación y las autoridades educativas de las Comunidades Autónomas como problema de Estado– y por último, la mejora del capital humano –pilar básico para conseguir la igualdad de oportunidades en nuestra sociedad y objetivo estratégico de crecimiento económico según establece el PNR.

— Plazo de entrega del encargo

El encargo de evaluación explícita, no sólo las fechas de entrega sino el doble cometido: a) la evaluación 2008, requería un diagnóstico del sistema general de be-

13 Disponible en internet en el sitio web de AEVAL:
<<http://www.aeval.es/comun/pdf/evaluaciones/E11.pdf>>.

14 Por artificial que esa concepción pueda parecer a algunos analistas no deja de tener una utilidad práctica nada desdeñable.

15 Tómesese en el sentido común del término, como sinónimo de estudio y no en el específico de objeto de investigación. En puridad, lo estricto sería hablar de objeto de evaluación, pero para refrescar el texto con otros sinónimos se utilizará indistintamente objeto de estudio, de investigación o de evaluación, con la salvedad de tener bien presente las diferencias que hay entre las distintas prácticas.

cas, así como la propuesta de nuevas modalidades de ayudas que pudieran contribuir a la disminución del abandono escolar prematuro; b) en un período posterior, se planteaba una evaluación *ex ante* de las alternativas seleccionadas.

— *El sistema general de becas*

Las políticas de becas se incardinan dentro de las políticas de educación compensatoria.

El Capítulo II del Título II de la Ley Orgánica 2/2006, de 3 de mayo, de Educación recoge las acciones compensatorias que corresponderán a las Administraciones públicas con el fin de hacer efectivo el principio de igualdad en el ejercicio del derecho a la educación.

Entre las políticas de educación compensatoria que reforzarán la acción del sistema educativo para evitar las desigualdades derivadas de factores sociales, económicos, culturales, geográficos, étnicos o de otra índole se encuentran las becas y ayudas al estudio (art. 83 de la citada Ley).

Es el Estado, según las leyes, el encargado de establecer, con cargo a sus Presupuestos Generales, un sistema general de becas y ayudas al estudio con el fin de que todas las personas, con independencia de su lugar de residencia, disfruten de las mismas condiciones en el ejercicio del derecho a la educación.

El Sistema General de Becas y Ayudas al Estudio se plantea un doble objetivo explicitado en el Real Decreto 1721/2007, de 21 de diciembre: a) lograr un sistema de becas y ayudas al estudio que garantice la igualdad en el acceso a las becas en todo el territorio y b) dar cumplimiento a la jurisprudencia del Tribunal Constitucional en esta materia (Sentencias 188/201, de 20 de septiembre, y 212/2005, de 21 de julio), todo ello manteniendo la eficacia y eficiencia del sistema.

— *Nuevas competencias de las CC.AA. en materia de becas*

La normativa básica del sistema de becas venía siendo competencia de la Administración General del Estado¹⁶, pero a partir de las sentencias citadas en el párrafo anterior se otorga a las Comunidades Autónomas la capacidad para desarrollar la normativa básica en materia de becas y ayudas, así como para realizar su convocatoria, gestión, resolución, adjudicación, pago y resolución de recursos.

¹⁶ La única Comunidad Autónoma con competencias plenas en materia de becas es el País Vasco, sin perjuicio de que tenga en cuenta en sus convocatorias los requisitos y umbrales mínimos establecidos con carácter general por el Estado.

En cualquier caso, dicha descentralización garantizará lo dispuesto en la Ley Orgánica 2/2006 de Educación, de 3 de agosto respecto al sistema general de becas: que todas las personas, con independencia de su lugar de residencia, disfruten de las mismas condiciones en el ejercicio del derecho a la educación.

El proceso de descentralización de la gestión del sistema de becas para el resto de las Comunidades se inició recientemente y plantea la necesidad de equilibrar esa mayor asunción de competencias por las CC.AA. con la vigilancia y control del principio de equidad por parte de la Administración central.

Asimismo, con el fin de articular un sistema eficaz de verificación y control de las becas y ayudas concedidas se deberán establecer los procedimientos necesarios de información, coordinación y cooperación entre las diferentes Administraciones educativas.

— *El abandono escolar prematuro*

El abandono escolar prematuro es un fenómeno complejo que se focalizó, en su origen, dentro del ámbito psicológico interpretado como desventaja individual del alumno, inadaptación o fobia escolar y que pasó, seguidamente, a considerarse un problema del ámbito educativo que debía atender a cuestiones como el refuerzo de las capacidades cognitivas de los alumnos y a factores que influyen en su rendimiento: calidad y diseños curriculares de la enseñanza. A esto se fueron añadiendo otras consideraciones como la organización y clima de los centros, control y disciplina de los alumnos, valor social de la educación, valor concedido por los alumnos, etc. En las últimas décadas, el problema empezó a tener en consideración múltiples factores en relación con las condiciones sociales y económicas y la relación entre los criterios de equidad y eficiencia aplicados a la educación, cobrando mayor interés la incidencia del AEP en la capacitación del capital humano y su repercusión en el proceso productivo que es el último objetivo al que alude la evaluación.

Aunque está constatado ampliamente que las diferencias sociales y culturales de los alumnos condicionan su progreso educativo y los resultados que obtienen (Informe de la OCDE-CERI, 1995), no está tan clara la correspondencia estricta entre las desigualdades sociales y las desigualdades educativas. Hay otros factores, como la familia, el funcionamiento del sistema y la propia escuela que pueden incrementar o disminuir estas desigualdades. En los últimos tiempos el acceso a internet se ha convertido ya en una poderosa causa de desigualdad.

El Abandono Escolar Prematuro (AEP) es un fenómeno sobre el que existen múltiples aproximaciones parciales, pocas investigaciones generales y en él se detecta, sobre todo, un déficit de análisis sistemáticos y rigurosos desde una perspectiva integral.

Entre las aportaciones recientes y destacadas cabe mencionar el *Estudio sobre Abandono Escolar Prematuro* encargado en 2005 por la Dirección General de Educación y Cultura de la Comisión Europea a la consultora GHK, cuyas conclusiones han orientado muchas de las intervenciones europeas en este campo¹⁷.

La lucha contra el abandono requiere de un esfuerzo conjunto de la Administración General de Estado y de las CC.AA., siendo de celebrar la Conferencia Sectorial de Educación que se celebró con carácter monográfico sobre el tema, convocada el 20 de noviembre de 2008. Como resultado de la misma se aprobó un Plan para reducir el Abandono Escolar a la mitad en 2012. Dicho Plan comprende un conjunto de 12 medidas y estará dotado con 121 millones de euros en 2009. Sobre el detalle de las medidas se remite al Informe de evaluación. La única salvedad que aquí se expone es que ninguna de esas medidas hace referencia a las becas.

— *Capital Humano*

Los buenos resultados educativos constituyen una riqueza social en sí misma y, a la vez, son una fuente esencial del capital humano de un país. En este sentido, algunos estudios¹⁸ atribuyen a la educación el crecimiento del capital humano en España en las últimas décadas.

Las inversiones realizadas en el sistema educativo español como la propia conciencia ciudadana frente a los estudios para adquirir mayor cualificación, ha provocado un incremento significativo en las dotaciones del capital del humano de las distintas regiones del país. De ahí la importancia de cuantificar el valor del capital humano generado por el sistema educativo y de las cuestiones que trata de relacionar el último objetivo de esta evaluación.

2.2. Enfoque evaluador, diseño y metodología

— *Enfoque*

El planteamiento evaluador responde a un enfoque de políticas públicas en su nivel macro y estratégico, combinando estudios de distintos campos disciplinares con la planificación estratégica, sin perder de vista nunca el campo administrativo en el que se originan los principales discursos y medidas educativas, y teniendo en cuenta la formulación nacional y europea.

¹⁷ El resumen ejecutivo de dicha Comunicación está disponible en español en *Revista de Educación* (2006).

¹⁸ Véase los diferentes estudios sobre Capital Humano elaborados por el Instituto Valenciano de Investigaciones Económicas (IVIE), algunos disponibles en su sitio web: <<http://www.ivie.es/banco/caphum/caphum.php>>.

— *Diseño*

El estudio parte de un diseño en el que se traza un eje de abscisas en el que figura el *Estudio sobre Abandono Escolar Prematuro* realizado por la Comisión en el año 2005 y otro de ordenadas en torno a datos nacionales proporcionados por el Instituto Nacional de Estadística (INE) en la Encuesta de Condiciones de Vida (ECV) de 2006 y la Encuesta de Transición Educativo-Formativa e Inserción Laboral (ETEFIL) de 2005, con el fin de ver los puntos de intersección de las variables y de las conclusiones europeas y los resultados que ponía en evidencia la evaluación.

En cuanto a la población sobre la que actuar con posibles intervenciones se ha tenido en cuenta un doble abordaje: “intervenciones para prevenir el AEP” –dentro de las etapas del sistema educativo– e “intervenciones para disminuir el AEP” –fuera de las etapas del sistema educativo.

Por último, se han estimado ineludibles los análisis comparativos, tanto para las becas como para el examen del problema de AEP. En el caso de las becas se han examinado los países que figuraban en las primeras posiciones del ranking educativo: Reino Unido, Holanda y Suecia. En cuanto al AEP, se ha seleccionado Finlandia (Julku-nen, 2007) porque es el de menor tasa de AEP en la UE y todos ellos por su posición en el *Programa Internacional para la Evaluación de Estudiantes* (PISA).

En todo momento, los análisis comparativos con otros países han tenido en cuenta el contexto nacional y, para el examen final de posibles intervenciones con becas se plantearon criterios de oportunidad política, social y económica que se examinarán en la segunda fase de la evaluación.

La concepción del trabajo plasmada en el Informe de evaluación se inicia con una parte introductoria que contextualiza, en el marco nacional y europeo, las intervenciones y cuestiones objeto de estudio: becas, AEP y Capital Humano, para pasar a continuación a describir con detalle los objetos a evaluar. Una vez establecidos los criterios que servirán de referentes, se plantea la estrategia evaluativa partiendo de las principales preguntas e hipótesis de trabajo, completando el diseño con las fuentes de información requeridas y la selección de técnicas a aplicar. Determinadas todas estas cuestiones se abre la fase de implementación en la que, tras la recogida de los datos, su tratamiento y análisis, desemboca en los resultados, con las conclusiones y recomendaciones del Informe (AEVAL, 2009).

— *Base metodológica del estudio*

La metodología ha sido abierta y pluralista¹⁹. Flexible en cuanto a las herramientas, combinando diferentes técnicas: el análisis documental (normativa básica y específica y literatura sobre la materia), la explotación de encuestas: principalmente, la

¹⁹ Se ha contado con un buen número de actores implicados en la política a evaluar, aunque el grado de participación en la evaluación haya sido muy desigual por razones que exceden este artículo.

Encuesta sobre Abandono Escolar Prematuro, encomendada por la Dirección General de Educación y Cultura de la Comisión Europea (DG EYC, diciembre, 2004-septiembre, 2005) realizada a numerosos países tanto de la UE como otros de la OCDE, la Encuesta de Condiciones de Vida (ECV) de 2006 y la Encuesta de Transición Educativo-Formativa e Inserción Laboral (ETEFIL) de 2005; y fuentes secundarias (estadísticas del Ministerio de Educación, Eurostat, OCDE, etc.).

La evaluación encargó también estudios propios²⁰ como uno de investigación, basado en modelos probabilísticos, sobre la eficacia del sistema de becas y grado de cobertura de las poblaciones objetivo según los umbrales de renta, en el que a través de la técnica de análisis bivariante o multivariantes sobre los datos de la ECV (con atención especial a la educación postobligatoria) se analizó la equidad del sistema, la probabilidad de ser becario, el cumplimiento de los requisitos legales, la distribución de becarios por grupos de renta, el perfil de becarios en deciles de renta baja, el análisis de fraude, el de potenciales becarios que no hacen uso de las becas y el perfil de los individuos que abandonan o permanecen en el sistema. El estudio se aplica a una muestra de 3651 individuos con edades de 17 a 25 y utiliza los datos de la ECV de 2006 y la referencia de variables de renta de 2005.

Además, se encargó otro estudio de *benchmarking* entre diferentes países europeos, analizando las características de algunos que, como Finlandia, presenta el mayor éxito en el Informe PISA y las tasas más bajas de abandono. Y se profundizó aún más en las experiencias de Holanda, Reino Unido y Suecia, con el fin de extraer información sobre las becas y ayudas que logran más éxito en estos países.

Se solicitó también un trabajo sobre el estado de la cuestión del Capital Humano en relación con los objetivos de la evaluación y se realizaron entrevistas a los agentes sociales, responsables de educación y expertos en materia educativa y otros vinculados con el sistema nacional de becas, tanto del Ministerio de Educación, Política Social y Deporte (MEPSYD) como del Ministerio de Ciencia e Innovación (MICINN) y de las CC.AA. Finalmente, se organizó un panel de expertos en materia educativa con los representantes de las instituciones más relevantes para “contrastar”²¹ las conclusiones del diagnóstico de la primera evaluación.

3. LECCIONES APRENDIDAS

3.1. En relación con el encargo

La primera lección a destacar en este encargo está en la distinción que supone evaluar “políticas públicas” o evaluar “programas”. Sobre este asunto convendría

²⁰ La coordinación de estos estudios estuvo a cargo del profesor Jorge Calero, experto en política educativa y gasto público.

²¹ Las comillas tratan de llamar la atención sobre el objetivo que perseguía la técnica del panel que fue contrastar los resultados obtenidos en la evaluación, nunca consensuar conclusiones con los participantes.

hacer una aclaración antes de continuar. ¿A qué llamamos evaluación de políticas públicas? En realidad, “todo lo que los gobiernos deciden hacer o no hacer constituye una política” (Dye, 1975; también Bachrach y Baratz, 1963). Según esta concepción se podría argumentar que todos los objetos a evaluar son políticas públicas. Sin embargo, no es lo mismo enfrentarse al “mandato” del Consejo de Ministro que al encargo de un administrador de programas. No es lo mismo que la evaluación haya de servir a los niveles estratégicos que a los operativos. Para examinar con mayor detalle estas diferencias, nos serviremos de la doble distinción que hace Dror (2003) de los niveles de acción pública, distinguiendo el correspondiente a:

1. *Funciones de orden superior (o macro)*: que se corresponde con la modificación de las trayectorias colectivas mediante decisiones que, en esencia, constituyen intervenciones en el proceso histórico y que son, cualitativamente, las más importantes de la acción de gobierno. Tienen su base en el binomio Estado-sociedad y requieren de los evaluadores buenos conocimientos de la ciencia política y administrativa.

Por el momento, este tipo de acciones han sido las menos evaluadas.

2. *Funciones de ejecución, gestión y servicios de los gobiernos (mezzo y micro)*: que se refiere al despliegue y a la operatividad de las políticas en objetos más concretos: desde los programas y proyectos a las actividades prestacionales y el funcionamiento de la gestión. Se trata de funciones cuantitativamente más numerosas basadas en el binomio: servicios-usuarios, que requiere de los evaluadores enfoques, organizativos y económicos.

Aunque esta diferenciación pueda resultar demasiado simple: “una política pública no es un fenómeno objetivo dotado de un perfil evidente y bien definido” (Hecló, 1972: 82), creemos que permite avanzar por una senda conceptual que está planteando confusiones. En todo caso, hablaremos de evaluación de políticas cuando el objeto sea de nivel estratégico y de evaluación de programas cuando el objeto sea de nivel operativo.

El nivel de gobierno que corresponde a las funciones de orden superior se preocupa más de responder a criterios de legitimidad con la ciudadanía y se plasma en la formulación de las políticas, sean globales sobre una materia, transversales o estratégicas. El segundo nivel, se centra más en criterios de funcionalidad y atañe a la parte operativa y de gestión de las políticas. El primero permite medir la calidad democrática y el segundo la calidad de las organizaciones y de los servicios públicos –asistidos, generalmente, por programas. La suma de la calidad en un nivel y otro dan como resultado la calidad de las políticas públicas.

Si bien el proceso evaluador y sus fases no varían, se trate de evaluar políticas o de evaluar programas, lo que varían son los pesos entre ellas. El análisis del contexto y la elección del enfoque tienen un mayor peso al evaluar el nivel macro de las políticas públicas y requieren conocimientos de la ciencia política y administrativa, mientras que en la evaluación de programas ha sido más frecuente el enfoque de calidad y el económico

como respuesta a la cultura de gestión de calidad y de resultados. Los resultados en el nivel estratégico son resultados sociales, que engloban cuestiones más allá de la gestión. Como apuntábamos antes, varían también las preguntas que interesan a un nivel u otro y el tipo de información que se requiere. Otras diferencias son las que afectan al entramado de criterios y su articulación entre niveles y, particularmente, las recomendaciones que habrán de estar en la línea del interés que requiere un Consejo de Ministros.

— *Las particularidades de los encargos del Consejo de Ministros*

Los encargos del Consejo de Ministros tienen para la Agencia la particularidad de actuar como “mandatos”²² y eso supone que la negociación evaluativa –esencial para precisar y delimitar el objeto, objetivos, necesidades informativas, usos de la evaluación, etc.– no puede hacerse directamente con los que realizan el encargo, sino con instancias administrativas que actúan por delegación, a veces en total sintonía con el mandato y otras, no tanto. Esto supone mayor independencia para el desarrollo de la evaluación pero también algunas dificultades a la hora de definir las preguntas o las necesidades informativas requeridas a un nivel político más sofisticado.

Eso obliga a explotar al máximo las intervenciones parlamentarias de los ministros, las comparecencias en los medios y cualquier documento que pueda dar pistas sobre las intenciones gubernamentales en la materia objeto de evaluación. Todo ello con el fin de plantear las preguntas de mayor interés para ese nivel de gobierno y analizar si las intervenciones puestas en marcha, o algunas de las previstas, son las adecuadas para los objetivos que se persiguen.

Se completa esta vía exploratoria con entrevistas a los responsables educativos que tienen encomendado el desarrollo e implementación de los objetivos políticos objeto de evaluación.

Este tipo de encargo sitúa la acción evaluadora –aunque sea considerándola una ayuda más entre otras para la decisión– en el centro de los problemas de estado, justificándola como palanca de gobierno (Subirats, 2005).

Un encargo como el de becas-abandono-capital humano supone el reconocimiento para la Agencia de su esencia transversal y de su campo de evaluación más específico: el nivel macro de la acción pública.

— *La importancia de los factores políticos en el desarrollo de la evaluación de políticas públicas*

Los factores políticos tienen mayor peso en la marcha de la evaluación de las políticas públicas que en la evaluación de programas. La legislatura anterior ha ofrecido

22 Ese mandato supone “obligación de hacer” en relación con la propia misión de la Agencia.

algunos ejemplos de ello. La remodelación de gobierno del verano de 2008 adjudicó las competencias universitarias al MICINN. Esto afectó al alcance de la evaluación que en su origen, primer trimestre del 2008, había incluido en su estudio las becas universitarias, competencia entonces del MEPSYD, que era quien solicitó la evaluación.

Un cambio posterior, con la remodelación gubernamental del mes de abril de 2009, supuso un movimiento de retorno. Los estudios universitarios volvieron al seno del Ministerio de Educación y esto afectó de nuevo al alcance del estudio que debía volver a incluir las becas universitarias.

Toda esta movilidad que afecta a la evaluación de políticas públicas, no afecta de igual manera a la evaluación de programas. El factor político es tan importante en el caso de la evaluación de políticas públicas que, con el nombramiento de nuevos responsables ministeriales tanto se puede impulsar más la evaluación como puede quedar en vía muerta. Sin embargo, en el nivel operativo, los programas de gestión de las becas universitarias, continuaron funcionando al margen de los cambios e independientemente del Ministerio al que estuvieran adscritos los servicios que los gestionaban.

— *Evaluación de carácter estratégico/ evaluación de carácter operativo*

La evaluación de la que aquí se trata es de carácter estratégico en razón de la utilidad que persigue: verificar los logros en los objetivos de educación 2010 del MEPSYD, en particular, en su lucha contra la reducción del AEP a través de las becas, así como evidenciar el logro de los objetivos de convergencia de la Estrategia de Lisboa. Además, los resultados perseguían orientar la formulación de futuras intervenciones.

“En el mundo de las Administraciones Públicas, el concepto de estrategia no ha de entenderse como respuesta a la competitividad entre instituciones y empresas, sino como una serie de acuerdos explícitos que reflejan las preferencias surgidas del debate ente los actores colectivos e individuales presentes en el marco de juego de esas Administraciones” (Stewart y Ranson, 1988). La estrategia apunta a los objetivos que pretenden alcanzar los gobiernos para mejorar las sociedades y suele plasmarse en Planes Estratégicos que difieren claramente de los Planes operativos en los que se insertan los programas que orientan su acción a la gestión.

Esta naturaleza estratégica, como se señalaba anteriormente, va a determinar muchas de las cuestiones que afectan a la evaluación: preguntas, criterios, indicadores, tipo de información que se requiere y, por supuesto, las recomendaciones. Las cuestiones estratégicas apuntan a una comprensión y planteamientos de índole diferente. Los resultados de la evaluación no sólo servirán para informar o mejorar los diversos niveles de gobierno, sino para apoyar las posibles intervenciones que requieren los nuevos problemas. Esta visión estratégica afectará, como es lógico, a la composición de los equipos evaluadores para los que se requerirá diferentes *expertizajes* y visiones más holísticas.

Hay que señalar que tanto la UE como las instituciones internacionales de evaluación reivindican desde hace tiempo la necesidad de realizar evaluaciones estratégicas y más integrales. Este tipo de evaluación es el menos explorado hasta el momento y supone todo un desafío para la AEVAL.

El caso que aquí se expone resultó interesante al incluir el mandato de evaluación resultados de utilidad para los dos niveles de gobierno: la exploración de un instrumento financiero como es la política de becas en la resolución de un problema social tan crítico como el abandono escolar prematuro.

El diagnóstico sobre el Sistema General de Becas planteaba al evaluador cuestiones operativas, de tipo prestacional y de gestión, pero sin perder de vista, que el estudio del diagnóstico solicitado funciona, a la vez, como el análisis DAFO de la formulación estratégica. Por ello, los resultados de la gestión del Sistema de Becas, fueron una de las conclusiones de la evaluación pero nunca el objetivo último. Se trataba más bien de conocer las fortalezas y debilidades del Sistema para poder actuar estratégicamente sobre la disminución del AEP. Queda claro, en este ejemplo, el intento de dar respuesta al binomio Estado-sociedad, a través de nuevas intervenciones con becas que lo que pretenden es introducir transformaciones y mejoras en el proceso histórico-educativo y, contribuir, a la vez, con ese instrumento a alcanzar los objetivos estratégicos de la política educativa europea. Por lo mismo se espera que quede claro, también, el mejor uso que pueden hacer los evaluadores al utilizar los conceptos “evaluación de políticas” y “evaluación de programas” o “evaluación de proyectos” –nivel aún más desagregado.

En cuanto a las herramientas de la evaluación de políticas públicas en el nivel estratégico, resulta obvio el fuerte peso, junto a otras técnicas de investigación social, del análisis de políticas públicas, así como de herramientas propias de la planificación estratégica.

No es que en la evaluación de programas no tenga importancia la planificación de futuro, sin embargo, no llega a alcanzar la dimensión estratégica a la que apunta la evaluación de políticas públicas en cuanto a propiciar transformaciones sociales²³.

— *“Evaluación de grandes conglomerados”*

Se trata de un mandato rico para el aprendizaje del análisis político, al plantear un conglomerado de acciones públicas que conforman sistemas que se interrelacionan unos con otros. La riqueza no está sólo en la evaluación de un campo complejo por sí mismo, como el de la política educativa, ni en el cruce con otras políticas conexas, las políticas de trabajo –este tipo de complejidad resulta ineludible ya en

²³ Estamos considerando la evaluación de programas en su faceta operativa, sin obviar programas que puedan tener una importancia estratégica en el nivel que venimos señalando.

cualquier evaluación de políticas públicas. El mayor valor del encargo está en el nivel en el que se sitúa el problema del estudio (estratégico y de convergencia) y la vinculación de tres objetos de distinta naturaleza. El mandato desborda los habituales encargos de evaluación de programas educativos concretos por complejos que estos sean.

En este caso se trata de un macro objeto conformado por tres anillos: el sistema de becas como instrumento financiero, el abandono escolar como problema social y el capital humano y su mejora como resultado de la interrelación. Círculos que no están perfectamente delimitados sino que encajan como las piezas de un puzzle. En razón a esta peculiaridad denominamos a este tipo de evaluación “evaluación de grandes conglomerados”.

— *Evaluación de políticas sociales: la política educativa*

El encargo se enmarca en el campo de “las políticas sociales”²⁴, temática educativa.

La evaluación de políticas educativas supone para el evaluador situarse en el campo disciplinar en el que surgió la evaluación, y enfrentarse a una de las políticas sociales con mayor legado en la que el proceso de aproximaciones sucesivas al objetivo deseado ha sufrido numerosas reconsideraciones y hasta enfoques opuestos para los que se viene reclamando un pacto de Estado²⁵. Se trata, también, en educación, de una política cuyas transformaciones necesita para dar resultados el largo plazo y que ha estado en las últimas décadas en continuo movimiento, definiendo y redefiniendo objetivos y criterios.

Tienen, además, estas políticas, una fuerte carga ideológica que se polariza en la defensa de dos principios: la equidad –para los gobiernos socialdemócratas– y la calidad-eficiencia –para los gobiernos liberales. Equidad entendida como igualdad de oportunidades para todos, independientemente de las diferencias socioeconómicas y eficiencia como aprovechamiento máximo de los recursos invertidos para obtener los objetivos deseados.

Aunque recientes informes de la UE concluyen que ambos conceptos deben aunar, el foco que cada orientación política pone en uno u otro principio está en la base del tipo de sociedad que se quiere construir y resulta un arma partidista a la hora

²⁴ Sobre la evaluación de las políticas sociales y su complejidad ha servido de referente el artículo ya clásico del profesor Subirats (2005) “*Catorce puntos esenciales sobre evaluación de políticas públicas con especial referencia al caso de las políticas sociales*”; si bien su ejemplo se centra en el campo de la inclusión social.

²⁵ En esta línea se viene trabajando y el Gobierno y la oposición aparecían a principio del otoño en los medios iniciando un nuevo diálogo para ese pacto educativo que tanto reclama la sociedad.

de enjuiciar los resultados de la educación. Sin embargo, las conclusiones del informe de la Comisión Europea concluyen que ambos objetivos se refuerzan mutuamente. Quizá no en el corto plazo, pero sí al considerar el impacto de la enseñanza dentro de un marco social lo más amplio y duradero posible.

Sin embargo, el abandono escolar prematuro se explica pero no se resuelve simplemente a través del binomio anterior, puesto que hay causas que están en la base de la propia estructura del sistema o en el desarrollo de sus políticas que plantean que las intervenciones no son suficientes: políticas de integración, política de becas, de gestión de centros, programas de apoyo, etc. Sin embargo, hay causas personales que vienen a sumarse y que afectan más a unos individuos que a otros aunque se hallen en situaciones similares. Y esto enlaza con una de las particularidades de las políticas sociales cada vez más evidente: la diversidad.

A la vez existen contradicciones difíciles de encajar cuando se establecen comparaciones con otros países. España obtiene en el Informe PISA una de las puntuaciones más favorables en cuanto a la equidad de su sistema educativo, sobre la que insiste el último informe de la OCDE, 2009: “los alumnos españoles con entornos menos favorecidos tienen mejores resultados que sus homólogos de la OCDE (...) hay menos diferencia entre los mejores y los peores resultados, o lo que es lo mismo, en España hay mayor equidad”, destacando que “la esperanza de vida escolar española ha aumentado en los últimos años hasta alcanzar un valor similar al de los promedios OCDE y UE, que sólo son superados por los países nórdicos, además, el porcentaje de españoles de 25 a 34 años titulados en bachillerato y ciclos formativos de grado medio sigue creciendo²⁶. Sin embargo, España sigue ofreciendo uno de los peores datos en cuanto a AEP²⁷.

— *Fuerte influencia del contexto europeo y de los organismos internacionales en la formulación de las políticas educativas nacionales*

Si hay un campo político donde los factores de contexto jueguen un papel lleno de negociaciones (polémicas y complejas), compromisos, estrategias, acuerdos (implícitos y explícitos) es el educativo. Si a ello añadimos que el enfoque educativo, en la mayor parte de los países, se entretene con el sistema productivo, la saturación de esta arena política en cuanto a información y actores resulta más que evidente. Esto obliga al evaluador de políticas públicas a una consideración decisiva del contexto global por el desplazamiento decisional que se ha ido produciendo hacia Bruselas. Hoy resultaría imposible comprender hacia dónde va la educación sin tomar como referentes los informes internacionales: Objetivos de Desarrollo del Milenio de Nacio-

²⁶ Las tasas de titulación en bachillerato se encuentran en España (45%) entre las de la OCDE (48%) y las de la UE (43%). En 2007 ha mejorado en 4 puntos el porcentaje de graduados en ciclos formativos de grado medio.

²⁷ Los datos que considera el Informe se refieren a 2006-2007.

nes Unidas, la Agenda Educativa de la UNESCO, los estudios PISA y otros de la OCDE. Todos ellos ejercen una considerable influencia en las políticas educativas y en la opinión pública, como consecuencia del sistema de indicadores que publican cada año. Las comparaciones internacionales sobre niveles de gasto, modelos de financiación, remuneración de los docentes, estudiantes y titulados, se han convertido en habituales en todos los países desarrollados. Ha sido también la OCDE la que ha prestado una atención destacada a los análisis sobre el impacto del capital humano en el crecimiento económico y en las características del mercado de trabajo.

Contexto especial e ineludible para los países miembros de la UE-27 son los acuerdos y objetivos de la Comisión y el Consejo. El programa educativo europeo se plasma en la Estrategia de Lisboa 2000 y en los Objetivos para 2010, encomendando a la Comisión el liderazgo de los principales cambios en esta materia, por lo que no podría entenderse ningún trabajo sobre educación que no tuviera en cuenta la explotación documental de las Directrices y Comunicaciones UE.

Tampoco puede olvidarse la lucha de intereses en pugna por un mercado, el de la educación, que es una fuente apetecible de poder y de beneficios, donde confluyen muchos intereses, se ponen en juego recursos muy preciados y se generan estrategias para su control por parte de algunos actores. En este sentido el papel de estos grupos en la formulación de las políticas europeas es una cuestión a tener muy en cuenta cuando se examinan medidas para la resolución de los problemas.

La lección más importante que ha dejado el encargo respecto al enfoque y contexto es la necesidad formativa de los evaluadores en el análisis político y en la explotación de los documentos institucionales, requisito indispensable para entender los encargos de las políticas educativas y el grado de responsabilidad de los diferentes niveles de gobierno implicados. No se puede olvidar el papel de las CC.AA. en la implementación de estas políticas.

— *Complejidad de la evaluación*

Hablar de complejidad al evaluar la acción pública resulta una obviedad. En el caso que se expone, cada una de las lecciones extraídas evidencian por sí mismas la complejidad de la evaluación: objeto compuesto, carácter estratégico, política social y conexiones de los sistemas productivos y educativos, multiplicidad de actores implicados en la formulación, intervención de diversas esferas de gobierno, reparto competencial, criterios de valor, entre otros. El tema excede el límite de estas páginas, destacando dentro de la complejidad general del estudio, la multiplicidad de objetos puestos en juego y la interrelación en cadena de sus resultados e impactos. Asimismo, el entramado de criterios que requiere esta política pública para emitir el juicio de valor.

3.2. Lecciones aprendidas en relación con la fase de planificación

— Necesidad de definir con mayor rigor los conceptos

Por lo que respecta a los aspectos técnicos de la evaluación, la primera consideración ha sido la necesidad de definir bien los conceptos.

En el caso de las becas aplicadas al AEP lo correcto sería hablar de “ayudas al estudio personalizadas”, por las siguientes razones: las becas responden a dos requisitos básicos: las circunstancias socioeconómicas y el aprovechamiento académico del solicitante. Si hablamos de utilizar estos beneficios económicos para alumnos que, generalmente, no podrían demostrar ese aprovechamiento, sino todo lo contrario, resulta inconveniente utilizar el término becas. Más correcto sería el de “Ayudas al estudio”, que recoge nuestra normativa para “ayudas que atiende únicamente a las circunstancias económicas del beneficiario”.

En cuanto al discurso del abandono escolar prematuro se está moviendo en dos direcciones superpuestas que tienen que ver con dos tipos de población distinta: Una de estas direcciones es la que marca la definición de Eurostat: “Se considera abandono escolar prematuro el porcentaje de personas de 18 a 24 años que han completado como máximo la primera etapa de la educación secundaria y no siguen ningún estudio o formación adicional” (PNR, 2005:23).

Esta definición alude, en el fondo, a un problema de tipo productivo pues se refiere a una población que está ya fuera del sistema educativo, de baja cualificación profesional al no haber alcanzado las titulaciones mínimas establecidas.

En paralelo, los objetivos educativos y puntos de referencia para el año 2010 del Ministerio de Educación avanzan un enfoque que apunta a una política de mayor titulación y permanencia del alumnado dentro del sistema educativo pero con el foco, no sólo en la adquisición de estudios obligatorios sino también en los postobligatorios. “Reducir el abandono escolar prematuro consiste en aumentar el número de jóvenes que, una vez finalizada con éxito la educación secundaria obligatoria, continúan y finalizan sus estudios de bachillerato o en ciclos formativos de grado medio” (MEPSYD, 2008a: 61).

Las estrategias aquí se dirigen a mantener al mayor número de alumnos dentro del sistema y a la obtención de titulados, no sólo en los niveles de obligatoria, sino de postobligatoria. “España necesita, de aquí a 2012, que 570.000 jóvenes más estudien más allá de la educación obligatoria para que tengan un futuro más próspero y para acercarnos a los países europeos” (MEPSYD, 2008b). A fin de propiciar la continuidad de los jóvenes en los estudios post-obligatorios de bachillerato, formación profesional y educación superior, se plantean las becas como un instrumento estimulante. El énfasis, en todo caso, debiera ponerse en que estas ayudas deben contribuir –y evidenciar– la disminución del abandono escolar prematuro.

— Enfoque evaluativo

Toda intervención pública responde a un enfoque, como tal la evaluación no está exenta de aplicar una perspectiva u otra a la hora de elaborar sus diseños. Es esta una cuestión que no siempre se evidencia y que, en muchos casos, se pretende hasta ocultar, pero lo cierto es que la calidad de la investigación no está sólo en las técnicas y en la recogida sistemática de información, sino en la claridad con la que se exponga el enfoque elegido. Porque sólo así podrá ser contrastado.

En la evaluación que se comenta, el discurso sobre el AEP parece ser el mismo pero encubre matices y diferencias según las instituciones, los informes y las políticas implicadas y las disciplinas que lo abordan.

Concluir sobre las intervenciones públicas referidas al AEP requiere análisis tanto políticos como sociológicos y económicos. Una evaluación comprensiva del asunto habrá de tener en consideración una visión multidisciplinar sin perder de vista que, desde la Administración Pública, el foco ha de ser “la atención que le han prestado las políticas públicas”. Por ello, ha sido indispensable construir la historia de estas intervenciones a través de los textos institucionales. Lo primero que se ha tratado de clarificar es quién dice qué sobre la cuestión y qué se ha hecho. Se ha tratado de reconstruir un discurso en el que se interrelacionan las diferentes voces del espacio público (europeo y nacional), tratando de clarificar cómo se superponen y, a veces, incluso, cómo se contradicen.

Respecto a los estudios comparativos, requieren también un planteamiento investigador que aprecie esa información, pero que considere las peculiaridades de cada situación. En este sentido se recogen las palabras de Mercedes Cabrera, anterior Ministra de Educación: “Las experiencias extranjeras no bastan”²⁸. Y las recomendaciones de *Education at a glance/Panorama de la Educación* (OCDE, 2008): “Los países deberán considerar las opciones –de otros– cuidadosamente y necesitarán mejorar los conocimientos de base en cuanto a cómo se vinculan esas opciones con la relación coste-beneficio si la eficiencia de los servicios educativos ha de aumentar”. Y aún insiste más respecto a los países que están experimentando aumentos en el gasto por alumno –caso de España–: “Necesitan examinar cuidadosamente cómo se hace uso de los recursos”.

En definitiva, el enfoque comparativo pondrá especial cuidado en no copiar por copiar, sino intentar dejar demostrada la eficacia al transferir experiencias de otros países y sobre todo, los niveles de equidad que subyacen en cada sistema político y educativo.

Otra cuestión a considerar a la hora de enfocar la evaluación es la siguiente: los enfoques macro-económicos y macro-sociológicos respecto al AEP son incontestables.

28 La evaluación se llevó a cabo bajo el mandato de la Ministra de Educación, Mercedes Cabrera.

Pero estas perspectivas, hasta el momento, sólo informan sobre lo que está pasando y el número de alumnos a los que afecta. Los datos cuantitativos resultan interesantes al cuantificar y sacar a la luz la gravedad del problema, pero éste persiste y se agrava desde hace años. La visión cuantitativa resulta insuficiente para explicar los porqués. Si se quiere que las becas vayan destinadas a resolver el núcleo del problema se debe abrir la “caja negra” del abandono y poder llegar a determinar lo que esconde. Abrir la caja del abandono es, simplemente, tener en cuenta a los afectados.

Se destaca la falta de estudios cualitativos sobre AEP y, en particular, etnográficos. Estudios de vida que vayan clarificando la acumulación de circunstancias que se pueden llegar a dar en el abandono para comprender la diferencia entre países y entre comunidades dentro de España. En este sentido, no se perderá de vista la orientación que propone la Comisión Europea (2006) a los países miembros para la resolución del AEP: “Hacer hincapié en la prevención, detección precoz y seguimiento individual de las personas en situación de riesgo”. Esto supone introducir un nuevo criterio requerido cada vez más en el análisis y evaluación de las políticas sociales: el criterio diferenciador. Y es en esa misma línea que deberían ir las ayudas financieras que se propongan para su resolución. Se requieren seguimientos individualizados.

— *Criterios de evaluación*

El asunto de los criterios de evaluación está estrechamente ligado con el enfoque de evaluación y ha sido uno de los más ricos en el caso que aquí se presenta. Puso en evidencia, de nuevo, las diferencias que existen entre establecer referentes para evaluar programas y los que requiere la evaluación de las políticas públicas. Los tradicionales criterios que se manejan en el campo de la evaluación de programas resultan insuficientes a la hora de valorar las políticas públicas. En realidad, cada uno de ellas genera sus propios criterios de valor, su jerarquía y sus tensiones. En el caso de la evaluación que nos ocupa se trató de poner en sintonía los dos niveles de acción de gobierno que se entrecruzan y se examinó un entramado de criterios rico y complejo: equidad, pertinencia, coherencia, cobertura, eficacia (gestora y social), cooperación administrativa, eficiencia y productividad, oportunidad.

Sólo a modo de ejemplo se avanzan algunos de los pilares básicos del “construc-to” de criterios.

— *Equidad*

El pilar fundamental de todos estos criterios en la evaluación que nos ocupa lo constituye el criterio de equidad cuyo cumplimiento debe examinarse tanto al hacer el análisis sobre el diagnóstico del sistema actual de becas como al plantearse las

nuevas modalidades que pudieran proponerse para reducir el AEP. Ninguna modalidad, por eficaz que resulte en otros sistemas, sería recomendable si no garantiza el principio de equidad.

En las últimas décadas, las interpretaciones sobre la desigualdad en la educación han pasado de un enfoque unidimensional más determinista a una visión multidimensional e interactiva (Marchesi, 2003).

— *Pertinencia*

Supeditado a este criterio estaría el de pertinencia. Por lo que se refiere al Sistema General de Becas, son numerosos los estudios que evidencian que esa política contribuye a compensar las desigualdades derivadas de factores sociales, económicos, culturales, geográficos étnicos o de otra índole. Más dificultad encierra probar este criterio respecto a las nuevas modalidades que prevé la planificación estratégica para reducir el AEP: ¿Son las becas un buen instrumento para reducir el AEP? El cuestionamiento tiene su base en una de las conclusiones del Estudio de la Comisión Europea que mantiene que en la mayor parte de los países las becas no han demostrado ser un instrumento significativo para frenar el problema, excepción hecha de algunas modalidades de becas que en el Reino Unido han tenido éxito. Naturalmente, la eficacia en ese país, debería contrastarse con otros factores esenciales del sistema educativo británico, tanto económicos como de primacía de la igualdad. Sólo así se podría valorar la eficacia de esas becas en un modelo educativo tan distinto como el español.

Lo anterior enlaza con las metodologías. Los estudios comparativos sobre sistemas políticos, que se apoyan en estadísticas y datos coyunturales sin atender a los estructurales, resultan deficientes y hasta engañosos cuando se trata de analizar políticas sustanciales. Y mucho más cuando se quieren aplicar experiencias de otros países a la solución de sistemas cuya estructura y coyuntura pueden ser bien diferentes.

Con la mejor voluntad, se apuntan a veces soluciones educativas que presentan éxitos en algunos países, pero sin explicar hasta qué punto esos países han ido desmantelando su sistema público de educación.

— *Eficacia*

La lección más contundente en este sentido es que en las políticas públicas hay que distinguir con mayor claridad la eficacia técnica y la eficacia social. Hay que desglosar y transparentar la eficacia en sus distintos componentes. Una intervención puede ser eficaz desde el punto de vista técnico y absolutamente ineficaz desde el punto de vista de la igualdad o la cohesión social.

— *Calidad*

El asunto de la calidad educativa en las intervenciones para reducir el AEP y en particular para las becas, examinó el problema desde la perspectiva de valorar aquellas intervenciones que no persigan sólo la titulación de los que abandonan. Este enfoque, interesante, tal vez, desde el punto de vista productivo al brindar la titulación mayores oportunidades, sin embargo, no puede ser el eje del sistema educativo. Las becas deben ayudar a proveer a los alumnos de abandono la capacidad de formarse adecuadamente como personas y como ciudadanos de este siglo, reforzando la calidad de todos, lo que sin duda revertirá a su vez en una mayor cualificación profesional y en una mejora del capital humano del país.

En este sentido sería de interés evaluar otras intervenciones complementarias dirigidas a disminuir el abandono como los Programas de cualificación profesional.

— *Productividad*

Este criterio resultará capital para establecer el juicio de valor sobre la incidencia del sistema educativo en el capital humano y, en concreto, para establecer los objetivos de productividad de la Estrategia de Lisboa. Es un criterio que tendrá más peso en la segunda fase de la evaluación.

— *Oportunidad*

De aplicación también en la segunda fase, este es un criterio poco utilizado en la evaluación de programas pero que resulta indispensable tener en consideración en la evaluación de las políticas públicas, en particular, cuando se requiere una evaluación *ex ante* que examine futuras medidas a poner en marcha. En ese caso, el evaluador no podrá dejar de examinar la oportunidad “política”, “económica” “social” de las propuestas a la hora de diseñar las medidas, si bien la decisión final ponderará al planificador.

— *Cooperación y colaboración entre administraciones.*

Los sistemas políticos descentralizados llevan consigo mayor responsabilidad en el control por parte del Estado de las condiciones necesarias que mantengan la unidad y la equidad en todo el territorio. Requiere asimismo por parte de las administraciones autonómicas responsabilidad en la colaboración y cooperación.

3.3. Lecciones aprendidas: fase de implementación

— *El papel de la información en la evaluación de políticas públicas*

Una diferencia fundamental entre la evaluación de políticas públicas y la evaluación de programas es la que se refiere a la información. Para los gestores públicos la información y los sistemas de seguimiento resultan fundamentales para saber sobre los resultados. Sin embargo, en el nivel político, la información no es un déficit. En realidad es el nivel más informado. En el caso sobre el AEP, el gobierno conoce desde hace años las tendencias de nuestro país. Dispone antes que nadie de los informes internacionales sobre educación y cuenta con el apoyo de sus unidades internas. Se apoya en el Instituto de Evaluación y cuenta con las oficinas estadísticas. Dispone, además, de un gran aporte de informes de expertos. Es de suponer que información no le falta. Sin embargo, surgen nuevos problemas y se necesita diseñar con el mayor acierto las intervenciones públicas. Es una cuestión de comprensión más profunda y holística en un mundo con cambios vertiginosos. Lo que ha de tener en cuenta el evaluador de políticas públicas es que la necesidad de información es mucho más sofisticada en el nivel estratégico que en el operativo y sus preguntas y sus conclusiones deberán estar en sintonía con esa necesidad.

— *Armonización de datos estadísticos y necesidad de estudios comparativos más allá de lo cuantitativo*

Una de las lecciones aprendidas en esta fase es la falta de armonización de datos estadísticos globales –estudios *International Indicators of Education System* (INES), PISA, Indicadores Eurostat– así como un marco coherente de indicadores que ha establecido la UE para medir el logro de los objetivos de la Estrategia de Lisboa que se encuentra en mejora continua (Comisión Europea, 2006). A ello se añade las disponibilidades informativas del sistema estadístico nacional que siguen ofreciendo algunos problemas para comparar los datos, tanto entre países como entre Comunidades Autónomas.

En cuanto a la medición, no todos los organismos ponen igual énfasis en medir las desigualdades, ni consideran en sus baterías de indicadores exactamente las mismas dimensiones. No podemos olvidar que los indicadores internacionales pueden ser un estímulo para la inversión en gasto educativo o para competir con niveles de calidad, pero también pueden obedecer a la selección de ciertos grupos de presión que resaltan unas dimensiones de la educación desde una posición más mercantilista. Cualquier estudio de evaluación debe considerar estas cuestiones a la hora de redactar las conclusiones.

Por eso las comparaciones entre países y las medidas de intervención propuestas debieran acompañarse de las explicaciones sobre la idiosincrasia de los distintos sistemas educativos (Prats y Raventós, 2005). Hay algunos que presentan eficacia en

cuanto a resultados, pero que generan desigualdad social o que avanzan hacia el desmantelamiento de sus sistemas públicos. Cualquier medida eficaz que se proponga en educación tiene que contrastarse con los valores que sustentan el sistema.

3.4. Resultados

— *Utilidad y uso*

La evaluación que examinamos, como se viene señalando, responde a una utilidad final estratégica. El juicio de valor sólo podrá expresarse tras realizar la evaluación *ex ante* de la segunda fase. Las conclusiones sobre el diagnóstico y las alternativas que se proponen en la primera entran más bien en la categoría de análisis prospectivo que servirá de base para la formulación estratégica.

El trabajo del primer año diagnostica y plantea algunas de las posibles soluciones que se dan en otros países –eso sí, con diferentes sistemas– y que sólo tras el siguiente estudio podrían proponerse para aplicarse con mayor seguridad en el nuestro.

Entre todas las tipologías de evaluación, la evaluación *ex ante* es una de las que tiene más posibilidades de evidenciar el uso e incorporación de sus conclusiones a la formulación política. Aunque nunca quedará claro si la formulación incorpora los resultados de la evaluación o del conjunto de informes que avalan la decisión pública²⁹ que compiten entre sí.

— *Competitividad de la evaluación con otros informes*

En relación con lo anterior hay que destacar que el uso de la evaluación dependerá de muchas cuestiones, pero una de ellas, es que los informes compiten con los de otras instancias. En este caso, el informe de la AEVAL (2009) corría en paralelo con otro sobre AEP que tenía que presentar el Instituto de Evaluación y es seguro que en el despacho ministerial habría otros más para contrastar orientaciones diversas. De especial relevancia son los informes de experto y los que emiten las comunidades epistémicas de prestigio: universidades, fundaciones, consejos, etc.³⁰ La evaluación puede encontrar en esos informes refuerzo o rechazo a sus conclusiones. En la fase final, los informes de evaluación vuelven a la arena política y a la influencia de dimen-

29 Durante las entrevistas realizadas a los responsables del Instituto de Evaluación se tuvo conocimiento de un encargo monográfico al Instituto sobre Abandono Escolar Prematuro.

30 En la Sesión ordinaria del Pleno de 3 de marzo de 2009, el Consejo Económico y Social presentó su Informe 01/2009 sobre *Sistema Educativo y Capital Humano* que analiza infinidad de cuestiones similares a las planteadas por la evaluación AEVAL. La diferencia mayor entre ambos tipos de trabajo es que las conclusiones del Consejo son consensuadas por el conjunto de actores mientras que en el caso de la evaluación, podrían consensuarse las preguntas de evaluación, pero en ningún caso las conclusiones a las que llegue el equipo evaluador. Teniendo en cuenta esta consideración, es innegable que los dos tipos de trabajo tienen su provecho. Y que pueden servir mutuamente para contrastar los hallazgos.

siones que a ésta le son propias: valores e ideología, aspiraciones, compromisos, experiencia y, en particular, la concepción del tiempo y la oportunidad para las transformaciones que se persiguen. Vuelve entonces la evaluación a la pugna entre diferentes intereses. Sus conclusiones entran en competición con orientaciones diversas: intelectuales, políticas y sociales que los evaluadores deberán tener en cuenta a la hora de considerar las decisiones públicas.

4. CONCLUSIÓN

A fin de no reiterar las conclusiones precisas que se extraen del epígrafe anterior, se exponen a continuación las generales que se derivan de la evaluación en su nivel estratégico. Este tipo de evaluación tiene un campo propio, de naciente estudio y aún de escaso desarrollo que es todo un desafío y una oportunidad para instituciones como la AEVAL. El reto está en encadenar mejor los niveles operativos y los estratégicos de la acción pública y para ello habrá que explorar nuevas herramientas de evaluación.

En este campo se le reclama cada vez más ocuparse de las políticas públicas en el nivel macro, por lo que se hace preciso distinguir las particularidades que hay entre la evaluación tradicional de programas y proyectos que han sido hasta el momento las más numerosas y las nuevas exigencias de los decisores públicos. Para el objeto macro resulta determinante la aplicación de métodos y técnicas de múltiples disciplinas, pero en particular, los de la ciencia política que, en cierta medida, hasta el momento, ha tenido menor protagonismo en la evaluación de programas. Los gobiernos no sólo demandan datos, sistemas de seguimiento, y resultados para la mejora, sino análisis de proyección y de escenarios en los que esos datos puedan traducirse en intervenciones públicas más eficaces para resolver los problemas de la sociedad. Reclaman cada vez más la evaluación como apoyo a la definición de nuevas estrategias.

En la evaluación de políticas públicas, la eficacia debe tener en cuenta, no sólo la eficacia técnica sino la eficacia social y los evaluadores deberán esforzarse por desglosar y analizar ambos aspectos antes de elaborar sus juicios.

En el mundo globalizado, los análisis comparativos son absolutamente necesarios, pero sólo resultarán útiles en la medida en que vengan contextualizados.

Cualquiera de las medidas propuestas para resolver el AEP a través de becas debería probar que los recursos destinados a esa población están encaminados a:

1. Proporcionar igualdad de oportunidades respecto a su promoción social y a una atención personalizada que tenga en cuenta la casuística de esa población.
2. Formar ciudadanos que puedan participar en la toma de decisiones en una sociedad democrática y multicultural.

3. Contribuir a emancipar a los jóvenes a fin de hacer de ellos personas adultas y responsables.
4. Asegurar que el refuerzo generará valor añadido en la búsqueda de empleo en la sociedad del conocimiento.

Para unas conclusiones más pormenorizadas de los resultados de la evaluación remitimos al lector al Informe completo de la evaluación que figura en la web de la AEVAL (2009). La finalidad de este trabajo era sumar valor al primero examinando las lecciones aprendidas que desde el *expertizaje* de evaluación ha dejado el proceso.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- AEVAL (2009), *Informe Evaluación del sistema general de becas educativas. Primera Evaluación: Diagnóstico de la situación actual y principales alternativas para mejorar su eficacia*. Ministerio de la Presidencia. Madrid: AEVAL. En línea: <<http://www.aeval.es/comun/pdf/evaluaciones/E11.pdf>> (Consulta 15 noviembre 2009)
- AEVAL (2009), *Informe: La percepción de los españoles hacia los servicios públicos (1985–2008)*. Observatorio de la Calidad de los Servicios Públicos. Madrid: AEVAL.
- Bachrach, P. y Baratz, S.M. (1963), “Decisions and Non-decisions: An *analytical* framework”, *American Political Science Review*, 3 (57): 632-642.
- Calero, J. (1993), *Efectos del gasto público educativo. El sistema de becas universitarias*. Barcelona: Servicio de Publicaciones de la Universitat de Barcelona.
- (2006a), *La equidad en educación. Informe analítico del sistema educativo español*. Madrid: CIDE, Ministerio de Educación y Ciencia.
- (2006b), *Desigualdades tras la educación obligatoria: nuevas evidencias*. Documento de trabajo de la Fundación Alternativas, 83/2006.
- Calero, J. y Bonal, X. (1999), *Política Educativa y Gasto Público en Educación. Aspectos teóricos y una aplicación al caso español*. Barcelona: Pomares-Corredor.
- Coleman, J.S. et al. (1966), *Equality of Educational Opportunity Study (EEOS)*, Washington, DC: Department of Health, Education, and Welfare, Office of Education.
- Comisión Europea (2001), *Futuros objetivos precisos de los sistemas de educación y formación*. Bruselas: COM(2001).0059 Final. Doc. 5980/01, Comisión Europea.
- (2006), “Resumen Ejecutivo sobre el estudio del abandono escolar prematuro”. *Revista de Educación, La cualificación profesional básica: competencias para la inclusión sociolaboral de jóvenes*, 341. Madrid: Ministerio de Educación y Ciencia: 899-916.
- Consejo de la Unión Europea (2001), “Sobre los futuros objetivos precisos de los sistemas de educación y formación”, Doc. 5680/01 EDUC 18. Informe del Consejo “Educación” al Consejo Europeo. Bruselas: Consejo de la Unión Europea.
- Díaz Malledo, J. y San Segundo, J.M. (2000), “La financiación de la enseñanza superior: Un análisis comparado de la situación española”, *Papeles de Economía Española*, 86: 249–265.
- Dror, Y. (2003), *Public Policymaking reexamined*. New Brunswick: Transaction Publishers.

- Dye, T.R. (1975), *Understanding Public Policy*. Englewood Cliffs: Prentice Hall.
- García Sánchez, E. (2004), “La institucionalización de la evaluación educativa no universitaria en España”, *Gestión y Análisis de Políticas Públicas*, 30/31
- Hecló, H. (1972), “Policy Analysis”, *British Journal of Political Science*, 12: 83-108.
- Iglesias de Ussel, J. et al. (2009), *Sistemas y políticas de educación superior. Premio de Investigación del Consejo Económico y Social 2005*. Madrid: Consejo Económico y Social.
- Consejo de Universidades (1998), “Informe sobre evaluación de la calidad de las universidades”, *Gestión y Análisis de Políticas Públicas*, 11/12: 157-181.
- INCE (1999), *Instituciones para la evaluación del sistema educativo*. Madrid: Ministerio de Educación y Cultura. En línea: <<http://www.ince.mec.es/jornad/index.htm>> (Consulta: 15 noviembre 2009).
- INE (2006), *Encuesta de Condiciones de Vida (ECV)*. Madrid: Instituto Nacional de Estadística. En línea: <<http://www.ine.es/jaxi/menu.do?type=pcaxis&file=pcaxis&path=%2Ft25%2Fp453%2F%2Fa2006>> (Consulta: 15 noviembre).
- (2005), *Encuesta de Transición Educativo-Formativa e Inserción Laboral (ETEFIL)*. Madrid: Instituto Nacional de Estadística. En línea: <http://www.ine.es/daco/daco42/etefil/etefilo5_tabav.htm> (Consulta: 15 de noviembre).
- Julkunen, I. (2007), “El abandono escolar prematuro en Finlandia ¿un problema resuelto?”, *Revista de Estudios de Juventud*, 7: 123-138.
- Marchesi, A. (2003), *El fracaso escolar en España*. Documento de Trabajo. Fundación Alternativas, Madrid.
- Marchesi, A. y Hernández Gil, C. (2003), *El fracaso escolar una perspectiva internacional*. Madrid: Alianza Editorial.
- MEPSYD (2007a), *Apuntes del Instituto de Evaluación*, 10. Resumen Informativo. Madrid: Ministerio de Educación, Política Social y Deporte.
- (2007b), *Apuntes del Instituto de Evaluación*, 11. Resumen Informativo. Madrid: Ministerio de Educación, Política Social y Deporte.
- (2008a), *Plan para la reducción del abandono escolar, El incremento de la formación postobligatoria en España*. Documento de trabajo. Madrid: Ministerio de Educación, Política Social y Deporte.
- (2008b), *Plan para reducir el abandono escolar a la mitad en 2012*. Madrid: Ministerio de Educación, Política Social y Deporte.
- (2008c), *Panorama de la Educación. Indicadores de la OCDE*. Madrid: Instituto de Evaluación, Ministerio de Educación, Política Social y Deporte.
- (2008d), *Datos y cifras. Curso escolar 2008/2009*. Madrid: Secretaría General Técnica, Ministerio de Educación, Política Social y Deporte.
- (2009), *Panorama de la Educación. Indicadores de la OCDE*. Madrid: Instituto de Evaluación, Ministerio de Educación, Política Social y Deporte.
- Ministerio de la Presidencia (2005), *Convergencia y Empleo. Programa Nacional de Reformas de España*. Madrid: Secretaría General Técnica, Ministerio de la Presidencia.
- Mora, J. G. (1996), “Equidad en el Acceso a la Educación Superior, ¿Para Quién Son las Becas?”, *Revista de Educación*, 309: 239-259.
- Observatorio Social de España (2007), *Informe. El estado de bienestar en España y las CC.AA. Análisis de Indicadores Clave*. Barcelona: Secretaría de Estado de la Seguridad Social.

- OCDE (2006), *PISA. Science Competencies for Tomorrow's World*. París: Organisation for Economic Co-Operation and Development. En línea: <http://www.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html> (Consulta 15 noviembre 2009).
- (2008), *Education at a Glance OECD Indicators /Panorama de la Educación. Indicadores de la OCDE*. París: Organisation for Economic Co-Operation and Development. En línea: <<http://www.oecd.org/dataoecd/16/59/41262207.pdf>> (Consulta: 15 noviembre 2009).
- (2009), *Education at a Glance. OECD Indicators*. París: Organisation for Economic Co-Operation and Development. En línea: <<http://www.oecd.org/dataoecd/32/34/43541373.pdf>> (Consulta 15 noviembre 2009).
- OCDE-CERI (1995), *Our children at risk*. París: Organisation for Economic Co-Operation and Development.
- Pastor, M., et al. (2006), *El rendimiento del capital humano en España*. Valencia: IVIE, BANCAJA. En línea: <<http://obrasocial.bancaja.es/publicaciones/publicaciones-ficha.aspx?id=124>>, (Consulta: 15 noviembre 2009).
- Prats, J. y Raventós, F. (2005), *Los sistemas educativos europeos. ¿Crisis o transformación?* Barcelona: Obra Social Fundación “La Caixa”. En línea: <http://obrasocial.lacaixa.es/estudiossociales/vol18_es.html> (Consulta: 15 noviembre 2009).
- Puente Azcutia, J. (1999), “Instituciones evaluadoras y administración educativa”, *Jornadas: Instituciones para la evaluación del sistema educativo*. Madrid: Instituto Nacional de Calidad y Evaluación.
- San Segundo, M. J. (2001), “El impacto nacional y regional del programa de becas” *X Jornadas de la Asociación de la Economía de la Educación*. Murcia: Asociación de Economía de la Educación.
- (2003), “El gasto público en educación: 1975-2000”, en J. Salinas y S. Álvarez, coords., *El gasto público en la democracia*. Madrid: Instituto de Estudios Fiscales.
- (2005), “Política de becas y ayudas al estudio en los presupuestos del año 2005”, *Presupuesto y gasto público*, 38: 7-30.
- (2006), “La Política educativa a comienzos del siglo XXI”. *Información Comercial Española (ICE). Productividad y competitividad de la Economía Española*, 829: 49-66.
- Stewart, J. y Ranson, S. (1988), “Management in the Public Domain”, *Public Money and Management*, 8 (1/2). CIPFA: 13-19.
- Subirats, J. (2005), “Catorce puntos esenciales sobre evaluación de políticas públicas con especial referencia al caso de las políticas sociales”. *Economías*, 60, (1): 19-37.
- Tiana, A. (2007), “La política de educación en los Presupuestos Generales del Estado para 2007: Objetivos y prioridades”. *Presupuesto y Gasto Público*, 46: 167-174.
- (2008), “La política de becas del Estado: un balance del período 2004-2008”, *Presupuesto y Gasto público*, 50: 191-200.

Autoevaluación y políticas públicas: una experiencia en escuelas primarias argentinas

Elena Duro
Olga Nirenberg

Resumen: *¿Es posible influir positivamente en la política pública mediante procedimientos de autoevaluación sobre la gestión de los establecimientos que brindan servicios públicos? Bajo la premisa de que la autoevaluación promueve “organizaciones inteligentes” (que aprenden y se transforman a partir de su propia práctica) tal camino resulta estratégico y posible, aunque no excluyente. En el presente artículo se relata la implementación de un Instrumento para la Autoevaluación de la Calidad Educativa (IACE) en escuelas primarias argentinas¹. Se consideró la autoevaluación como un complemento de las formas tradicionales de evaluación de la calidad educativa que se implementan desde los años 90 en los países de América Latina y Caribe. Esas evaluaciones de los sistemas educativos resultan necesarias pero no suficientes para elevar los niveles de aprendizaje de los estudiantes y las capacidades retentivas de las escuelas; su dirección es desde arriba hacia abajo tanto en su implementación como en las decisiones y en la formulación de políticas, mientras que la autoevaluación desarrollada va desde abajo hacia arriba: nace en las escuelas, detecta los problemas y promueve acciones de superación desde la óptica de los docentes, los directivos y los familiares de alumnos.*

Palabras clave: *Evaluación participativa, planes de mejora, gestión institucional, calidad educativa.*

1. SIGNIFICADO E IMPLICANCIAS DE LA EVALUACIÓN

En el Instrumento para la Autoevaluación de la Calidad Educativa (IACE), se define la autoevaluación como una actividad programada y sistemática de reflexión que hacen los protagonistas acerca de la propia acción desarrollada en cada estableci-

¹ Esa implementación fue entre el 2007 al 2009, en el marco institucional de UNICEF Argentina en convenio con el Centro de Apoyo al Desarrollo Local (CEADEL), Organización no Gubernamental de Argentina que trabaja en el campo de las políticas públicas, en especial en intervenciones evaluativas.

miento educativo, sobre la base de información confiable, con la finalidad de emitir juicios valorativos fundamentados, consensuados y comunicables (Nirenberg *et al.*, 2000).

El propósito de toda evaluación debiera ser siempre recomendar acciones orientadas a la mejora. En el caso del IACE, dado que la autoevaluación la realizan los propios integrantes de la escuela, principalmente los directivos, docentes y familiares a cargo de alumnos, a los que pueden sumarse otros miembros de la comunidad educativa, como alumnos, personal no docente y demás agentes de la misma, son esos mismos actores quienes recomiendan –mediante dinámicas grupales que permiten discusiones, acuerdos y consensos– actividades para la superación de los problemas detectados y priorizados. Así, las acciones orientadas a la mejora no son impuestas por las autoridades ni por nadie externo a las escuelas, lo que sin duda aporta más viabilidad a su ejecución ulterior.

La evaluación que sólo se queda en el enjuiciamiento, en particular si éste no resulta del todo positivo, suele convertirse en frustración y tornarse inmovilizador. Esa es una de las razones, aunque no la única, de las conocidas resistencias a la evaluación y de la poco frecuente utilización o aplicación de las recomendaciones evaluativas (Nirenberg, 2003).

Pero si el propósito explícito desde el comienzo es aprender en forma conjunta a partir de la propia práctica (tanto de los errores o las falencias como de los aciertos o las fortalezas) para extraer de allí conocimiento compartido y útil para mejorar la acción, entonces la significación de la evaluación es evidente para el fortalecimiento de todos aquellos que se involucran en sus procesos (Nirenberg, 2009).

En función de la concepción que las autoras sustentan acerca de la evaluación, como una instancia de enseñanza y aprendizaje para todos los que participan de ella, el método de la pedagogía socrática² resulta el más indicado para su aplicación. Este método consiste en plantear interrogantes correctos y procurar respuestas adecuadas generando espacios de intersubjetividad (Habermas, 1982)³ donde se puedan dar discusiones, opiniones, y se reflejen razonamientos, fundamentos y evidencias. Así, es posible aprender algo más sobre la estructura y el funcionamiento de la escuela con vistas a mejorar su calidad y generar un entorno protector para que todos los niños y las niñas tengan oportunidades de educarse.

2 Sócrates, en el diálogo con Fedro, su alumno, le explicaba que sólo a través de preguntas apropiadas es posible aprender. Así, la evaluación puede ser vista como un proceso de aprendizaje de estilo “socrático”, pues se centra en formular preguntas evaluativas y responderlas con argumentos, en forma fundamentada [Platón, 1970].

3 Este autor menciona procedimientos de creación de campos epistemológicos de intersubjetividad, contraponiendo este tipo de generación de conocimiento al que se realiza en forma aislada o individual y que corre el riesgo de la subjetividad individual, no sometida a crítica. Sostiene que la ciencia, incluyendo las llamadas “ciencias duras” o aún en el enfoque positivista, carece de objetividad absoluta, lo cual no impide procurar el rigor.

2. FUNDAMENTOS PARA PROMOVER LA AUTOEVALUACIÓN DE LA GESTIÓN ESCOLAR

Los sistemas educativos de la región de América Latina y el Caribe presentan persistentes déficit de calidad y equidad que impactan más negativamente sobre las poblaciones más pobres. Las razones son múltiples y abarcan causas extraescolares e internas a la escuela. Cada vez más las respuestas para alcanzar el derecho a una educación de calidad para todos, provendrán no sólo de cambios en las dinámicas escolares, sino de apoyos del conjunto de sectores de la política pública y de redes de apoyo local a las familias y a la institución escolar (UNESCO, 2007).

En la década de los 90 se inicia un proceso de evaluación de la calidad de los sistemas educativos, destinado a medir básicamente los logros de aprendizaje que alcanzan los estudiantes en distintos años y niveles de la escolaridad básica y secundaria. Estas políticas de evaluación instauradas en la mayoría de los países de la región, son absolutamente necesarias, sobre todo para la formulación de políticas educativas nacionales, pero no suficientes para provocar por un lado el arraigo de una cultura evaluativa en los actores escolares, y por otro, no alcanzan para producir los cambios necesarios definidos y activados por los principales actores del hecho educativo: los docentes, las familias y la comunidad.

Las razones de la mencionada insuficiencia de esos grandes operativos sistémicos para provocar transformaciones perdurables en las prácticas y las gestiones institucionales educativas son: en primer lugar que se centran fundamentalmente en los logros del aprendizaje y menos en otros aspectos que hacen a la calidad educativa, como la inclusión, la atención a la diversidad, la promoción de los derechos, el clima escolar y otras cuestiones de singular importancia. En segundo lugar, el carácter externo y predominantemente cuantitativo del enfoque evaluativo que utilizan, impiden profundizar en los factores que expliquen los problemas del fracaso escolar y las trayectorias truncadas de los alumnos. Y en tercer lugar, un factor relevante es la demora en las devoluciones de los resultados evaluativos para que permitan tomar mejores y más oportunas decisiones en los niveles locales e institucionales, ya que el tiempo que transcurre entre los operativos de recolección de los datos y la obtención de resultados procesados, analizados y publicados suele ser demasiado extenso.

Además, por lo general ese tipo de evaluaciones externas suelen finalizar en rankings o escalas clasificatorias de escuelas, que más que promover actitudes hacia la mejora en los planteles (así como en los alumnos y sus familiares) de las escuelas que se ubican en lugares bajos de esas escalas, provocan desaliento, impotencia o en el mejor de los casos, actitudes competitivas.

No nos proponemos reemplazar las evaluaciones externas y/o cuantitativas por otras de carácter participativo –autoevaluaciones– con enfoques cualitativos, sino procurar eficaces triangulaciones entre los diferentes abordajes (Nirenberg, 2000,

2003)⁴, y quizás influir también en la selección de indicadores, más integrales y de mayor riqueza que los que tradicionalmente se incluyen en esos operativos nacionales masivos.

Como se aclaró al comienzo, una fuerte hipótesis es que los procesos autoevaluativos que se desarrollan en cada establecimiento escolar permiten que los planteles docentes a la vez que toman conciencia de las problemáticas prevalentes programen e implementen (como protagonistas) procesos de mejora en sus ámbitos institucionales, que resulten viables en ese nivel de la acción. Pero además, con las consolidaciones y agregaciones que los niveles centrales de las jurisdicciones realicen de los problemas que las escuelas plantean y de las propuestas y planes de mejora que formulan, se espera una rápida y más oportuna incidencia en la formulación de líneas de acción y políticas educativas de esos niveles jurisdiccionales⁵, así como, por su intermedio, influir en las políticas educativas del nivel nacional. No se desconoce que para tales efectos son requeridas intensas actividades de abogacía⁶.

En resumen: la evaluación de la calidad de los sistemas educativos emerge como una condición necesaria para la definición de políticas pero no suficiente para activar las transformaciones que se requieren en pos de la mejora educativa no sólo en el plano de los logros sino en el de la inclusión social (Bolívar, 2009).

Actualmente más del 50% de los adolescentes de la región están fuera del ámbito educativo formal. En Argentina casi el 25% de la población escolar no culmina el nivel básico obligatorio y alrededor del 40% de los adolescentes no culminan la secundaria superior. Por otro lado, los promedios nacionales obtenidos en los operativos de la Evaluación de la Calidad en relación a los logros de aprendizaje en lengua, matemática y ciencias son bajos. Historias de fracaso y abandono temprano suelen acompañar las trayectorias escolares de sectores importantes de la población escolar más desaventajada (Niremberg, 2006). Mejorar la educación es una obligación permanente y continua de los Estados en un siglo de cambios, desafíos y cada vez más exigencias para las nuevas generaciones.

En este escenario se trabajó en el desarrollo de una herramienta de autoevaluación institucional, que orientada a la mejora de los desempeños y los logros puede resultar un proceso muy sinérgico (Max-Neef, 1986)⁷ para promover la calidad en las

4 Para mayor profundización en los tipos de evaluación según predominancias cuantitativas y cualitativas y según quienes la realizan, y para mejor entendimiento del concepto de triangulación.

5 Por “niveles jurisdiccionales” se hace referencia a los niveles provinciales (estatales) del gobierno educativo y en algunos casos también a los municipales.

6 El término abogacía traduce la palabra inglesa *advocacy*, que alude a las acciones que se desarrollan en favor de una causa, para colocarla en forma relevante en la agenda pública y en las políticas gubernamentales.

7 El autor considera sinérgicas las acciones que además de orientarse a la resolución de un problema puntual, contribuyen a la vez a la resolución o prevención de otros problemas actuales o futuros.

escuelas primarias, debido al protagonismo que otorga a los propios agentes escolares en la mejora de sus prácticas.

Fue por ello que UNICEF Argentina, desde el año 2007, en convenio con el Centro de Apoyo al Desarrollo Local (CEADEL), en forma articulada con los Gobiernos educativos provinciales y el aval político del Ministerio de Educación Nacional, ha desarrollado el IACE, en coherencia con los objetivos relativos a la evaluación que estipula la Ley 26.206 de Educación Nacional⁸, como una cooperación de UNICEF a los gobiernos educativos para facilitar procesos autoevaluativos en los establecimientos escolares del nivel primario.

3. EL PROCESO DE DISEÑO Y APLICACIÓN DEL IACE

El IACE es el producto de un proceso de generación colectiva de conocimiento, donde confluyeron distintas vertientes: algunas provenientes de la experiencia acumulada por las autoras, a partir de sus formaciones académicas y trayectorias profesionales, otras de la exhaustiva y focalizada búsqueda de bibliografía actualizada que realizaron sobre la calidad educativa y los temas asociados que fueron abordados, y otras provenientes de los aportes de agentes de las diferentes instancias del sistema educativo, es decir, funcionarios de los gobiernos educativos nacionales y jurisdiccionales, supervisores de distritos escolares, directivos de establecimientos y docentes de aula.

Se sintetizan a continuación los principales pasos seguidos para el diseño del IACE, los que ponen en evidencia ese proceso de creación colectiva.

- Se realizó un estudio de opinión en tres provincias en las cuales UNICEF ya venía trabajando en otros temas (educativos y/o de salud): Buenos Aires, Misiones y Tucumán; se indagó acerca del significado que los actores adjudicaban a la calidad educativa y cuáles consideraban sus aspectos más relevantes; también se relevaron las percepciones acerca de la necesidad (y viabilidad) de encarar procesos de autoevaluación en las escuelas, donde se contemplan los puntos de vista de los diferentes actores. Mediante grupos focales, cuestionarios autoadministrados y entrevistas, el estudio indagó opiniones de funcionarios nacionales y provinciales, directivos y docentes, alumnos de los grados superiores, personal no docente y también de familias a cargo de alumnos.

En tal sentido, la mejora de la educación es una acción de las más sinérgicas desde el punto de vista del desarrollo humano y social.

⁸ En el capítulo III, artículo 94 de dicha ley, se dice: “El Ministerio de Educación tendrá la responsabilidad principal en el desarrollo e implementación de una política de información y evaluación continua y periódica del sistema educativo para la toma de decisiones tendiente al mejoramiento de la calidad de la educación, la justicia social, en la asignación de recursos, la transparencia y la participación social”; y en el mismo capítulo, el artículo 96 precisa: “...asimismo, apoyará y facilitará la autoevaluación de las unidades educativas con la participación de los/as docentes y otros/as integrantes de la comunidad educativa”.

- Se elaboró el marco conceptual acerca de la calidad educativa en escuelas primarias y se especificaron sus principales dimensiones.
- En función de ese marco conceptual, se determinó el abordaje evaluativo a adoptar y se diseñaron los instrumentos y ejercicios del IACE para poder realizar la autoevaluación.
- Se sometieron esos productos preliminares a la técnica denominada “juicio de expertos”, para su corrección y ajuste, convocando para eso a directivos y docentes de veinte escuelas primarias de las provincias de Buenos Aires y Tucumán. A lo largo de intensas y animadas jornadas de trabajo conjunto con el equipo a cargo del proyecto revisaron en forma concienzuda las herramientas diseñadas y realizaron aportes que fueron atendidos y tomados en cuenta para la primera edición del IACE.
- El IACE fue aplicado, en una primera instancia de puesta a prueba, entre mayo y octubre del 2008, en 70 escuelas de las provincias de Buenos Aires, Misiones y Tucumán. De allí emergieron aprendizajes que permitieron introducir ajustes que fueron plasmados en una segunda edición.
- A finales del año 2008 y comienzos del corriente año 2009 se llevaron a cabo nuevas aplicaciones en 252 escuelas de las provincias de Tucumán, Misiones y Chaco, completando así 322 escuelas.
- En acuerdos con las autoridades del Ministerio de Educación Nacional (MEN), y en base a los aprendizajes emergentes de las aplicaciones ya efectuadas, se introdujeron nuevos ajustes y modificaciones⁹ que se plasmaron en una tercera edición (Duro y Nirenberg, 2009), con el propósito de facilitar la generalización de su utilización (intención que fue expresada por las autoridades del MEN).
- Durante la segunda mitad del año 2009, UNICEF – CEADEL apoyó nuevas aplicaciones en otras 255 escuelas de las tres provincias mencionadas (en el momento de escribir este artículo esas aplicaciones están llevándose a cabo), alcanzando así un total acumulado de 577 escuelas.
- También en el segundo semestre del corriente año 2009 se desarrollaron dos “separatas” del IACE, adecuadas para su aplicación en escuelas de contextos rurales y para aquellas que atienden alumnos de pueblos originarios (indígenas)¹⁰. Las mismas fueron solicitadas por las autoridades provinciales.

9 Esas modificaciones fueron de dos tipos: primeramente se redujeron de seis a tres las dimensiones consideradas en la Calidad educativa, lo cual implicó un reordenamiento de las variables consideradas en cada dimensión; en segundo lugar se revisaron y simplificaron los ejercicios básicos y opcionales propuestos. Los aspectos sustanciales del IACE sin embargo se mantuvieron, así como el rigor metodológico.

10 Tales escuelas se denominan EIB (de educación intercultural bilingüe). Pueden verse ambas separatas en www.ceadel.org.ar/unicef-iace.

4. EL ALCANCE DE LA APLICACIÓN DEL IACE

En la tabla que sigue se estiman las cantidades de docentes y alumnos cubiertos, así como las familias de alumnos indagadas (mediante cuestionario autoadministrable) en las aplicaciones realizadas y en realización hasta la fecha.

TABLA 1

Escuelas, docentes, alumnos y familias, según jurisdicción-2009

| Provincias | Escuelas cubiertas | Docentes estimados | Alumnos estimados | Familias indagadas |
|--------------|--------------------|--------------------|-------------------|--------------------|
| Buenos Aires | 20 | 665 | 10.000 | 1.600 |
| Chaco | 150 | 3.000 | 53.000 | 12.000 |
| Misiones | 177 | 3.540 | 62.000 | 14.000 |
| Tucumán | 230 | 5.500 | 80.500 | 18.400 |
| Total | 577 | 12.705 | 205.500 | 46.000 |

Fuente: Elaboración propia.

Merece reiterarse que las provincias donde el IACE se aplicó (o se encuentra en aplicación) son aquellas en las que UNICEF contaba en forma previa con acuerdos marco de cooperación y por ende, donde ya venía trabajando en diversos temas (educativos, de derechos y/o de salud) relacionados con el desarrollo integral de la infancia, circunstancia que otorgó viabilidad a los primeros procesos aplicativos; lo cual en un principio de instalación del dispositivo fue importante de considerar para los efectos demostrativos buscados¹¹. Además, esas tres provincias muestran los niveles de pobreza más altos e indicadores educativos más desfavorables, en términos comparativos con los del resto del país.

Los gobiernos educativos de las provincias de Mendoza y Neuquén (que no cuentan con convenios previos de trabajo con UNICEF) han solicitado la inclusión gradual de sus escuelas (400 de un total de 800 en Mendoza y 120 de un total de 350 en Neuquén); aún no se han celebrado los correspondientes convenios, por lo cual ambas aplicaciones se realizarían durante el primer semestre del año 2010.

Dado que UNICEF es un organismo de cooperación con gobiernos nacionales y provinciales y no una instancia gubernamental, no necesariamente busca coberturas “universales” sino que se preocupa por instalar prácticas con fines “demostrativos” y por esa vía incidir en las políticas de Estado que se orientan a niños/as y adolescentes. De allí también el trabajo conjunto del equipo IACE con el Ministerio de Educación

¹¹ En cada provincia, la selección de las escuelas fue responsabilidad de las autoridades educativas provinciales.

Nacional, para arribar a un consenso en materia de autoevaluación educativa en el nivel primario.

Más allá de esa advertencia previa se brinda a continuación una estimación de la cobertura¹² de escuelas alcanzada hasta la fecha con la aplicación del IACE en esas tres provincias:

TABLA 2

Escuelas cubiertas acumuladas y % de cobertura-2009

| Provincias | Total escuelas primarias públicas | Escuelas cubiertas | % de cobertura |
|------------|-----------------------------------|--------------------|----------------|
| Chaco | 351 | 150 | 43 |
| Misiones | 832 | 177 | 21 |
| Tucumán | 631 | 230 | 36 |

Fuente: DINIECE, Ministerio de Educación Nacional (2007) y elaboración propia.

Vale aclarar que las autoridades educativas de las tres provincias han manifestado sus intenciones de aplicar el IACE en todas las escuelas primarias de sus dependencias, en un proceso gradual, de modo tal que la expectativa es que esas jurisdicciones alcancen la cobertura total de sus escuelas entre el año 2010 y el 2011. Y aunque aún no han comenzado, esa misma intención fue expresada por las autoridades de Mendoza y Neuquén.

Por otra parte, se reitera que las autoridades del Ministerio de Educación Nacional (MEN) expresaron su interés por aplicar el IACE en las escuelas públicas primarias de diferentes provincias argentinas que están bajo programas nacionales de inclusión educativa¹³, las que suman alrededor de 2500 escuelas. Con tal motivo se acordaron con el MEN las modificaciones introducidas en el dispositivo de modo que su generalización e “ida a escala” se vea facilitada.

5. EL SIGNIFICADO DEL CONCEPTO DE CALIDAD EDUCATIVA

La misión de la escuela y sus logros guardan estrecha relación con la idea de calidad que cada establecimiento sustente; sin desmedro de que a nivel jurisdiccional y nacional existan lineamientos sobre el sentido de tales cuestiones, sin embargo en cada escuela, sea en forma explícita o implícita, existe una especificación de la misión

¹² No se estima la cobertura en la provincia de Buenos Aires por ser ésta insignificante, dado que allí la aplicación se realizó como parte de la “puesta a prueba” del instrumento.

¹³ Esos programas cuentan con financiamiento bilateral (del BID y BIRF, con contraparte de fondos del Tesoro Nacional).

y consecuentemente de lo que significa la calidad educativa y qué logros son deseables en los alumnos.

En el IACE se considera que la calidad educativa es un concepto multidimensional, que son diversos los aspectos a los que hace referencia y múltiples los factores que la determinan.

Este concepto puede ser analizado en relación con el sistema educativo como un todo o bien en relación con cada escuela. En el primer caso, si se quisiera resumir en un solo párrafo el complejo concepto de calidad educativa, podría decirse que un sistema (provincial o nacional) con una educación inclusiva y de calidad es aquel que logra que todos los niños y las niñas ingresen en la escuela y tengan trayectorias escolares completas, cumpliendo la edad teórica deseada y obteniendo adecuados logros de aprendizaje.

Precisando lo dicho, un sistema educativo de calidad tiene:

- Objetivos curriculares relevantes y compartidos¹⁴.
- Eficacia, es decir que logra acceso universal, permanencia y egreso de los alumnos acorde con los objetivos de aprendizaje previstos.
- Impacto, en términos de que los alumnos adquieren conocimientos y conductas duraderas o permanentes, orientadas a una ciudadanía plena y cabal.
- Eficiencia, o sea que cuenta con recursos humanos y materiales suficientes y los aprovecha de la mejor manera posible.
- Equidad, es decir que toma en cuenta la desigual situación de los alumnos y las comunidades en que éstos y sus familias viven, y brinda apoyo especial a quienes lo requieren, para que los objetivos educativos puedan ser alcanzados por todos.

Teniendo en cuenta lo expresado para el sistema educativo como un todo y considerando la calidad educativa de cada escuela –que es lo que aborda realmente el IACE– puede decirse que un establecimiento brinda educación de calidad si:

- Logra que todos los alumnos y las alumnas aprendan lo que tienen que aprender.
- Posee objetivos de enseñanza pertinentes y actualizados, plasmados en un proyecto educativo institucional.

¹⁴ Tales objetivos en Argentina se reflejan en los Núcleos de Aprendizaje Prioritarios (NAP), que constituyen el cuerpo de contenidos comunes y obligatorios que deben incorporarse al currículo y desarrollarse en todas las escuelas del país, ya sean éstas de gestión pública o de gestión privada.

- Implementa estrategias para prevenir el fracaso y la deserción escolares.
- Logra el acceso, la permanencia y el egreso de los alumnos, de acuerdo con los objetivos de aprendizajes deseables y programados.
- No discrimina y trabaja la diversidad.
- Tiene en cuenta la desigual situación de los alumnos, las familias y las comunidades en que viven, y promueve apoyo especial a quienes lo requieren (por sí misma o a través de otras organizaciones locales).
- Fomenta y posibilita el desarrollo integral de los niños y las niñas.
- Genera un clima escolar favorable y respetuoso.
- Genera un entorno protector de los derechos de la infancia.
- Cuenta con recursos humanos y materiales suficientes y acordes a las necesidades.
- Promueve la participación de los miembros de la comunidad educativa.

Como puede verse en el IACE, esas son variables consideradas clave, que van más allá de los logros de aprendizaje, aunque sin duda repercuten en ellos, y que se incluyen en los dispositivos autoevaluativos.

6. LOS OBJETIVOS DEL IACE

Mediante la utilización del IACE en las escuelas se procuran lograr los siguientes objetivos básicos:

- Instalar una cultura evaluativa en las instituciones escolares, mediante procedimientos e instrumentos sistemáticos y validados que son cumplimentados por los propios agentes escolares.
- Formular e implementar Planes de Acción para la Mejora de la Calidad Educativa en cada establecimiento, que resulten viables y actualizables y que se orienten a superar los problemas detectados durante los procesos autoevaluativos.
- Dejar capacidad instalada en las jurisdicciones para la posibilidad de una aplicación más autónoma de la herramienta autoevaluativa.
- Brindar insumos para la formulación de líneas de acción y políticas educativas jurisdiccionales a partir de las problemáticas y demandas priorizadas por las escuelas y sus propios agentes.

Ninguno de esos objetivos es sencillo de lograr, pero sin duda el más arduo es el último ya que implica la posibilidad de incidir en la formulación de las políticas públicas educativas de las jurisdicciones, compensando¹⁵ su tradicional direccionalidad desde arriba hacia abajo con esta otra desde abajo hacia arriba.

7. CARACTERIZACIÓN DEL IACE

El IACE brinda orientación y apoyo a los directivos y docentes de escuelas primarias para mejorar en forma continua las oportunidades de aprendizaje de la población escolar que tienen a su cargo. Es una herramienta amigable y flexible, concebida en la modalidad de autoevaluación a través de una serie de pasos y ejercicios.

Es considerada sinérgica (en el sentido antes apuntado) dado que de su utilización son esperables otros efectos adicionales al cumplimiento de los objetivos detallados en el ítem previo; por ejemplo: contribuir a la formación de los planteles escolares, mejorar el trabajo en equipo y el clima laboral dentro de los establecimientos educativos, puesto que se basa, sobre todo, en dinámicas grupales, en reflexiones conjuntas y en búsquedas de acuerdos, y promover mejores articulaciones con el contexto (sobre todo con las familias de alumnos y organizaciones sociales locales).

La herramienta fue concebida desde un *enfoque inclusivo y de los derechos de niños y niñas*; tales derechos pueden ser vistos desde dos planos de análisis: uno es el relacionado con el acceso a una educación de calidad que viabilice su desarrollo individual y colectivo y otro se vincula con la generación de un entorno protector de los derechos que permita detectar, orientar y atender o derivar casos de vulneración (real o potencial) de los mismos.

También contempla los *derechos de los adultos (docentes y familiares)* en términos de la participación que los mismos deben tener en cuanto a la identificación y priorización de problemas y la formulación de propuestas para la implementación de acciones de mejora en las escuelas donde trabajan (los docentes) o donde asisten sus hijos (en el caso de los familiares).

El IACE no agota la multifacética estructura y dinámica del fenómeno educativo, pero a efectos de viabilizar los procesos autoevaluativos fue necesario priorizar algunas dimensiones y variables. Así, las grandes *dimensiones o ejes analíticos* que se consideraron dentro del concepto de Calidad Educativa son las tres que siguen¹⁶:

¹⁵ Es necesario enfatizar en que no se piensa en sustituir esa direccionalidad, sino en complementarla.

¹⁶ Como se anticipó, de seis dimensiones que se consideraban en las dos ediciones previas del IACE, se pasó a un esquema tridimensional. A título ilustrativo, se detallan las seis dimensiones antes consideradas, cuyas variables fueron "reagrupadas": 1) La gestión institucional democrática; 2) Los logros de la escuela: aprendizajes e inclusión; 3) La gestión y orientación pedagógica de la escuela; 4) El clima escolar: vínculos y convivencia; 5) El entorno escolar protector de los derechos de los niños/as y 6) El espacio escolar: infraestructura y equipamientos.

- I. Logros y trayectorias educativas de los alumnos
- II. Perfiles y desempeños docentes
- III. Capacidades y desempeños institucionales

La primera de esas dimensiones se refiere al impacto de la educación en los niños, las niñas y adolescentes, reflejado en variables vinculadas a sus logros en las áreas de matemáticas, lengua y otras, así como en la adquisición de valores y en sus actitudes hacia el estudio; también se incluyen variables tales como el ausentismo, la repitencia y el abandono, que dan forma a las trayectorias educativas de los alumnos.

La segunda dimensión tiene que ver sobre todo con los perfiles de formación, las capacidades y los desempeños pedagógicos de los docentes. Se incluyen variables tales como la adecuación del desempeño docente respecto de las metas de la respectiva escuela y los objetivos de enseñanza, sus capacidades para trabajar en la diversidad –tomando en cuenta los distintos ritmos e intereses de los alumnos–, sus habilidades para trabajar con alumnos con capacidades diferentes, la inclusión de las actuales tecnologías de información y comunicación (TIC), las metodologías que usan para la evaluación de aprendizajes, sus conductas no discriminatorias en el aula, sus capacidades para detección y derivación de casos de vulneración de derechos de niños/as y adolescentes y su concurrencia a actividades de formación y actualización.

La tercera y última dimensión se refiere a la gestión institucional, particularmente la existencia en cada escuela de una misión compartida por los diferentes integrantes de la comunidad escolar; el estilo de la conducción, más o menos democrático y promotor de la participación (de docentes, del personal no docente, de alumnos y sus familiares); el clima escolar imperante, en términos de los vínculos entre los diferentes actores involucrados; la comunicación e información que el establecimiento brinda acerca de la gestión escolar (tanto hacia el interior como a su contexto exterior); la promoción de un entorno escolar protector de los derechos de niños/as y adolescentes; la articulación con otros niveles educativos, con otras instituciones sectoriales y con organizaciones locales diversas; se consideran también en esta amplia dimensión los aspectos estructurales relacionados con la planta física, los mobiliarios y los equipamientos.

El IACE prevé en cada escuela la realización de *cinco Ejercicios Básicos* diseñados en torno al esquema tridimensional recién sintetizado, los que se llevan a cabo en un mínimo de cuatro reuniones plenarias de trabajo del plantel docente (incluidos sus equipos directivos)¹⁷, complementadas por momentos de trabajo en pequeños grupos, de a pares, y/o en forma individual. Los títulos de esos ejercicios básicos son bastante autoexplicativos (Duro y Nirenberg, 2009):

17 Cada una de esas reuniones plenarias demandan media jornada escolar (alrededor de 4 horas).

1. Datos de la escuela para los últimos cinco años.
2. La misión de la escuela.
3. El significado de la Calidad educativa en la escuela.
4. Valoración de las dimensiones y variables de la Calidad educativa.
5. Plan de Acción para la Mejora de la Calidad educativa.

Se agrega también una *Encuesta de opinión a los Familiares de los alumnos*¹⁸, que indaga sobre los distintos aspectos de la Calidad Educativa, y que las propias escuelas distribuyen, recogen, procesan¹⁹ y analizan.

El quinto de los ejercicios básicos mencionados y con el cual culmina el proceso, brinda un sencillo formato de programación; cada escuela formula su plan en base a los problemas detectados y priorizados a través de la realización de los cuatro ejercicios básicos anteriores. Es ese el principal aporte de la autoevaluación.

EL IACE propone además un conjunto de *Ejercicios Opcionales* que las escuelas realizan según sean sus características y temas de preocupación prioritarios. Los temas que abordan esos ejercicios opcionales son: la planificación pedagógica, el conflicto, la violencia, las estrategias de inclusión y la atención a la diversidad, los vínculos con las familias y con las organizaciones sociales locales, y la protección de los derechos de niños/as y adolescentes (Duro y Nirenberg, 2009).

También se diseñaron instrumentos útiles para realizar el seguimiento del proceso aplicativo; se trata de las Grillas de registros (que son planillas con indagaciones a completar luego de realizar cada ejercicio básico u opcional) y la Grilla de apreciaciones sobre el proceso autoevaluativo en general (Duro y Nirenberg, 2009)²⁰, que cada escuela completa y entrega a sus supervisores/as.

El IACE postula que mientras el proceso autoevaluativo en cada escuela debería tener carácter autónomo y confidencial, sin interferencias por parte de autoridades u otros agentes, el producto de ese proceso, reflejado en el Plan de Acción para la Mejora de la Calidad Educativa requiere ser difundido y adquirir carácter público, de modo de lograr viabilidad para su ejecución y a la vez impactar en la formulación de políticas educativas.

¹⁸ Se trata de un cuestionario anónimo que los propios familiares completan; en caso que sean analfabetos, algún agente de la escuela (que no sea el maestro del niño/a respectivo/a), brinda ayuda para el llenado.

¹⁹ Un agente de la propia escuela realiza el procesamiento mediante una base de datos donde se vuelcan los formularios completados y así se obtienen en forma automática los tabulados de frecuencias (absolutas y relativas) y los gráficos.

²⁰ Véanse los formatos de las mencionadas Grillas en el documento IACE.

8. PAUTAS ORIENTADORAS PARA LOS NIVELES JURISDICCIONALES

Se redactaron unas Pautas orientadoras dirigidas a los funcionarios educativos de los niveles jurisdiccionales, cuyo propósito es que a la vez que ellos puedan contribuir autónomamente a una más eficaz utilización del IACE por parte de las escuelas a su cargo, también puedan consolidar y aprovechar los procesos y resultados del conjunto de establecimientos, para así incidir positivamente en la toma de decisiones y en la formulación de las políticas educativas respectivas.

Se subdividen esas pautas en tres grandes capítulos que tienen que ver con distintas fases o momentos del proceso de autoevaluación escolar:

- El primero se refiere a la transferencia del IACE²¹ a los directivos y docentes de las escuelas que se involucren en los procesos autoevaluativos; también se incluyen orientaciones acerca de la conformación de los equipos técnicos jurisdiccionales y una descripción de sus principales funciones o responsabilidades para un eficaz acompañamiento a las escuelas durante la implementación del IACE.
- El segundo provee la metodología y los instrumentos para realizar la sistematización y metaevaluación de los procesos de aplicación del IACE, una vez que los mismos se completaron en las escuelas de la jurisdicción²². Así se posibilitan dos cuestiones básicas: una es el aprendizaje emergente de tales aplicaciones para ajustar en forma permanente la herramienta autoevaluativa y los procedimientos aplicativos; y la otra es la obtención de una mirada holística, del conjunto de las aplicaciones en las escuelas de cada jurisdicción, de modo de extraer de allí insumos útiles para las decisiones y la formulación de políticas educativas que promuevan cambios positivos. Otra consecuencia relevante de la sistematización / metaevaluación es que posibilita realizar la *devolución*²³ de los hallazgos a los agentes escolares que se involucraron en la aplicación del IACE, para que puedan obtener una mirada

21 Se adoptó el método de “capacitación en cascada” que consiste en capacitar a un número de entre 50 a 60 escuelas seleccionadas, las que luego de concurrir a una jornada inicial y con el apoyo de profesionales del equipo técnico jurisdiccional, replican la transferencia / capacitación recibida no sólo a los planteles de sus propios establecimientos sino además a representantes de otras escuelas de su región educativa, logrando así mayor alcance.

22 Para entender la complejidad de ese proceso de sistematización / metaevaluación se enumeran sus principales pasos: 1) consolidación y análisis de las Grillas de Registro que cada escuela completó luego de efectuar cada ejercicio y de la Grilla de Apreciaciones sobre el proceso de aplicación que cada escuela completó al finalizar el mismo; 2) aplicación, procesamiento y análisis de cuestionarios autoadministrados a docentes, directivos y supervisores y 3) Realización de talleres de reflexión, mediante una guía orientadora, con docentes, directivos y supervisores, acerca del proceso de aplicación del IACE en sus escuelas, (se redacta una relatoría de cada taller y se sintetiza todo en un informe consolidado).

23 Si bien en la jerga evaluativa se usa el término devolución, en verdad se trata de la socialización de los hallazgos, como producto de la consolidación de los productos emergentes de los procesos autoevaluativos en cada escuela.

abarcativa del conjunto de las escuelas y poner en contexto lo ocurrido en el propio establecimiento.

- El tercer y último capítulo se refiere al monitoreo y acompañamiento de la ejecución de los Planes de Acción para la Mejora de la Calidad Educativa que formularon las escuelas, de modo de facilitar y optimizar su cumplimiento, así como prever las acciones de reprogramación que sean necesarias.

Lejos de ser taxativas, debe remarcarse que esas pautas son orientadoras para la acción de los niveles jurisdiccionales y deberán adecuarse a los contextos específicos donde el IACE se aplique.

9. LA APLICACIÓN DEL IACE EN LAS ESCUELAS: PRINCIPALES APRENDIZAJES

A partir de las sistematizaciones / metaevaluaciones realizadas de las aplicaciones mencionadas surgieron importantes aprendizajes, que a continuación se sintetizan.

— *Acerca de los tiempos para la aplicación*

- Con la excepción de las escuelas de Buenos Aires, la mayor parte logró cumplir, sorteando dificultades, con los cronogramas que elaboraron inicialmente. Los tiempos requeridos oscilaron entre los 2 y los 4 meses.
- En la mayoría de las escuelas fue necesario utilizar tanto tiempos institucionales como extra institucionales para poder completar la tarea.
- La falta de tiempo institucional específicamente adjudicado fue considerado el principal y mayor obstáculo que han tenido que afrontar para la realización “a conciencia” del IACE.

— *Sobre la modalidad de trabajo*

- En más de la mitad de las escuelas se conformó un grupo promotor que cumplió un rol importante tanto para las convocatorias como para la realización de los ejercicios²⁴.
- Sugieren reforzar la forma de instalar el IACE en las escuelas incluyendo acciones preliminares orientadas a los planteles docentes de cada establecimiento

²⁴ La conformación de los Grupos promotores surgió de la primera experiencia aplicativa en Tucumán; de allí que en la segunda y la tercera edición se lo incluyó como una pauta recomendada para el proceso aplicativo del IACE.

para su sensibilización, información y transferencia de los ejercicios, sus significados y procedimientos²⁵.

- Elogiaron la modalidad de trabajo que propone el IACE, que combina momentos de trabajo individual o en pequeños grupos con reuniones plenarias para discusiones y acuerdos.
- Resaltaron que en algunos casos fue arduo llegar a acuerdos en las reuniones plenarias, aunque reconocen como positiva la puesta en común de los diferentes puntos de vista.
- Hubo algunos casos en que se señaló como “dificultad” del proceso aplicativo la existencia de diferentes concepciones y opiniones en los planteles (por ejemplo: alrededor del concepto de “calidad educativa” o en relación con la “misión” de cada escuela).
- En varios casos reclaman mayor acompañamiento y asistencia técnica durante el proceso aplicativo; sobre todo para la tarea de moderación de las dinámicas grupales, pues no siempre se consiguió dar con el perfil y las capacidades apropiadas dentro de cada escuela para realizar esa tarea. Fueron varios los que plantean que ese rol de moderadores grupales lo realice alguien externo a la escuela.
- Algunos señalamientos fueron hechos acerca de la falta de espacios físicos adecuados en las escuelas para llevar a cabo las reuniones plenarias.
- Hubo quienes sugirieron ampliar la encuesta a todos los familiares e incluir, además de la encuesta, otras modalidades para lograr mayor y mejor compromiso de los familiares, no sólo en los procesos autoevaluativos y en las acciones para la mejora de la CE, sino también en los procesos educativos de sus hijos/as; al respecto han sugerido realizar reuniones del plantel docente con familiares²⁶.

— *Sobre la participación e involucramiento de los actores*

- La participación de los docentes en los procesos aplicativos fue considerada muy alta, aunque también hubo críticas a algunos por falta de compromiso

²⁵ La modalidad usada para instalar el IACE en cada jurisdicción se sintetiza: a partir de acuerdos con las autoridades, se realiza una jornada intensa de transferencia de la herramienta, con la concurrencia de un representante del equipo directivo y un docente por cada escuela a involucrarse. Las ulteriores acciones de multiplicación y convocatorias al resto de los planteles en cada escuela fue responsabilidad de los concurrentes a esa jornada inicial. A partir de estas recomendaciones, se decidió implementar una modalidad de “capacitación en cascada” donde se capacita a un grupo inicial que luego replica esa capacitación no sólo al plantel de la propia escuela sino a las restantes escuelas (en subgrupos).

²⁶ Vale aclarar que en el diseño preliminar del IACE (del año 2007, no publicado) se preveían dinámicas grupales entre docentes y familiares, las que a instancias de los docentes y directivos convocados para realizar el “juicio de expertos” para el ajuste del IACE preliminar, fueron eliminadas y cambiadas por la encuesta mediante cuestionario autoadministrable (sugerida por esos directivos y docentes como modo de minimizar el contacto personalizado entre docentes y familiares).

(reflejada no tanto en ausencias sino sobre todo por asistir a las reuniones sin una cabal lectura previa del cuadernillo).

- Hubo quienes plantearon que la participación docente fuera de carácter obligatoria para poder abarcar la totalidad del plantel.
- Para estimular el involucramiento de los y las docentes un factor que los mismos consideran de alta relevancia es el rol que asume la conducción de cada establecimiento y la importancia que el mismo adjudica al IACE, así como la coherencia entre sus discursos y sus acciones al respecto.
- Surgió como algo importante reforzar el protagonismo de los/las supervisores/as tanto para facilitar la realización del IACE como para su institucionalización.
- La importancia que la conducción educativa jurisdiccional otorga al IACE y el apoyo consecuente para su aplicación es crucial para que los planteles de los establecimientos se involucren. Ello debería reflejarse en el otorgamiento de tiempos y espacios institucionales y en la adjudicación de puntajes a los docentes (en términos de formación docente).

— *Sobre los ejercicios IACE*

- Todas las escuelas completaron los cinco ejercicios básicos. Fueron pocas las que además hicieron ejercicios opcionales (aduciendo falta de tiempo, no de interés).
- Las consignas para la realización de los ejercicios fueron consideradas fáciles de interpretar.
- Todos los ejercicios fueron calificados como muy útiles y relevantes para provocar la reflexión crítica y replantear la tarea pedagógica. Consideran que en conjunto aportan una mirada amplia y abarcativa de la vida institucional y que otorgan la posibilidad de adecuarse a las distintas realidades institucionales.
- Señalaron la existencia en los ejercicios de reiteraciones en variables y ciertas ambigüedades que deberían omitirse en el primer caso y precisarse en el segundo²⁷.
- Especialmente los docentes, valoraron los datos aportados por el ejercicio sobre datos acerca de los logros y la matrícula en las escuelas para la planificación

27 Se refieren dichas críticas a las primeras dos ediciones; se tomaron en cuenta esas recomendaciones para la tercera dedición, que fue revisada para evitar reiteraciones, superposiciones o ambigüedades.

de acciones en el aula²⁸, expresando que con los datos en la mano, contrastaban sus percepciones acerca de la calidad de la enseñanza; eso se refleja en el siguiente testimonio de una docente: “Pensábamos que hacíamos una tarea muy buena y nos dábamos cuenta que este año teníamos más alumnos que repetían”; “además que cada una de nosotras sabíamos de nuestros dos o tres alumnos con problemas, en la reunión con las estadísticas nos enterábamos que había alumnos con dificultad en casi todos los grados”.

— *Acerca de los Planes de Acción para la Mejora de la Calidad Educativa*

- Surgió la propuesta de incorporar el Plan de Acción para la Mejora de la calidad educativa al Proyecto Educativo Institucional (PEI)²⁹, lo cual facilitaría su concreción, dada la aprobación formal que el PEI implica.
- Los Planes de Acción para la Mejora de la calidad educativa fueron considerados altamente viables, dado que su formulación fue participativa y se hizo en forma adecuada a las problemáticas percibidas y priorizadas por los propios agentes y acorde a las características de cada establecimiento.
- La expectativa por la implementación de los Planes de Acción para la Mejora de la calidad educativa es alta. Sugieren realizar un meticuloso seguimiento y acompañamiento de esos procesos, para estimular su concreción, lo cual consideran un importante factor que facilitaría la institucionalización del IACE.

— *Sobre la institucionalización del IACE*

- Una sugerencia mayoritaria y reiterada fue acerca de la necesidad de institucionalizar el IACE, lo cual señalan que implica decisiones expresas por parte de las autoridades sectoriales, tales como la asignación de tiempos institucionales para realizarlo y el otorgamiento de puntaje docente como estímulo para una mayor participación de los planteles.
- Tanto los directivos como los docentes aconsejan la aplicación del IACE a escala en todas las escuelas, teniendo en cuenta los tiempos de implementación y el compromiso de parte de todos/as (docentes, docentes de materias especiales, directivos, supervisores, responsables jurisdiccionales, etc.).

28 No así los directivos, pues éstos conocen bien los datos de sus escuelas; se pudo detectar entonces que falta realizar difusión de esos datos hacia los docentes.

29 El PEI es el Proyecto Educativo Institucional que elabora cada establecimiento educativo en función de su misión y visión y objetivos a cumplir por cada escuela en un ciclo lectivo. Se supone que el PEI incorpora las prioridades educativas, contenidos, procedimientos didácticos y sistemas de evaluación y acreditación como así también aspectos de la escuela en relación con la comunidad y proyectos especiales.

- Algunos/as consideran que es posible aplicar el IACE cada tres años o realizar determinados ejercicios con más frecuencia y evaluar el Plan de Acción anualmente para ver los resultados obtenidos y en caso necesario ajustar la programación.
- Opinaron que los efectos de la implementación del IACE podrán mejorar la calidad educativa, y esto provocará una mejora en la escuela no sólo en cuanto a la gestión institucional sino a nivel pedagógico, siendo los alumnos los principales beneficiarios.

— ***Efectos inmediatos percibidos en la escuela por la sola aplicación del IACE, es decir sin aún aplicar el Plan de Acción para la Mejora de la Calidad Educativa***

- Mejoró la comunicación entre los docentes y de éstos con los directivos.
- Más contacto y mejor vínculo con los familiares de los alumnos (adjudicado esto a la realización de la encuesta).
- Satisfacción de los padres por sentir que sus opiniones son tomadas en cuenta en la escuela donde concurren sus hijos/as.
- Permitió tomar conciencia de la realidad de la propia escuela, una “mirada hacia adentro” que provocó considerar algunos aspectos que no eran reconocidos, hacerse planteos sobre la propia tarea, diferenciar aquellas cosas que obstaculizaban o facilitaban la calidad educativa, buscándolas a través de los consensos.
- Mayor comunicación, interacción e integración entre los distintos turnos de las escuelas.
- Se incrementó el sentimiento de pertenencia a la escuela.
- Se instalaron modalidades más democráticas de actuación.
- Se generó un espacio reflexivo y de autocrítica para el análisis de las prácticas docentes. Ese fue tal vez el efecto más valorado del IACE: proponer una reflexión sobre la práctica y la puesta en común de las miradas y experiencias.

Para finalizar este ítem vale incluir dos testimonios textuales de directoras de escuelas, que parecen resumir adecuadamente las apreciaciones recogidas sobre el IACE:

“La utilidad de los ejercicios a nuestro criterio consiste en direccionar la reflexión enfocándola, evitando dispersiones y en hacer un punteo exhaustivo de un número de variables bastante más amplio y abarcativo del que se suele tratar en las jornadas institucionales. Permite dirigir la mirada hacia problemáticas que preocupan y ocupan

en mayor o menor medida a todas las escuelas. Incluye la participación de la comunidad a través de la encuesta de opinión. En general consideramos que los ejercicios fueron muy útiles y adecuados para la autoevaluación de la calidad educativa”.

“Me voy enriquecida porque más allá de lo que uno pueda construir desde dentro de la Escuela, también tener estas herramientas, estas ayudas externas es valioso. A veces nos quejamos que estamos solos en el mundo y nos sentimos en pequeñas islas, pero poder participar de esta experiencia nos brindó un instrumento más que nos permite mejorar en nuestro trabajo, hacerlo un poco mejor cada vez (...) Salimos de las jornadas pensando, reflexionando y diciendo: bueno a ver esto que hasta ahora no había meditado, no había pensando, lo puedo poner en palabras y es un puntapié inicial hacia una cultura evaluativa diferente a la que vinimos aplicando en las Escuelas”.

10. LOS PLANES DE ACCIÓN PARA LA MEJORA DE LA CALIDAD EDUCATIVA DE LAS ESCUELAS Y SUS INFLUENCIAS EN LAS POLÍTICAS PÚBLICAS

Tal como ya se aclaró, los procesos de autoevaluación deben tener características de máxima autonomía y confidencialidad en el marco institucional de cada escuela, evitando interferencias de agentes externos o funcionarios jerárquicos del sistema educativo, para que los involucrados en tales procesos puedan opinar con absoluta libertad y espontaneidad. Sin embargo, el principal producto de ese proceso autoevaluativo, reflejado en el Plan de Acción para la Mejora de la calidad educativa, debe tomar estado público no sólo por motivos de transparencia de la gestión, sino además para concitar las adhesiones y apoyos necesarios y sobre todo para posibilitar la influencia en la formulación de las políticas educativas jurisdiccionales.

Por lo dicho, el IACE promueve que las escuelas realicen acciones de abogacía y de difusión amplia de sus planes, adquiriendo para ello un rol relevante los y las supervisores/as distritales, que son el nexo entre cada escuela y el nivel jurisdiccional.

Un primer comentario que surge de la revisión de esos Planes es que las escuelas han puesto el foco sobre todo en problemas posibles de resolver en forma interna, por ellas mismas. Seguramente la intención fue la de otorgar a la implementación de los mismos mayor viabilidad, evitando depender de decisiones y recursos externos. Un ejemplo de esto fue que si bien en todas las jurisdicciones un problema manifestado fue la carencia de profesionales que conformen los gabinetes psicopedagógicos, el mismo no fue elegido (o priorizado) para los Planes de Acción, debido a la presunción (basada en realidades) de las restricciones presupuestarias de los gobiernos educativos para resolución de tal problema. Como la resolución no es viable, el problema no se prioriza. Sin embargo, si un propósito relevante del IACE es influir en las políticas públicas educativas, debería promoverse especialmente el planteamiento de ese tipo de problemas a las autoridades.

Una segunda apreciación se refiere a las debilidades detectadas en materia de programación por parte de algunas escuelas. Pese a que en el IACE se brindó un for-

mato muy sencillo para el ejercicio de formulación del Plan, en varios casos se evidenciaron dificultades para formular los problemas o las líneas de acción en forma adecuada. Se concluye que no necesariamente las habilidades para la planificación pedagógica y la evaluación de aprendizajes (que sin duda poseen los docentes) son equivalentes o útiles a la hora de programar acciones o proyectos de intervenciones orientadas al cambio.

Otro aspecto que estaría reflejando la debilidad antedicha es la falta de integración de los planes presentados, que resultaron muy fragmentados como si los problemas priorizados no se vincularan entre sí y las acciones no tuvieran efectos integrales o sinérgicos, dado que suelen referirse a más de un problema a la vez³⁰.

Algunos de los *problemas priorizados* en el conjunto los planes fueron: repitencia, ausentismo, dificultades de aprendizaje en áreas básicas, escasa formación docente en tecnologías de información y comunicación (TIC), carencias en la formación continua de los/las docentes, falta de formación y capacitación docente para el tratamiento en el aula con niños que tienen “necesidades especiales” o “capacidades diferentes”, escasas capacidades y práctica docente en materia de evaluación del desempeño docente y de alumnos, carencias en las pautas de convivencia, no compartidas ni institucionalizadas, vínculos conflictivos, bajas capacidades en el tema de promoción de derechos, deficiencias en el vínculo de la escuela con las familias y con la comunidad local, fallas en la comunicación interna (dentro del plantel), baja participación de alumnos, insuficiencia e inadecuación de espacios físicos, del mobiliario, carencias en el mantenimiento y la higiene.

En muchos casos los planes hacen referencia a la falta de formación docente para el tratamiento en el aula con niños que tienen “necesidades especiales” o “capacidades diferentes”. Sería deseable precisar si esta problemática hace referencia a una sobreabundancia de niños/as con discapacidades psicofísicas (lo cual plantea ciertas dudas) o tiene que ver con las dificultades en el campo del desarrollo infantil (2 a 5 años) en las áreas de lenguaje verbal y no verbal y en los procesos intelectuales relacionados con el cálculo, la escritura y la lectura. En buena medida ese fenómeno tiene que ver con las “condiciones de educabilidad”³¹ en los ámbitos familiares y también a una carencia de recursos pedagógicos adecuados por parte de los docentes.

30 Es probable que el formato de programación ofrecido en las dos primeras ediciones haya favorecido tal fragmentación, por lo cual en la tercera edición dicho formato se sustituyó por otro que permite una mejor integración programática, de modo de superar esa dificultad.

31 La escuela, en tanto experiencia educativa formal, requiere de la presencia y eficacia de la denominada “educación primera” para su desarrollo. Cuando un niño ingresa a la educación básica debe haber pasado por esta formación previa, la cual está sobre todo en manos de sus familias. Si bien cada vez más los niños son escolarizados desde muy pequeños, en la mayoría de los países de América Latina la educación es obligatoria a partir de los cinco o seis años de edad, por lo que el paso de los niños por jardines infantiles o salas de preescolar es parte de las decisiones tomadas por los familiares en este proceso de educación inicial. El conjunto de aptitudes y disposiciones adquiridas o gestionadas en el seno del hogar conforman la base que condiciona fuertemente y hace posible los aprendizajes posteriores en la escuela. Para poder educar, hoy las escuelas esperan niños “ya educados”; ese es el sentido del concepto de “educabilidad”. Para mayor profundización véase López (2003).

En cuanto a las *actividades que los Planes proponen para superar los problemas priorizados*, la mayoría plantea generar espacios de diálogo, intercambio y reflexión conjunta ya sea entre los docentes, entre ellos y sus directivos y en algunos casos hacen extensiva esa propuesta para las familias y las organizaciones sociales. Varias plantearon la generación de espacios de encuentro entre docentes y familiares: clases abiertas, actividades recreativas y culturales, proyectos conjuntos, ferias, etc.

Esto puede estar indicando la necesidad de las escuelas de construir y fortalecer vínculos y de haber identificado como resultado del proceso de autoevaluación, la potencialidad del trabajo colaborativo.

Otra línea de acción que plantean en forma frecuente para la superación de los problemas priorizados, consiste en desarrollar estrategias comunicacionales tanto internas, orientadas a los planteles, como externas, hacia las familias, organizaciones sociales e instituciones sectoriales.

En los casos de falta de algún recurso humano se proponen reasignaciones de funciones de los propios planteles, como también elevar notas de pedidos formales a las respectivas autoridades.

Son reiterados los casos en que se refieren a formular o actualizar los códigos de convivencia, con protagonismo de familiares y alumnos.

Varios son los establecimientos que proponen refuncionalizar o aprovechar mejor espacios existentes, aunque también solicitan a las autoridades y/o procurarían donaciones para refacciones y ampliaciones de sus plantas físicas así como adquisición de mobiliarios y equipamientos.

La disponibilidad de PC, el acceso a Internet y la formación en cuestiones informáticas han sido también necesidades que se reflejarían en solicitudes concretas a las autoridades.

Todos sugieren la necesidad de acompañar estos procesos de programación e implementación con adecuada e intensa asistencia técnica y seguimiento, para no obturar el proceso de autoevaluación propositiva realizado y contribuir a la optimización y concreción de los planes formulados en procura de los cambios buscados. El compromiso de los gobiernos educativos jurisdiccionales en tal sentido es imprescindible.

11. COMENTARIOS FINALES

En primer lugar merece ser destacado que el IACE resultó una herramienta efectiva para promover una reflexión crítica acerca de la propia gestión y a la vez propositiva para el cambio en los establecimientos escolares.

Como se anticipó, un efecto secundario aunque relevante del IACE, es su componente formativo para aquellos que se involucran en sus aplicaciones³². Esa formación, que se podría encuadrar dentro de la modalidad de “capacitación en servicio” o “capacitación en la propia escuela”, no es exclusivamente en cuestiones de evaluación sino también en otras relacionadas tanto con la gestión institucional como con las funciones pedagógicas y la Calidad Educativa.

Con el IACE se contribuye además a romper ciertas concepciones y prácticas erróneas de muchos agentes educativos, acerca de que la autoevaluación consiste meramente en una serie de reuniones espontáneas e interactivas (frecuentemente de carácter catártico), sin un guión ni técnica preestablecida, para reflexionar en forma asistemática sobre los problemas, sin un encuadre metodológico básico donde se determinen con claridad los objetivos, las metodologías, las dimensiones, las variables e interrogantes orientadores, que permitan arribar a una mayor clarificación de los problemas y logros, así como proponer acciones para la mejora.

El IACE, tal como cualquier otro instrumento de evaluación, debe ser considerado como algo flexible y dinámico, en contraposición a algo cerrado o estático. Esa concepción se vincula con la que considera a la evaluación como una instancia de enseñanza-aprendizaje, pero en este caso no sólo para quienes son evaluados sino también para quienes diseñan la evaluación. En ese sentido uno de los principales aprendizajes es el que se refiere a la necesidad de revisión constante de los propios diseños evaluativos y por ende de las herramientas usadas.

La tercera edición surge en gran parte de las recomendaciones y sugerencias recogidas de quienes utilizaron el IACE en las jurisdicciones, pero además por la necesidad del nivel político nacional de aplicarlo en escala en todo el territorio argentino, para cuya viabilidad requirieron una versión más sintética de la herramienta.

Se pueden identificar ya algunas incidencias relevantes del IACE en las políticas educativas. En primer lugar, en los establecimientos donde ya fue aplicado, se han instalado prácticas de autoevaluación orientadas a la mejora, las que son altamente apreciadas como motores de cambio; a su vez, esas prácticas fortalecen las gestiones institucionales a través de reconocer aquellas acciones superadoras que cada establecimiento puede realizar por sí mismo, sin tener que aguardar intervenciones externas o depender en forma absoluta de recursos ajenos (Nirenberg, 2003).

En segundo lugar, los gobiernos educativos provinciales pudieron visibilizar la potencia de los procesos y resultados autoevaluativos y sobre todo, a partir de los hallazgos emergentes de las actividades de sistematización/metaevaluación, así

³² Es por ese carácter claramente formativo del IACE que el pedido de los agentes escolares a las autoridades jurisdiccionales, surgido en la mayor parte de las escuelas donde se lo aplicó, acerca de que se otorgue puntaje para aquellos que se involucran en los procesos autoevaluativos, resulta razonable, aunque su implementación no carezca de dificultades.

como de la consolidación de los Planes de Mejora, contaron con insumos válidos para mejorar los procesos de toma de decisiones y adoptar líneas de acción tendientes a la superación de aquellos problemas que resultaron denominadores comunes y que no son resolubles por la acción que cada escuela pueda llevar adelante por sí misma.

La herramienta facilitó una nueva modalidad de vinculación entre los niveles de los sistemas educativos, donde desde cada escuela (mediante dinámicas grupales y consensos) los problemas son detectados y priorizados y las propuestas de acciones superadoras son planteadas hacia sus niveles de supervisión y las autoridades jurisdiccionales.

Aunque su alcance aplicativo no muestre hasta la fecha altos valores absolutos y porcentuales, dado que para la ida a escala se requieren procesos graduales de utilización y apropiación del dispositivo por parte de los actores, puede afirmarse que el IACE ha sido adoptado como herramienta eficaz en materia de evaluación para la política pública en las tres provincias argentinas que lo utilizaron; asimismo la intención manifestada por las autoridades educativas nacionales ha sido promisorio en tal sentido, como fue señalado anteriormente. Se espera que la incidencia del IACE aporte a los esfuerzos colectivos para hacer real el derecho a una educación inclusiva y de calidad para todos y cada uno de los niños, niñas y adolescentes que asisten a la escuela primaria. Como se dijo al inicio, ese era el más ambicioso propósito del IACE y sus autoras.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Bolívar, A. (2000), *Los centros educativos como organizaciones que aprenden. Promesas y realidades*. Madrid: La Muralla.
- Duro, E. y Nirenberg, O. (2008), “Instrumento de Autoevaluación de la Calidad Educativa (IACE): un camino para mejorar la calidad educativa en escuelas primarias”. UNICEF Argentina. En línea: <http://www.ceadel.org.ar/unicefiace/> (consulta: 15 septiembre 2009).
- Habermas, J. (1982), *Conocimiento e Interés*. Vol. I. Buenos Aires: Taurus.
- López, N. (2003), “Educación y Equidad: algunos aportes desde la noción de educabilidad”, Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura, OEI. En línea: http://www.oei.org.ar/noticias/03062004_04.pdf (consulta: 17 septiembre 2009).
- Max-Neef, M. (1986), *Desarrollo a Escala Humana*. Uppsala: CEPAUR, Fundación Dag Hammarskjöld.
- Nirenberg, O. (2006), *Participación de adolescentes en proyectos sociales: aportes conceptuales y para su evaluación*. Buenos Aires: Paidós.
- (2008), “Aportes de la evaluación para el desarrollo de las organizaciones de la sociedad civil”, *Cuaderno de CEADEL 41*. En línea: http://www.ceadel.org.ar/cuadernos/Evaluac_Desarrollo_OSC-41.pdf (consulta: 17 septiembre 2009).
- (2009), “Modelos para la evaluación de políticas: insumos para una mejora de la gestión”, en M. Chiara y M.M. Di Virgilio, orgs., *Gestión de la Política Social: con-*

ceptos y herramientas. Buenos Aires: Universidad de General Sarmiento y Prometeo Libros.

Nirenberg, O., Brawerman, J. y Ruiz, V. (2000), *Evaluar para la Transformación innovaciones en evaluación de programas y proyectos sociales*. Vol. 8. Buenos Aires: Paidós.

— (2003), *Programación y Evaluación de Proyectos Sociales. Aportes para la racionalidad y la transparencia*. Vol. 19. Buenos Aires: Paidós.

Platón [1970], *Fedro*. Madrid: Instituto de Estudios Políticos.

UNESCO (2007), *Situación educativa de América Latina y el Caribe. Garantizando la Educación de Calidad para Todos. Informe Regional de Revisión y Evaluación del Progreso de América Latina y el Caribe hacia la Educación para todos en el marco del Proyecto Regional de Educación (EPT/ PRELAC)*. Santiago de Chile: OREALC/UNESCO.

Metaevaluación del sistema de evaluación de las titulaciones universitarias andaluzas*

Baltasar Fernández-Ramírez
Enrique Reboloso Pacheco
Pilar Cantón Andrés

Resumen: *El proceso de convergencia iniciado en Bolonia asigna un papel clave a la evaluación, como garante de la calidad de los títulos, para asegurar la confianza mutua entre los distintos sistemas educativos nacionales. Sin embargo, los modelos de evaluación no han sido revisados para incluir esta perspectiva legitimista, y tampoco se han incluido los requisitos que exige la internacionalización asumida en las declaraciones ministeriales como objetivo deseable del cambio universitario. En este artículo, se presenta una metaevaluación del sistema de evaluación de las titulaciones universitarias andaluzas, a través de la opinión del profesorado. Fueron entrevistados mediante cuestionario un total de 81 profesores con experiencia como miembros de comités de autoevaluación de las titulaciones. El cuestionario se compone de 82 criterios con los que se juzga la calidad del proceso de evaluación a lo largo de todas sus fases y elementos clave. Los resultados se interpretan como las características y valores deseables que debería cumplir el sistema de evaluación. Sugieren la existencia de una perspectiva certificacionista y convencional en el modo de entender la evaluación por parte del profesorado, obviando las posibilidades de un planteamiento político y las oportunidades que brinda para apoyar el proceso de internacionalización y convergencia europea.*

Palabras clave: *Evaluación universitaria, evaluación estratégica, proceso de Bolonia, calidad universitaria.*

1. EL PROCESO DE BOLONIA Y SUS OBJETIVOS. EL PAPEL DE LA EVALUACIÓN EN EL PROCESO DE CONVERGENCIA

La Declaración de Bolonia es un hito en las reformas universitarias de las últimas décadas. Un total de cuarenta y seis países están introduciendo cambios visibles de

* Esta investigación fue financiada por el Ministerio de Ciencia y Tecnología, en el marco del Proyecto SEC2003-03230/PSCE.

manera conjunta. La literatura es muy extensa, y el proceso continúa abierto, desde aquella segunda reunión ministerial en Bolonia, en 1999, hasta la última de Londres, en mayo de 2007, y la próxima convocatoria que será organizada por los países del Benelux en 2009¹. La documentación necesaria para comprender la filosofía, los objetivos, las iniciativas y el estado del proceso reformista es más asequible y reducida (abundan las repeticiones y los textos prescindibles que poco añaden).

Bolonia respondía a una situación y un diagnóstico concretos de los problemas de la universidad europea. La desaparición de la universidad de masas y la defensa del libre mercado en la educación superior, parecían situar a la universidad pública en una posición anquilosada, incapaz de competir frente a la oferta de las universidades a distancia, los títulos mixtos, los estudios profesionales privados y las delegaciones de prestigiosos centros de estudio extracomunitarios. La desbandada de alumnos que entonces se podía esperar ponía a nuestro sistema público en riesgo de desaparición o de merma seria. Las declaraciones ministeriales que siguen la ruta de Bolonia establecen casi desde el inicio una filosofía de trabajo que define la situación de partida, identifica los problemas de la universidad del momento y propone una lista relativamente breve de objetivos y estrategias para el cambio del modelo universitario europeo. La implantación extensa de los sistemas de evaluación forma parte del núcleo central de estrategias, con un papel clave en el aseguramiento de la calidad y en el aumento de la confianza mutua entre los estados firmantes.

Hoy la situación ha cambiado en algunos puntos. La crisis mundial del sistema financiero, la pérdida masiva de empleos y la recesión económica, han creado un nuevo escenario frente al cual los análisis son todavía confusos y caducan con rapidez, sobrepasados por los acontecimientos y la alarma continua de los indicadores macroeconómicos. Sin duda, la nueva situación tendrá repercusiones en el sistema universitario público, ya visibles en el fracaso inicial y la desaparición de un buen número de iniciativas de posgrado surgidas al amparo del discurso de la convergencia europea, y que no han conseguido atraer a suficiente número de alumnos para financiarse y justificar su continuidad. En consecuencia, al hablar del proceso de Bolonia, debemos tener un ojo puesto en las declaraciones y propuestas elaboradas durante esta última década, y el otro en el incierto futuro que nos depara la compleja situación actual de crisis económica y financiera.

Presentaremos en este artículo una metaevaluación del sistema de evaluación de las titulaciones universitarias andaluzas. Mediante cuestionario, fueron requeridas las opiniones de un grupo de profesores andaluces con experiencia como miembros de comités de autoevaluación. El cuestionario se utilizó para juzgar, desde el punto de vista del profesorado, la calidad del proceso de evaluación a lo largo de todas sus

¹ El calendario de actividades y los documentos centrales pueden ser consultados en la página oficial del proceso: <http://www.ond.vlaanderen.be/hogeronderwijs/bologna/>. Es obligada la lectura de las declaraciones ministeriales, que fijan los objetivos, instrumentos y plazos, y la lectura de los *trends*, informes periódicos sobre la implantación del proceso encargados por la EUA (*European Universities Association*).

fases y elementos clave. Los resultados se interpretan como las características y valores deseables que debería cumplir el sistema de evaluación. Antes de exponerlos, repasaremos brevemente los objetivos del proceso de Bolonia, el papel que dicho proceso reserva a las actividades de evaluación, y la situación y características de la evaluación de las titulaciones en las universidades andaluzas.

Según la propuesta oficial de Bolonia, el primer problema es la variedad de ofertas, estructuras y modelos de formación, así que el primer objetivo es crear una estructura común que aumente la legibilidad de la oferta educativa europea, dotándola de identidad y visibilidad ante alumnos y socios potenciales (*bachelor/master degrees*, ECTS, competencias). El siguiente paso es eliminar fronteras y flexibilizar el mercado de la educación superior para facilitar la movilidad estudiantil y profesional, promover el flujo e intercambio de experiencias y conocimientos y la libre elección de programas. La desconfianza inicial entre nuestros países² limita la eliminación de fronteras, así que el segundo objetivo es aumentar la confianza introduciendo o reforzando los sistemas de evaluación y acreditación existentes en cada país, entendidos como sistemas de aseguramiento y garantía de la calidad (ENQA, 2005), y mejorando los mecanismos de convalidación (*recognition*)³.

La preocupación por la empleabilidad de los egresados, la dimensión social (defensa del sistema universitario público), introducida en respuesta a las presiones iniciales de la organización europea de asociaciones de estudiantes (ESU), y otras cuestiones complementarias (*lifelong learning...*) completan el esquema central de los objetivos del proceso de convergencia. Las declaraciones ministeriales incluyen la lista oficial de objetivos, y desde aquí recomendamos la lectura de Haug y Tauch (2001) para un resumen argumentando de la cuestión de los objetivos e instrumentos principales del proceso.

Sólo dos cuestiones más de matiz. Muchos cambios que se están produciendo a nuestro alrededor no responden ni al espíritu ni a las recomendaciones iniciales, aunque se utilice el marco de Bolonia como argumento para justificarlos, mientras que algunos aspectos centrales no están teniendo eco en la reforma, toda vez que la competencia sobre legislación educativa pertenece a cada país (y a cada región, en el nuestro), y cada uno está interpretando y redefiniendo a su modo los requisitos de la convergencia europea⁴. Y en segundo lugar, en gran medida, el proceso de reforma se fundamenta en una estrategia mercantilista de la política universitaria, un intento de frenar las tendencias expansionistas de otros modelos de universidades alternati-

2 Las universidades y las autoridades estatales del país x no se fían de los títulos expedidos por las universidades de los países y , e incluso existen percepciones de este tipo entre universidades del mismo país. Ponga cada cual en x e y los países que formen parte de su experiencia o de su estereotipo.

3 La documentación está actualizada en: http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/Lisbon_Recognition_Convention.htm

4 Es justo decir que la definición inicial del proceso era escueta y ambigua, abierta a múltiples posibilidades. Visto de otro modo, quizá los cambios actuales no responden a las expectativas que muchos nos habíamos planteado al comienzo.

vas al modelo público, que intentan ser contrarrestadas haciendo más atractiva y competitiva nuestra oferta de estudios, para captar recursos económicos de un alumnado potencial que demanda determinado tipo de formación profesionalmente orientada⁵. Es un planteamiento que no convence a muchos (aunque la actual crítica de nuestros estudiantes es más bien ignorante de esta problemática), pero debemos ser conscientes de que este es el tema central que da significado a la reforma, al menos en sus años iniciales.

Ya hemos mencionado el importante papel que se asigna a la evaluación dentro del proceso de convergencia europea: asegurar la calidad de la oferta de formación de cada universidad y de cada país, para mejorar la confianza en términos de reconocimiento mutuo de los títulos expedidos en cada centro de estudio y de creación de redes universitarias de colaboración. Supuestamente, la movilidad profesional y la colaboración serán una realidad cuando cada país deje de dudar de las competencias profesionales de los egresados en otros países, permitiendo que un egresado pueda completar sus estudios o desarrollar una carrera profesional en un país diferente al de origen. Las fórmulas posibles son variadas: una universidad puede obtener un sello de calidad privado que la acredite ante las demás como una institución digna de crédito; el Estado puede crear o fortalecer mecanismos propios para elevar los requisitos que se imponen a una universidad antes de darle licencia para expedir títulos; un conjunto de universidades puede asociarse y definir un conjunto de requisitos comunes que les obliguen a mejorar en algún área relevante (investigación, empleabilidad, financiación, etc.), que aumenten su atractivo como centro de estudios o que sirvan para imponer normas de control a nuevas instituciones que quieran formar parte de la red. Todas estas opciones han sido ensayadas, y subyace en la diversidad la idea de acreditación múltiple (Van Vught, 1994), es decir, que cada universidad decida con cuál o cuáles sistemas desea ser acreditada, en la confianza de que eso mejorará su atractivo como socio para la creación de redes universitarias o para captar alumnos. La opción española parece vencerse hacia un refuerzo del papel ministerial (y autonómico, en virtud de las competencias transferidas) que define los requisitos y criterios para diseñar un título oficial de grado o de postgrado⁶.

5 Nuestras autoridades ministeriales han acuñado la expresión *profesionalizante*, para referirse a estos títulos de perfil más preocupado por dotar al alumno de competencias que optimicen su empleabilidad.

6 Este punto merece un comentario. Algunas ideas explícitas en los documentos iniciales del proceso de Bolonia abogaban por fórmulas flexibles en la creación de títulos, incluyendo títulos conjuntos entre diversas universidades, dentro y fuera de un mismo país, títulos mixtos, títulos especializados en campos multi- o interdisciplinares de alta demanda social, etc. En nuestro país, la burocratización ha impuesto un modelo rígido, donde toda flexibilidad y muchas oportunidades de innovación se han visto limitadas en la práctica casi hasta su desaparición. (Se podría decir, simplificando, que nos estamos limitando a pasar de títulos de cinco años, a títulos de seis, con cuatro años de grado y dos de posgrado. O sea, estamos volviendo atrás.) Sólo los títulos “oficiales” tienen ayudas en las convocatorias públicas, sólo sus alumnos pueden solicitar becas en estas convocatorias, y el afán invasor es tal que el ministerio pretende incluso reservar para sí la exclusividad en la utilización del término “máster” para denominar a un título de posgrado, obviando que el éxito social de los másteres durante estos últimos años han venido de la mano de iniciativas privadas, flexibles e independientes de los cauces formales ministeriales. Auguramos a la reforma un futuro incierto.

2. LA EVALUACIÓN INSTITUCIONAL EN ESPAÑA

La evaluación institucional se realiza en España desde los años noventa, siguiendo el modelo de cuatro fases extendido en Europa (Mora, 2004; Van Vught y Westerheijden, 1994). La agencia nacional responsable (Agencia Nacional de Evaluación de la Calidad y la Acreditación-ANECA) se coordina con las distintas agencias regionales para promover planes de evaluación de la calidad, con el objetivo de que todos los servicios de administración y todos los programas de formación sean evaluados en plazos aproximados de cinco años. En el caso andaluz, la Agencia Andaluza de Evaluación (AGAE)⁷ es responsable de definir y apoyar los planes de evaluación que se implantan en las universidades de la comunidad. La realización de evaluaciones es uno de los objetivos incluidos en el contrato programa firmado entre cada universidad pública y la Junta de Andalucía. Parte del presupuesto del centro está condicionado a la realización anual de cierto número de evaluaciones del profesorado, los servicios de administración y los programas de grado y postgrado.

Todos conocemos o hemos participado ya en algún proceso de evaluación universitaria, lo cual no quiere decir que todos tengan una idea clara y competencia crítica sobre la cuestión. La universidad española utiliza múltiples procedimientos de evaluación, bien se trate del proceso de selección del profesorado, la acreditación para la promoción profesional, la evaluación de los tramos de investigación, la acreditación de los nuevos títulos o la evaluación de las titulaciones y servicios universitarios. En este artículo sólo se discutirá sobre la evaluación de las titulaciones, que es el contexto en el que se aplica la metaevaluación que posteriormente describiremos.

Básicamente, se trata de un sistema de acreditación, revestido de un discurso político sobre la mejora y la calidad, y con ciertas peculiaridades en sus procedimientos y consecuencias. En primer lugar, la iniciativa de evaluarse no parte de los equipos rectores, sino de instancias políticas superiores, en contra de las recomendaciones usuales para la implantación de modelos de calidad. Tampoco los equipos decanales o los futuros participantes en los comités internos de autoevaluación tienen capacidad de decisión, y no está claro que tengan una idea ajustada sobre los objetivos del proceso, más allá de la mencionada retórica sobre la calidad y la responsabilidad social de la universidad⁸.

Después de recibir una jornada mínima de formación de dudosa efectividad, se constituye el comité interno de autoevaluación (formado usualmente por representantes del profesorado, estudiantado y personal de servicios), el cual se encargará

⁷ La AGAE sustituyó en el año 2005 a la anterior Unidad para la Calidad de las Universidades Andaluzas (UCUA), consorcio participado por las universidades y el gobierno regional, entonces a cargo del sistema de evaluación objeto de estudio en el estudio que aquí presentamos.

⁸ Una descripción crítica del proceso se encuentra en Reboloso, Fernández-Ramírez y Cantón (2009a).

de analizar la titulación y preparar el plan de mejoras. El trabajo del comité se centra en la reflexión acerca de un listado de áreas sobre los procesos de gestión de la titulación, tal como vienen propuestas en las conocidas guías de evaluación. La información “objetiva” disponible para apoyar la reflexión es poca y no siempre acertada, de tal modo que las reflexiones deben apoyarse en observaciones intuitivas e informales de los propios miembros del comité. Cuando la dinámica de trabajo resulta operativa, se plantea una reflexión de este tipo: “la guía sugiere aquí que deberíamos estar gestionando de este modo, o disponiendo de ciertas fuentes de información; no lo hacemos; por tanto, anotemos un punto débil”, el conjunto de los cuales conformará el plan de mejora que se incluya en el informe final. El planteamiento es abiertamente isomórfico (DiMaggio y Powell, 1983), es decir, se trata que la evaluación oriente al comité de autoevaluación en el esfuerzo de modificar las actuales prácticas de gestión de la titulación para aproximarlas a la norma que se hace explícita en la guía, con independencia de las características, necesidades y peculiaridades de la titulación o de la universidad en cuestión (Reboloso, Fernández-Ramírez y Cantón, 2005).

La elaboración del informe de autoevaluación es seguida de la visita de un segundo comité de evaluación externa, cuyo valor está por demostrar⁹. Tras una estancia corta y una serie de reuniones o audiencias públicas de muy escasa participación, el comité elabora un informe en el que comenta, critica y sugiere la introducción de cambios para mejorar el informe interno, en un tono de “crítico amistoso”, por utilizar la expresión de Lee J. Cronbach. Esta es una diferencia crucial con los sistemas de acreditación verdaderos, puesto que el informe externo no es vinculante, y la finalización del proceso sólo depende de la entrega de un posterior informe interno remodelado, que debe recoger las sugerencias del informe externo, según entendimiento y opinión del propio comité de autoevaluación.

Esta visión crítica sitúa la evaluación en un planteamiento superficial, en el que importa menos los resultados que el proceso en sí mismo. Lo importante es demostrar ante los responsables rectorales y políticos, que la evaluación ha sido realizada, es decir, entregar un informe formalmente correcto. La implantación de las mejoras es algo que trasciende a la evaluación en sí misma, más allá de la retórica sobre la calidad y la mejora continua que rodea al proceso. Si bien los evaluadores profesionales y académicos discuten a menudo sobre el problema del impacto de su trabajo, sabemos que las influencias que ejerce la evaluación en el cambio organizacional son muchas veces difusas, no intencionadas, indirectas y no esperadas (Reboloso, Fernández-Ramírez y Cantón, 2005). La cuestión de la utilidad del sistema de evaluación de las titulaciones queda abierta a investigaciones específicas que analicen de qué manera se está produciendo el impacto de la evaluación, cómo puede incrementarse, y si la mera ganancia en legitimidad ante la sociedad, nues-

⁹ Aunque Lawrenz, Keiser y Lavoie (2003) defienden la utilidad de las visitas externas y su validez como procedimientos de evaluación, son comunes las críticas a la competencia, la representatividad y la posible falta de posición crítica que asumen los comités de revisión por pares.

tros clientes y nuestros potenciales socios, es suficiente justificación para el esfuerzo de realizar las evaluaciones.

3. EL ESTUDIO EMPÍRICO REALIZADO

El estudio se enmarca en un proyecto de metaevaluación más amplio, aprobado como proyecto de investigación por el Ministerio español de Ciencia y Tecnología. Los destinatarios finales de la información eran los equipos de gobierno y responsables de procesos de evaluación de las universidades andaluzas, entre quienes se difundió un resumen ejecutivo de los principales resultados. Aunque hubo contacto y apoyo de estos para la recogida de información, en ningún momento se discutieron con ellos las cuestiones, objetivos y métodos de la evaluación. El proyecto pretendía juzgar la calidad de los procesos de evaluación de las titulaciones que se han venido llevando a cabo en las universidades andaluzas durante la última década.

El objetivo del presente artículo es analizar críticamente la perspectiva de los miembros del profesorado como participantes clave en los sistemas de evaluación y en el cambio estructural inscrito en el proceso de convergencia. Una perspectiva de metaevaluación es útil porque se intenta juzgar la calidad de la evaluación mediante la aplicación de un conjunto de criterios de valor que ayudarán a destacar cuáles son las principales virtudes y defectos del modelo actual. El estudio sigue el planteamiento conceptual y metodológico propuesto por Scriven (2000) y Stufflebeam (2001a) para la realización de metaevaluaciones. El resto del estudio puede consultarse en otras publicaciones de los autores (Reboloso, Fernández-Ramírez y Cantón, 2008, 2009b).

En este artículo se describe la parte cuantitativa del proyecto de metaevaluación desarrollado por el equipo de investigación desde el año 2004 en las universidades andaluzas. Se utiliza una técnica de encuesta on-line para facilitar el acceso a un número suficiente de docentes con experiencia en los comités de autoevaluación de sus titulaciones. Se describen a continuación los usuales apartados de muestra, instrumento, procedimiento y análisis.

— Muestra

Fueron entrevistados ochenta y un académicos de seis universidades andaluzas mediante un cuestionario voluntario *on-line*. 48.1% son mujeres, 83.9% tienen contrato a tiempo completo, 67.9% son profesores titulares o catedráticos de universidad, distribuidos homogéneamente entre las distintas Facultades y campos científicos, en representación de un variado número de titulaciones. La antigüedad media en la universidad es de 14.82 años (9.9% menos de cinco años, 23.4% entre seis y diez años de antigüedad). Todos han formado parte de comités de autoevaluación entre los años 1997 y 2005. 60.4% de estas evaluaciones se realizaron después del año 2000, finali-

zando la mayoría en el curso siguiente. Sólo en tres casos (3.7%), la evaluación se concluyó dentro del mismo año. 18.5% asumieron el papel de presidente o secretario de su comité de autoevaluación, mientras que 58.0% actuaron como vocales (los restantes no informan de este dato).

— *Instrumento*

El cuestionario incluye 82 ítems que representan a otros tantos criterios de metaevaluación. Está dividido en nueve bloques de información, definidos a partir de la guía para la evaluación de servicios utilizada en aquel momento por la UCUA (Unidad para la Calidad de las Universidades Andaluzas), responsable última de los procesos de evaluación (Reboloso, Fernández-Ramírez y Cantón, 2003). Los primeros cinco bloques contemplan criterios generales referidos a las funciones de la evaluación, contexto, mecanismos de información y control, resultados y guía de evaluación. Los restantes bloques repasan exhaustivamente las fases del proceso de evaluación, incluyendo la autoevaluación (interroga sobre la composición del comité, el proceso de autoanálisis y el informe inicial de autoevaluación), evaluación externa (composición del comité, proceso e informe externo), la elaboración de los informes finales (distinguiendo entre el proceso de elaboración y el resultado final) y las acciones emprendidas como consecuencia de la evaluación.

La selección de los criterios estuvo determinada por la utilización de cuatro metacriterios de orden superior (responsabilidad, mejora, actitudes democráticas y viabilidad del proceso). Se pretendía con ello que la metaevaluación fuera útil para orientar la revisión del sistema de evaluación, incrementando su utilidad para la mejora de las titulaciones, aportando valor ante clientes y potenciales entidades colaboradoras, reforzando el funcionamiento democrático de las evaluaciones y asegurando su viabilidad práctica. (Ver Reboloso, Fernández-Ramírez y Cantón, 2009a, para una discusión sobre el concepto de metacriterio y su utilidad para orientar el cambio organizacional en una línea deseada.)

En todos los casos, la cuestión es indicar hasta qué punto se piensa que la evaluación de las titulaciones debería ser realizada en línea con la cuestión mencionada en el ítem (escala de siete puntos, donde el valor siete indica el mayor acuerdo posible). Se intenta así conocer la posición valorativa de los académicos que han participado en procesos de evaluación, es decir, cómo creen que debería ser un proceso de evaluación ideal. (En todos los casos, también se les pedía que indicaran hasta qué punto la evaluación en la que participaron fue realizada del modo indicado en el ítem, aunque estos resultados no se expondrán aquí).

Este trabajo forma parte de un proyecto en el que también se utilizaron técnicas de investigación de índole cualitativa a través de la participación de los investigadores en grupos de autoevaluación (Reboloso, Fernández-Ramírez y Cantón, 2009b). Evidentemente, el formato de cuestionario es rígido y restrictivo en sus formas, pero

fue utilizado por sus ventajas para acceder *on-line* a una muestra amplia de personas pertenecientes a otras universidades del sistema andaluz. El cuestionario no incluyó ítems con formato abiertos de respuesta.

— *Procedimiento*

Se realizó un primer contacto con los responsables rectorales de los procesos de evaluación en cada una de las universidades andaluzas. En todos los casos, se remitió la dirección *on-line* de la encuesta junto a una carta de presentación del trabajo. Los propios responsables de cada universidad tuvieron la cortesía de difundirla entre los miembros del profesorado que hubieron participado en procesos de evaluación de las titulaciones en convocatorias anteriores. Las contestaciones a las encuestas se registraban directamente en una base de datos construida al efecto con el programa de encuestas de SPSS (*Statistical Package for the Social Sciences*). Antes de finalizar el proceso, se enviaron varias notas de recordatorio con el ánimo de recopilar un mayor número de casos. Los datos se recogieron entre los meses de enero y febrero de 2006.

— *Análisis*

Los resultados se presentan en sendas tablas para cada uno de los bloques del cuestionario. Cada tabla incluye el nombre del criterio de metaevaluación, el ítem que lo operacionaliza, así como las puntuaciones media y desviación típica resultantes. Para estimar la existencia de diferencias estadísticamente significativas entre las puntuaciones medias de cada bloque, la tabla incluye también la puntuación *alfa* y la *T² de Hotelling*. La primera mide la consistencia interna de la escala (oscila entre 0 y 1; el valor 1 indica un grado de coincidencia completo); la segunda contrasta la hipótesis nula de que todos los elementos de la escala tienen la misma media (una probabilidad $p < .05$ indica que las diferencias entre ítems no se deben al azar).

Como se verá, las evaluaciones ideales no alcanzan la puntuación máxima en ninguno de los criterios del cuestionario, permaneciendo siempre en valores positivos, pero no extremos. Igualmente, la consistencia interna es muy elevada en todos los casos, con puntuaciones superiores a 0.85 e incluso 0.90 en algunos casos.

4. RESULTADOS

4.1. Criterios generales

— *Funciones de la evaluación*

Además de las tradicionales funciones de mejora y responsabilidad (rendición de cuentas), se incluyen otras posibilidades relevantes en el marco actual de la evalua-

ción universitaria, caracterizado por el discurso de la convergencia europea y los procesos de acreditación (Sursock, 2003).

TABLA 1
Funciones de la evaluación

| criterio | ítem | ~ | S _x |
|--------------------|--|------|----------------|
| 1. Mejora | El proceso global de evaluación es relevante y útil para mejorar las actividades docentes, de investigación y de administración | 5.79 | 1.49 |
| 2. Autonomía | La evaluación aumenta la capacidad de la universidad para tomar decisiones de forma autónoma | 5.11 | 1.61 |
| 3. Responsabilidad | La evaluación es útil para facilitar información pública y transparencia a los estudiantes, el mercado laboral, el gobierno y la sociedad en general | 5.69 | 1.43 |
| 4. Innovación | La evaluación facilita la introducción de nuevos métodos docentes no convencionales | 4.62 | 1.65 |
| 5. Crítica | La evaluación ha servido para criticar las estructuras de poder universitario que generan situaciones de abuso | 4.16 | 2.21 |
| 6. Acreditación | La evaluación es útil para preparar los mecanismos para la acreditación de la universidad y sus titulaciones | 5.49 | 1.50 |

$\alpha = .732$.

$T^2 - Hotelling = 61.016, F = 11.577, p < .001$.

Fuente: Elaboración propia.

A primera vista, parece que el discurso oficial de la evaluación ha calado entre los participantes del estudio. Las evaluaciones deberían realizarse, en este orden, para mejorar las actividades (docentes, de investigación y de administración), para ofrecer información a la sociedad sobre el funcionamiento de la universidad y para apoyar a las titulaciones en sus procesos de acreditación. Incluso el criterio de autonomía es coherente con esta línea “tradicionalista”, respondiendo a la extendida idea de que la universidad debe rendir cuentas a la sociedad a través de la evaluación para justificar la autonomía en la toma de decisiones.

No está claro que estos resultados reflejen un conocimiento completo de lo que puede aportar la evaluación. Reboloso, Fernández-Ramírez y Cantón (2008 y 2009a) han mostrado que las evaluaciones están sirviendo aceptablemente para la función de responsabilidad, pero sólo de manera limitada la función de mejora. No obstante, los implicados parecen ignorar el potencial que tiene para las funciones de innovación y la crítica, especialmente relevantes en la situación actual de cambio. El proceso de convergencia se puede interpretar como un cambio superficial, ajustándose a las demandas mínimas de adaptación de la estructura de las titulaciones y de utilización

de sistemas de garantía de la calidad. O puede entenderse (en mi opinión, de manera más ajustada al espíritu inicial de Bolonia) como un reto para el sistema universitario, que tiene que adaptarse y sacar ventaja en un contexto social globalizado y altamente competitivo (ahora incluso en recesión económica), utilizando estrategias de internacionalización que potencien la movilidad, el intercambio y la creación de redes interuniversitarias. Cuando muchos de nuestros colegas demuestran un comportamiento reactivo, en espera de que las instancias políticas les dicten lo que deben hacer, se echa en falta mayor anticipación, y sobre todo, apoyo a iniciativas innovadoras que pueden dar protagonismo a nuestras universidades en el contexto internacional. La innovación y la crítica son funciones clave en el cambio organizacional, en las que la evaluación podría tener un papel que, a la vista de los datos, está siendo ignorado o relegado.

— *Contexto de la evaluación*

Son muchas las decisiones y preparativos anteriores al inicio de la evaluación, muchas de los cuales no serán trasladadas a los participantes en el proceso, a pesar de ser determinantes para el éxito del trabajo. Entendemos por contexto esta situación de partida, incluyendo cuestiones como la adecuación a las características e intereses de la titulación (Chen, 1990), la oportunidad para realizar la evaluación en el momento apropiado (Chen, 1990), la suficiencia de los recursos destinados (JCS, 1988; Stufflebeam, 2001b), la competencia de los implicados o el respeto a los participantes y sus intereses (AEA, 1995).

Los resultados sugieren que, idealmente, una evaluación no debería tener consecuencias negativas para los participantes, en términos de respeto a la seguridad, dignidad y autoestima de los mismos. Es decir, con una puntuación relativa muy elevada, la primera preocupación es personal, y no relacionada con los problemas de la gestión a los que la evaluación presta su interés. En segundo lugar, la evaluación debe estar en condiciones de desarrollarse de manera correcta, lo cual implica disponer de los recursos apropiados, y que los participantes tengan la formación y competencias adecuadas.

A pesar de sus elevadas puntuaciones, las cuestiones sobre el papel de la evaluación en los procesos generales de gestión (adecuación, ajuste, coordinación), quedan en un segundo término. El extremo se encuentra en el criterio de centralidad, según el cual parece que la evaluación no se percibe como un elemento clave de la gestión de la titulación, hasta el punto de que no parece relevante el momento en que se realice. Es un resultado llamativo, dado que el enfoque racionalista de planificación que da sentido a este modo de entender la evaluación exige precisamente que gestión y evaluación estén coordinadas, que la evaluación tenga un papel relevante en la toma de decisiones y que se realice en el momento en que se necesita, y no en otro (Patton, 1997). Paradójicamente, nuestros colegas parecen asumir que la evaluación sólo es un acontecimiento puntual periódico que responde a la iniciativa de instancias políticas suprauniversitarias, ajena a la realidad, los problemas y las características de la titulación, a la cual no sirve de gran cosa. No extraña así que lo importante para nuestros colegas sea que la

evaluación no les afecte de manera personal; como hemos visto, basta que los implicados sean competentes, y adecuados los recursos destinados al proceso.

TABLA 2

Contexto de la evaluación

| criterio | ítem | ~ | S_x |
|---------------------------|---|------|-------|
| 7. Adecuación | La evaluación respeta el perfil característico propio de la Titulación | 5.32 | 1.56 |
| 8. Ajuste | La evaluación se ajusta a las necesidades de la comunidad universitaria en materia de mejora | 5.30 | 1.53 |
| 9. Política de calidad | La evaluación está coordinada y forma parte de la política de calidad oficial de la institución | 5.58 | 1.59 |
| 10. Centralidad | La evaluación ocupa un lugar central dentro de las actividades generales de la Titulación | 4.65 | 1.73 |
| 11. Periodicidad | Existen procesos formales periódicos de evaluación, reforma y seguimiento de las titulaciones | 5.64 | 1.51 |
| 12. Igualdad | La evaluación ha tenido en cuenta los intereses de todos los grupos relacionados con la Titulación | 5.62 | 1.58 |
| 13. Competencia | Los evaluadores y los miembros de los comités de autoevaluación y evaluación externa cuentan con la formación y competencia adecuadas para realizar su tarea con la mayor calidad | 5.89 | 1.57 |
| 14. Recursos | Se ha dedicado el tiempo y los recursos necesarios a las actividades de evaluación | 5.89 | 1.33 |
| 15. Viabilidad del plan | El plan inicial de evaluación era viable en términos generales | 5.67 | 1.38 |
| 16. Respeto a la dignidad | Los responsables de la evaluación respetaron la seguridad, dignidad y autoestima de todos los participantes y demás personas relacionadas con el proceso | 6.48 | 0.88 |
| 17. Oportunidad | Dada la situación actual de la Titulación, la evaluación debería haber sido realizada en otro momento | 4.06 | 2.22 |

$\alpha = .861$.

$T^2 - \text{Hotelling} = 206.445, F = 17.081, p < .001$.

Fuente: Elaboración propia.

— Mecanismos de información y control de la evaluación

Para cumplir su función de mejora y responsabilidad, es necesario que la evaluación y sus resultados sean dignos de confianza para los participantes y para los distin-

tos destinatarios de la información de evaluación. En buena parte, la confianza se incrementa estableciendo controles sobre los datos y ofreciendo información sobre las decisiones centrales del proceso. La credibilidad y la transparencia son clave desde una perspectiva que subraya la dimensión política y pragmática de la evaluación (Guba y Lincoln, 1989; Patton, 1997). Para este tercer bloque, hemos reunido un conjunto heterogéneo de criterios derivados de los principios éticos de honestidad y responsabilidad hacia el bien general (AEA, 1995), la participación libremente aceptada (Reboloso, Fernández-Ramírez y Cantón, 2003) y el control de los datos (Bustelo, 2002; ENQA, 2005; JCS, 1988).

Los resultados de la tabla 3 muestran una valoración elevada y similar en casi todos los criterios de este bloque, si bien el estadístico de contraste es significativo ($p = .004$), lo cual sugiere que existen diferencias, al menos entre los valores extremos. Así, la autonomía y el control sobre los datos serían las cuestiones principales en una evaluación ideal. La disponibilidad de recursos y la independencia de los responsables del proceso es el criterio más valorado, junto a cuestiones que aseguran la calidad de la información y los datos utilizados (correcto procesamiento y análisis, coordinación en la recogida de datos, veracidad e imparcialidad). Ahora bien, los criterios con menor puntuación se nos antojan cruciales en una evaluación respetuosa con las formas políticas. Claro, no es que quiénes responden rechacen la importancia de conocer las características del proceso antes de su inicio, ni de resolver posibles conflictos previos de interés, pero su menor importancia relativa frente al problema de la autonomía y la credibilidad de la información (control de datos), nos remite nuevamente a una visión convencional de la evaluación como investigación, donde los criterios de validez tienen que ver con el discurso de la objetividad y la validez de los datos, y menos con las implicaciones políticas.

— *Resultados de la evaluación*

Con estos cuatro criterios, analizamos el valor que otorgan los participantes a los resultados generales de la evaluación. Como se aprecia, son valores muy genéricos, que hacen referencia a la calidad de la estructura y el proceso de evaluación (Stufflebeam, 2001b), así como a la satisfacción general (con dos ítems usuales en los estudios de satisfacción). Algo más concreto, hemos incluido un segundo ítem de la lista de control de Stufflebeam (2001b), el cual identifica la calidad con la obtención de resultados correctos, relevantes, seguros y de bajo coste.

Los cuatro criterios obtienen puntuaciones elevadas, algo menores que en anteriores tablas. El contraste estadístico indica que no hay diferencias significativas, así que concluiremos que todos son considerados deseables por igual. Cabe mencionar únicamente el segundo criterio de satisfacción, por cuanto su menor media y su mayor desviación típica indican que un porcentaje de respuestas anotan una puntuación inferior, es decir, que parte de ellos no desearían participar nuevamente en procesos

TABLA 3

Mecanismos de información y control de la evaluación

| criterio | ítem | ~ | S _x |
|---|---|------|----------------|
| 18. Difusión de información | Toda la información relevante durante el proceso ha sido difundida activamente y de manera comprensible entre todos los grupos implicados en la Titulación | 5.93 | 1.45 |
| 19. Exposición previa | Las características, ventajas y limitaciones del procedimiento de trabajo fueron expuestas antes de comenzar el proceso de evaluación | 5.72 | 1.59 |
| 20. Difusión de la cultura de la evaluación | Los responsables de gobierno realizaron actividades visibles al comienzo del proceso para promover la evaluación y asegurar la participación de los departamentos | 5.85 | 1.52 |
| 21. Libre aceptación | La participación en la evaluación parte de un compromiso libremente aceptado por todos | 6.02 | 1.43 |
| 22. Exposición de intereses | Los responsables de la evaluación informaron sobre qué se juega con la evaluación cada una de las partes implicadas (incluidos ellos mismos) | 5.84 | 1.72 |
| 23. Viabilidad política | La evaluación ha sido planificada y se ha desarrollado de manera colaborativa entre los grupos interesados | 6.05 | 1.23 |
| 24. Conflicto de intereses | Los conflictos de intereses entre los distintos grupos implicados en la Titulación han sido solucionados mediante la discusión, la negociación y la información | 5.74 | 1.52 |
| 25. Autonomía | La autonomía, independencia y recursos para los responsables de la gestión y realización de la evaluación estuvo garantizada en todo momento | 6.38 | 1.03 |
| 26. Implementación | El desarrollo de la evaluación se ha ajustado a la planificación inicial o se han incluido las correcciones oportunas cuando se necesitaron | 6.11 | 1.21 |
| 27. Recogida de datos | La coordinación de distintas personas y secciones facilitó la recogida de los datos necesarios para la evaluación | 6.20 | 1.16 |
| 28. Elaboración de la información | Toda la información recogida para la evaluación ha sido procesada y elaborada suficientemente | 6.31 | 1.02 |
| 29. Control de datos | La evaluación cuenta con mecanismos conocidos para controlar que la información utilizada por los comités es veraz e imparcial | 6.14 | 1.31 |

$\alpha = .950$.

$T^2 - \text{Hotelling} = 39.550, F = 2.996, p = .004$.

Fuente: Elaboración propia.

de evaluación si se les pidiera. Lo cual no quita, como decimos, que el deseo general de todos ellos es que las evaluaciones tengan resultados de alta calidad en diversos términos.

TABLA 4
Resultados

| Criterio | Ítem | \bar{x} | S_x |
|--------------------------------------|--|-----------|-------|
| 30. Valores sociales: excelencia | La estructura y el proceso de evaluación han alcanzado altos niveles de calidad | 5.98 | 1.36 |
| 31. Valores CIPP: resultados loables | Los resultados de la evaluación sirven a los usuarios correctos, son relevantes, seguros y han supuesto un coste razonable | 5.77 | 1.53 |
| 32. Satisfacción | La evaluación fue un éxito | 5.64 | 1.45 |
| 33. Satisfacción II | Volvería a participar en un proceso de evaluación si me lo pidieran | 5.44 | 2.03 |

$\alpha = .811$.

$T^2 - \text{Hotelling} = 8.024, F = 2.587, p = .061$.

Fuente: Elaboración propia.

— La Guía de evaluación

La guía detalla el proceso de evaluación, centrándose fundamentalmente en el procedimiento de autoevaluación. Su papel es indicar las áreas de información que deben ser objeto de análisis por parte del comité, facilitando y limitando al mismo tiempo las discusiones de trabajo. En este sentido, se constituye en el marco de referencia para la gestión de la titulación, imponiendo implícitamente que una gestión de calidad es aquella que coincide con los planteamientos de la propia Guía. Asumiendo esta perspectiva “certificacionista”, entendemos que una buena Guía es aquella que ofrece información suficiente y detallada para resultar útil a los diversos implicados (los diversos comités, los destinatarios de los informes) (AEA, 1995), y cuyos contenidos son comprensibles y relevantes, sin que la tarea requiera un esfuerzo más allá de lo necesario (Reboloso, Fernández-Ramírez, Cantón y Pozo, 2002).

Nuestros encuestados parecen coincidir sobre la importancia relativa de todos los criterios incluidos en este apartado. La diferencia indicada en el estadístico de contraste es debida al valor inferior asociado al criterio de esfuerzo. Dado que es un ítem formulado en sentido inverso (mayor puntuación debería significar un nivel de esfuerzo no deseado), no nos queda claro si el dato indica que los ítems de la Guía

requieren cierto nivel elevado de esfuerzo o si la formulación ha introducido alguna confusión en la respuesta. A pesar de lo cual, nos mantenemos en la idea de que la Guía ideal, efectivamente, debe informar sobre los procedimientos de evaluación, y resultar comprensible y útil en cada uno de sus contenidos.

TABLA 5
Guía de evaluación

| criterio | ítem | ~ | S_x |
|---|--|------|-------|
| 34. Enjuiciamiento | La guía describe con detalle el proceso de enjuiciamiento (cómo se valora la calidad, qué significa “suficiente” o “insuficiente”, cómo se llega a juicios finales, etc.) | 5.92 | 1.34 |
| 35. Contenido | La guía incluye información detallada sobre los objetivos y las actividades que hay que realizar en cada fase de la evaluación | 6.08 | 1.17 |
| 36. Objetividad: comprensión | Los conceptos y criterios que deben ser evaluados se comprenden fácilmente | 5.90 | 1.44 |
| 37. Responsividad: relevancia | Los conceptos y criterios que deben ser evaluados son claves para comprender el funcionamiento de la Titulación | 6.02 | 1.30 |
| 38. Eficiencia: esfuerzo | Se requiere un esfuerzo elevado para contestar cada cuestión | 4.08 | 1.78 |
| 39. Investigación sistemática: información sobre el procedimiento | Existe información sobre el procedimiento escrita con suficiente detalle como para que cualquiera pueda comprender, criticar e interpretar los resultados de la evaluación | 5.70 | 1.58 |

$\alpha = .761$.

$T^2 - Hotelling = 67.187$, $F = 12.610$, $p < .001$.

Fuente: Elaboración propia.

4.2. Fases del proceso de evaluación

— Fase 1. Autoevaluación

La autoevaluación es el elemento central de la evaluación de las titulaciones. En cierto modo, todos los demás elementos se ordenan para facilitar la autoevaluación y asegurar su éxito. Para analizar esta fase con detalle, distinguimos tres elementos: las características del comité de autoevaluación, las dinámicas de trabajo del comité y el informe de autoevaluación. Trataremos cada uno de estos apartados por separado. No obstante, para no recargar excesivamente el texto, hemos agrupado todos estos datos en la tabla 6. (Los valores alfa y T^2 de Hotelling de cada uno de los bloques de criterios se han incluido en la parte inferior.)

TABLA 6

Fase de autoevaluación

| Criterio | Ítem | ~ | S _x |
|--|---|------|----------------|
| <i>Sobre la composición del comité de autoevaluación^a</i> | | | |
| 40. Representatividad | Todos los estamentos y grupos implicados están representados en el comité (profesorado de distintas áreas y categorías, personal de administración, alumnos...) | 6.49 | 1.00 |
| 41. Respeto a las personas: equidad | Los miembros del comité han recibido compensaciones justas por su participación en el proceso | 5.20 | 2.30 |
| 42. Operatividad | El tamaño del comité facilita que las reuniones y discusiones de trabajo sean operativas | 6.34 | 1.08 |
| <i>Sobre el proceso de autoanálisis^b</i> | | | |
| 43. Actitudes democráticas | En las reuniones priman las actitudes democráticas y los valores de negociación y consenso | 6.43 | 1.31 |
| 44. Participación pública | Los miembros de la comunidad participan de distintos modos en el proceso aportando información, opiniones o críticas | 6.21 | 1.19 |
| 45. Deberes del personal: ejecución de la tarea | Todos los miembros del comité han cumplido con las tareas asumidas bajo su responsabilidad | 6.42 | 1.17 |
| 46. Gestión del tiempo | La duración de las actividades de la autoevaluación ha sido correcta | 6.01 | 1.53 |
| <i>Sobre el informe de autoevaluación^c</i> | | | |
| 47. Legibilidad | El informe está escrito de tal modo que se puede hacer una lectura relativamente fácil (lenguaje claro, tamaño reducido, capacidad de síntesis) | 6.20 | 1.32 |
| 48. Exhaustividad | El informe repasa todas las áreas de gestión relevantes y útiles para comprender y mejorar la Titulación | 6.22 | 1.32 |
| 49. Información basada en evidencias | Los resultados de la autoevaluación tienen como base datos suficientemente documentados | 6.25 | 1.34 |
| 50. Publicidad | Se dio suficiente publicidad al informe para conocer las opiniones de la comunidad | 6.31 | 1.22 |
| 51. Participación | Las opiniones de la comunidad fueron recogidas en el informe final para lograr mayor grado de aceptación y consenso | 6.22 | 1.25 |
| 52. Confianza: confirmación | Todos los resultados de la evaluación podrían ser confirmados por un evaluador independiente si fuera necesario | 6.40 | 1.02 |

^a $\alpha = .668$; $T^2 - \text{Hotelling} = 23.687$, $F = 11.655$, $p < .001$.

^b $\alpha = .898$; $T^2 - \text{Hotelling} = 10.259$, $F = 3.302$, $p = .027$.

^c $\alpha = .927$; $T^2 - \text{Hotelling} = 11.245$, $F = 2.096$, $p = .080$.

Fuente: Elaboración propia.

Representatividad y operatividad son los valores fundamentales en la composición ideal del comité de autoevaluación, de tal modo que estén representados todos los estamentos y grupos de interés implicados en la titulación, sin que el tamaño resultante sea excesivo (diversos autores recomiendan que el tamaño de los grupos oscile entre los siete y los once miembros, aunque a nosotros se nos antoja algo excesivo). La representatividad es un criterio subrayado por distintas agencias, entre las que se cuentan la Asociación Americana de Evaluación (AEA, 2005) o la Red Europea para el Aseguramiento de la Calidad (ENQA, 2005), con objeto de que todos los grupos afectados por los resultados de la evaluación estén informados y puedan hacerse escuchar.

Es llamativo el resultado del criterio de equidad, que interpretamos como la compensación por el tiempo y esfuerzo que los miembros del comité dedican a la evaluación (AEA, 2005). Nuestra experiencia es que los participantes realizan un esfuerzo complementario sin ver reducidas sus responsabilidades cotidianas, por lo que parece lógico esperar cierto tipo de compensación, y de hecho hemos encontrado quejas al respecto en numerosas ocasiones. Sin embargo, el promedio y la desviación típica resultantes en este caso, indican que, al menos cierto número de respuestas no consideran este criterio al mismo nivel de importancia que los relativos a la composición del comité.

Respecto al proceso de autoanálisis, destacan los criterios relacionados con la dinámica de trabajo del comité. Es deseable lograr un clima democrático, donde las discusiones se desenvuelvan en un ambiente de consenso y negociación, y que todos los miembros del comité participen activamente, cumpliendo con las responsabilidades asumidas ante los demás. En un estudio de seguimiento de las sesiones de autoevaluación, Reboloso, Fernández-Ramírez y Cantón (2009b) han observado la importancia que los miembros de un comité dan a la presencia de personas que eluden y descargan sobre los demás tareas que deberían ser comunes, así como las actitudes y sentimientos negativos que se desarrollan en un ambiente de discusiones aciagas donde las opiniones no pueden expresarse con libertad de manera relajada. También Stufflebeam (2001b) recomienda la asunción compartida de responsabilidades como una buena práctica en los equipos de evaluación.

En un segundo plano, se considera deseable implicar a la comunidad universitaria y que aporte información y críticas. Las guías recomiendan este apartado de manera especial, para evitar que los trabajos del comité queden como un ejercicio ignorado y ajeno para la mayoría de los implicados en la titulación, reduciendo su potencial para tener efectos en la mejora y para que la evaluación sea vivida como una actividad central en los procedimientos de gestión (Reboloso, Fernández-Ramírez y Cantón, 2003). Por otra parte, la cuestión de la gestión del tiempo (Bustelo, 2002) es un problema de operatividad, con la idea de que el proceso no se extienda en exceso y pierda el ímpetu inicial, aunque nuestros resultados indican que es un criterio menor en comparación con los anteriores.

Finalmente, todos los criterios planteados para valorar el informe de autoevaluación reciben puntuaciones similares, sin diferencias estadísticamente significativas. El criterio más importante es la fiabilidad, es decir, la idea de que el resultado de la evaluación podría ser confirmado por un evaluador externo. Es uno de los clásicos criterios de validez al uso en las ciencias sociales, en relación con valores tradicionales de objetividad, posibilidad de réplica e independencia del observador, y puede ser importante para aumentar la credibilidad que los destinatarios del informe otorguen a los resultados de la evaluación (Guba y Lincoln, 1989). El resto de los criterios incluidos en este punto reciben una consideración similar, como decimos. El informe debe leerse sin dificultad (legibilidad), incluir todas las áreas de gestión relevantes (exhaustividad), utilizar datos contrastados (información basada en evidencias) e incluir las opiniones de la comunidad (publicidad y participación). Nosotros entendemos el informe de evaluación como un documento informativo que tiene destinatarios y objetivos muy concretos, entre los cuales no sólo está informar, sino también convencer, con un sentido estratégico, para lograr el efecto deseado (*intended uses for intended users*, Patton, 1997). Desde este punto de vista, coincidimos en la importancia de los valores que aseguren la credibilidad del informe y la viabilidad de las acciones propuestas (Reboloso, Fernández-Ramírez y Cantón, 2005).

— Fase 2. Evaluación externa

Los evaluadores externos desarrollan la visita con el objetivo de valorar si la autoevaluación se ha desarrollado según lo previsto. El procedimiento es similar a los sistemas de acreditación (recogen datos, observan las instalaciones, realizan entrevistas y cumplimentan sus hojas de control), pero el informe no es determinante. Asesoran y hacen recomendaciones para mejorar la autoevaluación, pero su opinión puede ser más o menos ignorada en el informe final y, más importante aún, la finalización del proceso de evaluación depende de la entrega del informe final de autoevaluación y el correspondiente plan de mejora, que no vuelve a ser revisado por el comité externo ni está obligado a incluir sus recomendaciones.

Los encuestados esperan del comité de evaluación externa que sea imparcial, objetivo y que se muestre sensible a las características y peculiaridades de la titulación evaluada (Jeliazskova, 2002). Se atiende así a la función prevista para la visita externa, es decir, supervisar que el procedimiento y la información que ha dado lugar al informe de autoevaluación, lo hacen digno de confianza. Con una puntuación algo inferior, se considera deseable también que el comité de autoevaluación esté informado sobre los objetivos de la evaluación externa, el procedimiento a seguir durante la visita y los criterios que se utilizarán para redactar el informe externo (como recomendación, por ejemplo, ENQA, 2005). Estos elementos serán útiles para reducir posibles temores, una vez que el comité de autoevaluación acepte y conozca las características, funciones y actitudes del equipo de evaluación externa, legitimado a su vez ante los destinatarios finales del proceso de evaluación (Lawrenz, Keiser y Lavoie, 2003; van Vught y Westerheijden, 1994).

TABLA 7
Fase 2. Evaluación externa

| criterio | ítem | ~ | S _x |
|---|--|------|----------------|
| <i>Sobre el comité de evaluación externa^a</i> | | | |
| 53. Aceptación | Los miembros del comité de autoevaluación y sus representantes conocían previamente los objetivos, el procedimiento y los criterios de la evaluación externa | 6.16 | 1.34 |
| 54. Imparcialidad | La imparcialidad y objetividad de los miembros del comité externo está fuera de duda | 6.44 | 1.20 |
| 55. Sensibilidad | Los miembros del comité externo han sido sensibles a las características y necesidades específicas de la Titulación evaluada | 6.29 | 1.28 |
| <i>Sobre el proceso de evaluación externa^b</i> | | | |
| 56. Orientación evaluativa | La visita externa estuvo orientada a la emisión de juicios sobre la calidad del proceso y el informe de autoevaluación | 6.07 | 1.31 |
| 57. Duración | La duración de las actividades de la evaluación externa ha sido correcta | 6.20 | 1.28 |
| 58. Participación pública | Todos los grupos implicados en la Titulación dispusieron de medios para hacer llegar sus opiniones y críticas durante la visita de los evaluadores externos | 6.35 | 1.14 |
| 59. Coordinación | Los procesos de evaluación interna y externa están coordinados | 6.30 | 1.27 |
| <i>Sobre el informe de evaluación externa^c</i> | | | |
| 60. Función de mejora | El informe externo ayuda a la Titulación a mejorar su calidad | 6.14 | 1.36 |
| 61. Consenso | El informe externo refleja el consenso de todos los miembros del comité externo | 6.11 | 1.21 |
| 62. Información basada en evidencias | Las opiniones y juicios de los expertos externos tienen como base datos suficientemente documentados y contrastados | 6.30 | 1.20 |
| 63. Legibilidad | El informe está escrito de tal modo que se puede hacer una lectura relativamente fácil (lenguaje claro, tamaño reducido, capacidad de síntesis) | 6.24 | 1.25 |
| 64. Oportunidad | El informe de evaluación externa fue recibido tras un plazo razonable después de la visita del comité | 6.38 | 0.97 |
| 65. Participación | El comité de autoevaluación y la comunidad tuvieron oportunidad de revisar y criticar el informe externo antes de su entrega definitiva | 6.31 | 1.19 |
| 66. Confianza: confirmación | Todos los resultados de la evaluación podrían ser confirmados por un evaluador independiente si fuera necesario | 6.39 | 1.04 |

^a $\alpha = .852$; $T^2 - \text{Hotelling} = 13.142$, $F = 6.445$, $p = .004$.

^b $\alpha = .891$; $T^2 - \text{Hotelling} = 3.598$, $F = 1.152$, $p = .338$.

^c $\alpha = .878$; $T^2 - \text{Hotelling} = 7.360$, $F = 1.090$, $p = .385$.

Fuente: Elaboración propia.

No se encuentran diferencias estadísticas entre los criterios utilizados para valorar el proceso de evaluación externa. Se considera muy deseable que toda la comunidad tenga ocasión de opinar y estar presente de algún modo durante la visita, contribuyendo a mejorar las cualidades democráticas del proceso (Lawrenz, Keiser y Lavoie, 2003). Lo cual no deja de ser un deseo algo irreal, ya que las reuniones previstas entre el comité externo y los distintos grupos implicados suelen ser ignoradas por la mayoría, y apenas acuden grupos muy reducidos de personas interesadas. No tenemos claro si es un problema de falta de acciones que motiven a la participación, de dificultades de agenda, o sencillamente si la evaluación externa pasa desapercibida y no es valorada como una oportunidad para hacerse oír e influir en los cambios que se derivarán de la evaluación.

Además, por razones operativas, es deseable que la duración de la visita sea la correcta y que esté coordinada con el ritmo de la evaluación interna, para no generar disfunciones o pérdidas de tiempo (Bustelo, 2002). Igualmente, es importante hacer notar que la visita externa tiene por objetivo supervisar el proceso y el resultado de la evaluación interna, y no tanto poner en cuestión la calidad de las prácticas de gestión de la titulación evaluada.

En cuanto al informe externo, todos los criterios propuestos reciben valoraciones elevadas. El informe debe ser útil, lo que supone que esté escrito de manera comprensible (ENQA, 2005), que sea entregado dentro de un plazo razonable (Rebolloso, Fernández-Ramírez y Cantón, 2003) y que ayude a la mejora de la titulación. Además, debe ser digno de confianza, es decir, que las opiniones del comité externo estén basadas en datos contrastados y documentados, que sean fruto del consenso entre sus miembros y que pudieran ser confirmadas por un evaluador independiente (Guba y Lincoln, 1989; Lawrenz, Keiser y Lavoie, 2003). La confianza mejora también si los miembros del comité interno han tenido oportunidad de revisarlo críticamente y de hacer llegar sus comentarios al comité externo antes de la entrega del informe definitivo (Van Vught y Westerheijden, 1994).

— *Fase 3. Elaboración del informe final*

Una vez recibido el informe externo, el comité de autoevaluación revisa y corrige su informe propio en aquellos puntos que les parezcan convenientes. El texto final incluye un resumen de los principales resultados, señalando debilidades y fortalezas, y propone un plan de mejoras. Finalmente, el borrador se difunde para promover la participación de la comunidad y recoger sus opiniones antes de cerrar el informe final. Los criterios incluidos en este apartado inciden, pues, en la participación, en cómo se integran ambos documentos, y en cómo se definen, se priorizan y se asegura la implantación de las mejoras propuestas (Van Vught y Westerheijden, 1994).

TABLA 8

Fase 3. Elaboración del informe final

| Criterio | Ítem | ~ | S_x |
|--|---|------|-------|
| <i>Sobre el proceso de elaboración del informe</i> | | | |
| 67. Participación pública | El borrador y la versión definitiva del informe final han sido publicitados para recoger opiniones y comentarios de los grupos implicados | 6.30 | 1.12 |
| <i>Sobre el informe final de evaluación de la Titulación</i> | | | |
| 68. Difusión | El informe final se difundió a través de diversos medios entre los diversos grupos interesados | 6.21 | 1.36 |
| 69. Exhaustividad | El informe final incluye información exhaustiva sobre el contexto y el proceso de evaluación, los puntos fuertes y débiles de la Titulación y los planes de mejora | 6.48 | 0.95 |
| 70. Perspectiva integral | El informe final es una síntesis de los informes de autoevaluación y evaluación externa | 6.62 | 0.72 |
| 71. Legibilidad | El informe final está escrito de tal modo que se puede hacer una lectura relativamente fácil (lenguaje claro, tamaño reducido, capacidad de síntesis) | 6.46 | 0.95 |
| 72. Perspectiva crítica | El informe final señala las discrepancias y coincidencias de valoración surgidas entre los comités de autoevaluación y evaluación externa | 6.22 | 1.07 |
| 73. Perspectiva amplia del cambio | Las propuestas de mejora se centran en la realización de cambios estructurales, organizativos y tecnológicos, sin limitarlos al ámbito interno del servicio y más allá de la mera demanda de recursos adicionales | 6.14 | 1.09 |
| 74. Priorización | Las propuestas de mejora ponen en relación las deficiencias detectadas en los distintos apartados de la guía, priorizando los cambios más urgentes | 6.38 | 0.97 |
| 75. Viabilidad | Las propuestas de mejora son viables en términos políticos, prácticos y económicos | 6.28 | 1.16 |
| 76. Previsión | En las propuestas de mejora se analizan los obstáculos previsibles y se proponen estrategias para superarlos | 6.10 | 1.38 |

^a No hay valores de comparación, puesto que sólo un ítem forma este bloque de criterios.

^b $\alpha = .922$; $T^2 - Hotelling = 19.033$, $F = 2.001$, $p = .074$.

Fuente: Elaboración propia.

Nuestros encuestados consideran muy deseable el cumplimiento de todos los criterios propuestos (las puntuaciones *alfa* y T^2 indican un elevado nivel de similitud entre todos los ítems). Así, las opiniones recogidas en el documento deberían ser ampliamente difundidas, y escuchadas y asumidas las voces críticas del resto de la comunidad y de los expertos externos. El riesgo de sesgar el informe se reduciría ante los ojos de sus destinatarios, reforzando la credibilidad del informe final y la viabilidad del plan de mejora.

En segundo lugar, una concepción amplia del cambio (no reducido a la usual demanda de medios económicos, sino plasmado en propuestas concretas de actuación), la priorización de las reformas urgentes, la previsión de dificultades para implantar las actuaciones, además de la propia idea de viabilidad política, práctica y económica, aseguran un plan de mejora con posibilidades de implantación y éxito. No obstante, debe aclararse que las actuaciones que finalmente se llevarán a cabo, no son una decisión unilateral del comité de autoevaluación, sino el resultado de un posterior proceso de negociación con los responsables rectorales, como veremos en el apartado siguiente.

— *Fase 4. Acciones post-evaluación*

Es la etapa más descuidada del proceso, hasta el punto de que las guías de evaluación apenas le dedican unas breves sugerencias. De hecho, la evaluación se considera terminada con la entrega del informe final, y se echa en falta una guía que ayude a los comités responsables en la negociación, planificación, implantación, seguimiento y evaluación de los resultados de las acciones de mejora. Los criterios incluidos en este último bloque promueven la autonomía de los comités de mejora y el apoyo técnico y político necesario para su buen funcionamiento.

TABLA 9
Fase 4. Acciones post-evaluación

| Criterio | Ítem | \bar{x} | S_x |
|-------------------------------------|---|-----------|-------|
| 77. Autonomía | El comité de calidad de la Titulación dispone de autonomía para iniciar planes y estrategias de mejora | 5,87 | 1,59 |
| 78. Utilidad: seguimiento e impacto | La evaluación continúa para ayudar a los participantes a comprender los resultados y emprender las acciones adecuadas | 6,16 | 1,19 |
| 79. Mejora basada en evidencias | El plan de mejora se deriva de los resultados de la evaluación y no responde a un esfuerzo premeditado por captar los fondos de ayuda a la mejora | 6,08 | 1,50 |
| 80. Carácter institucional | Los equipos de gobierno y otros grupos de la institución se comprometen con el desarrollo del plan de mejora | 6,18 | 1,39 |
| 81. Recursos | Los recursos necesarios para implantar el plan de mejora están disponibles | 5,82 | 1,74 |
| 82. Control | Los planes de mejora disponen de procedimientos de seguimiento previstos e implementados de manera consistente | 6,13 | 1,21 |

$\alpha = .861$.
 $T^2 - \text{Hotelling} = 2.303, F = .417, p = .834$.
Fuente: Elaboración propia.

Los datos estadísticos muestran que las opiniones son homogéneas. Se considera que el comité responsable de las acciones de mejora debe disponer de autonomía, de apoyo por parte de los equipos de gobierno y otros grupos de la comunidad, así como de los recursos necesarios para llevar a cabo el plan. En su vertiente técnica, es necesario definir procedimientos de seguimiento consistentes (ENQA, 2005) y el comité necesita asesoramiento en forma de evaluación continua a lo largo del proceso (JCS, 1988). Además, el plan de mejora debe entenderse como la consecuencia de los resultados de la evaluación, y no como una estrategia para la captación de los fondos previstos para las actuaciones.

Finalmente, estos resultados merecen dos comentarios generales. Primero, el notable descenso en el número de encuestados que contestan estos ítems, hasta alcanzar en torno al 50% (estos datos no se muestran, si bien hemos encontrado un incremento progresivo de las respuestas en blanco a partir de la fase de evaluación externa). Y segundo, el descenso en las puntuaciones de los criterios en comparación con los anteriores bloques. Pensamos que ambas cuestiones son consecuencia del peso reducido de la fase de mejoras dentro del proceso global de evaluación. Ya hemos dicho que se le presta poca atención en las guías, el retraso en la negociación sobre las mejoras financiables es considerable en muchos casos, y no forma parte de los compromisos que cada universidad está firmando con la Junta de Andalucía (los contratos-programa)¹⁰.

5. CONCLUSIONES

La amplia lista de criterios se concibió como un instrumento exhaustivo para mejorar el sistema de evaluación de las titulaciones en cuatro aspectos o metacriterios: el valor de las evaluaciones como herramienta para la rendición de cuentas (legitimidad), su capacidad para contribuir a la mejora de las titulaciones, la promoción de actitudes democráticas en los procesos de gestión y la viabilidad global de las propia evaluaciones (Reboloso, Fernández-Ramírez y Cantón, 2008 y 2009a).

El estudio permite hacerse una imagen detallada de las opiniones del profesorado sobre los procesos de evaluación. De forma muy simplificada, creemos que el profesorado está mediatizado por una perspectiva convencional de investigación, según la cual, la evaluación es un proceso de análisis objetivo y científico de la realidad organizacional, fruto del cual se dispone de sugerencias evidentes para la mejora de las titulaciones. De manera alternativa, nuestra opinión es que la evaluación es un proceso fundamentalmente político, cuyo principal objetivo es la legitimación de la titula-

¹⁰ Cada Universidad ha negociado un “contrato” particular con la Secretaría General de Universidades, según el cual se compromete a mejorar una serie de indicadores de gestión, a cambio de lo cual recibirá cierto porcentaje de su presupuesto, que queda así condicionado. En este caso, el indicador correspondiente compromete a la realización de cierto número de evaluaciones anuales, las cuales no incluyen el proceso de negociación e implantación del plan de mejoras.

ción ante la comunidad, los responsables políticos y, potencialmente, ante posibles socios en la creación de redes interuniversitarias. No se trata de descubrir *ninguna verdad objetiva* de los procesos de la organización universitaria, sino de demostrar a distintos socios y grupos de interés que la titulación es gestionada según ciertos estándares normativos nacional e internacionalmente aceptados. No en vano, el sistema de evaluación cuenta con la aprobación directa de la agencia nacional de evaluación universitaria (ANECA) y tanto la agencia nacional como la agencia andaluza forma parte del ente europeo sobre la materia (ENQA).

La pregunta que resta es de qué modo sirve al proceso de convergencia un sistema de evaluación como el nuestro, que ha sido concebido y diseñado desde una óptica pre-Bolonia. Nada se menciona en él sobre aspectos clave en el proceso de internacionalización y convergencia, y no está definido a priori cómo el sistema de evaluación ayudará a la implantación de los cambios estructurales que ya se están introduciendo en nuestras universidades. No nos cabe duda de que la evaluación podría jugar un papel central en estas cuestiones, pero queda por hacer un esfuerzo de actualización de los sistemas de evaluación, el cual no parece formar parte de las agendas políticas correspondientes.

A pesar de sus méritos (la dinámica democrática, por ejemplo), el sistema de evaluación actual no deja de tener cierto aire *administrativista* o *certificacionista*, que impone cierta rigidez. Dada la variedad de universidades, titulaciones y grupos implicados, se necesitan mecanismos que flexibilicen las exigencias y las áreas relevantes de análisis, para que el sistema pueda ser adaptado en situaciones variadas, por ejemplo, las que impone la existencia de diferentes ritmos, acciones y compromisos con la adaptación a la filosofía de Bolonia.

Ahora bien, no somos optimistas al respecto, y más bien tememos un futuro incierto para este tipo de sistemas de evaluación universitaria. La falta de centralidad de las evaluaciones, aparentemente ajenas a los procesos de gestión a los que supuestamente deben servir, y la dejadez con que parecen afrontarse los procesos de mejora, nos hacen pensar en una pérdida progresiva de motivación entre los equipos rectorales y los participantes directos en los comités de trabajo. Si lo único que se persigue es el “sello de calidad”, ya tenemos antecedentes variados que muestran cómo desaparece el interés por un sistema de evaluación entre los equipos directivos con la aparición de nuevos sellos y nuevas modas en el sector correspondiente.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

AEA [American Evaluation Association] (1995), «Guiding principles for evaluators», en W.R. Shadish, D.L. Newman, M.A. Scheirer y C. Wye, eds., *Developing the guiding principles. New Directions for Program Evaluation*, 66: 19-26. San Francisco, Ca: Jossey-Bass.

- Reboloso, E., Fernández-Ramírez, B., Cantón, P. y Pozo, C. (2002), «Metaevaluation of a total quality management evaluation system», *Psychology in Spain*, 6(1): 12-25.
- (2003), *Guía de evaluación de servicios universitarios*. Almería: Unidad para la Calidad de las Universidades Andaluzas.
- (2005), «The influence of evaluation on change in management systems in educational institutions», *Evaluation*, 11(4): 465-481.
- (2008), «Responsibility of Educational Institutions for Strategic Change», *Journal of MultiDisciplinary Evaluation*, 5 (10): 5-20. En línea: <http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/181/197> (consulta: 15 noviembre 2009).
- (2009a), *Crítica y mejora de los sistemas de evaluación universitaria*. Almería: Education & Psychology I+D+i.
- (2009b), «Quality criteria for self-evaluation in higher education», *Journal of MultiDisciplinary Evaluation*, 6(11) : 16-31. En línea: <http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/203/211> (consulta : 15 noviembre 2009).
- Bustelo, M. (2002), «Metaevaluation as a tool for the improvement and development of the evaluation function in public administrations», ponencia presentada en la *V Biennial Conference of the European Evaluation Society*, Sevilla, 10-12 de octubre.
- Chen, H.T. (1990), *Theory-driven evaluations*. Newbury Park, Ca: Sage.
- DiMaggio, P.J. y Powell, W.W. (1983), «The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields», *American Sociological Review*, 48(2): 147-160.
- Enqa [European Association for Quality Assurance in Higher Education] (2005), *Standards and Guidelines for Quality Assurance in the European Higher Education Area*. En línea: <http://www.enqa.eu/files/ENQA%20Bergen%20Report.pdf> (consulta: 5 marzo 2009).
- Guba, E. G. y Lincoln, Y. S. (1989), *Fourth generation evaluation*. Newbury Park, Ca: Sage.
- Haug, G. y Tauch, C. (2001), *Towards the European higher education area: Survey of main reforms from Bologna to Prague. Summary and conclusions*. En línea: http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/EUA_Trends_Reports/TRENDS_II-April2001.pdf (consulta: 5 marzo 2009).
- JCS [Joint Committee on Standards for Educational Evaluation] (1988), *The personnel evaluation standards: How to assess systems for evaluating educators*. Newbury Park, Ca: Sage.
- Jeliazskova, M. (2002), «Running the maze: Interpreting external review recommendations», *Quality in Higher Education*, 8(1): 89-96.
- Lawrenz, F., Keiser, N. y Lavoie, B. (2003), «Evaluative site visits: A methodological review», *American Journal of Evaluation*, 24(3): 341-352.
- Patton, M.Q. (1997), *Utilization-focused evaluation*. Newbury Park, Ca: Sage.
- Rossi, P.H., ed. (1982), *Standards for evaluation practice*. San Francisco, Ca: Jossey-Bass.

- Scriven, M. (2000), *The logic and methodology of checklists*. En línea: <http://www.wmich.edu/evalctr/checklists/papers/logic&methodology_oct05.pdf> (consulta: 5 marzo 2009).
- Stufflebeam, D.L. (2001a), *Evaluation contracts checklist*. En línea: <<http://www.wmich.edu/evalctr/checklists/contracts.pdf>> (consulta: 5 marzo 2009)
- Stufflebeam, D.L. (2001b), «The metaevaluation imperative», *American Journal of Evaluation*, 22(2): 183-209.
- Sursock, A. (2003), «Reflections from the higher education institutions' point of view. Accreditation and quality culture», en D.F. Westerheijden y M. Leegwater, eds., *Working on the European dimension of quality*. Zoetermeer: Ministerie van Onderwijs, Cultuur en Wetenschappen.
- Van Vught, F.A. (1994), «Intrinsic and extrinsic aspects of quality assessment in Higher Education», en D.F. Westerheijden, J. Brennan y P.A.M. Maassen, eds., *Changing contexts of quality assessment. Recent trends in West European Higher Education*. Utrecht: Lemma B.V.
- Van Vught, F.A. y Westerheijden, D.F. (1994), «Towards a general model of quality assessment in higher education», *Higher Education*, 28(3): 355-371.

G_aP *ABSTRACTS*

Studies

Development of Evaluation Centers and Training of Evaluators

ROBERT E. STAKE

Key words: Standards based evaluation, responsive evaluation, evaluation centers, training in evaluation, binocular vision.

Abstract: Evaluation practice, as a formal activity, has been and is still characterized by three key elements: (i) the supremacy of a way of reasoning based on criteria and standards, (ii) the promotion of the use of tests and other quantitative techniques and (iii) the idea of quality as an intrinsic attribute of programs. Training in evaluation has followed the same line and it has mainly focused on models and methodological approaches. Many problems arise with such a perspective when identifying the effectiveness of an educational system. Another (often devalued) kind of analytical thinking is also needed: that of episodic thinking, which emphasizes (i) that program should be properly contextualized and (ii) that program quality depends greatly on who experiences it. Both kinds of thinking are required. They must be combined as if they were two sides of the same coin. Therefore, “binocular vision” is the best option (commitment) for evaluation.

Searching for an interpretation of educational achievements: limits and possibilities of international evaluation studies

ALEJANDRO TIANA FERRER

Key words: Comparative education, evaluation of educational achievement, large-scale evaluation, index of economic, social and cultural status (IESC), school effectiveness

Abstract: International studies aimed at evaluating educational achievement, started in the mid-20th century and extensively developed in the two last decades, had a two-fold purpose: on the one hand, they intended to support a rigorous comparison of results achieved by students in different education systems, providing as a result a valuable set of data which has improved our knowledge about educational achievement. On the other hand, they intended to contribute to the explanation of

educational achievement through the settlement of relationships between the main factors identified. This paper analyses the latter issue, focusing on the possibilities opened by these studies and their contributions to such an explanation, and the limits and difficulties found for explaining the effect of the main factors influencing achievement. Two main sets of variables are identified. A first group includes some external factors to education systems, and a second group focuses on internal variables related to educational organisation and activities. Taken into account the macroscopic perspective adopted by those international studies, it is argued that the information provided for that kind of explanation is forcefully limited. Granted that conclusion, they should be complemented by additional research and evaluation studies allowing more detailed analyses.

The administrative capture of evaluation

SAVILLE KUSHNER

Key words: Evaluation independence, program evaluation, politics

Abstract: The past 30 years has seen a gradual internalisation of program and policy evaluation into the machinery of government. This happens either through the evaluation function being taken on by administrators themselves, or by those administrators stipulating and proscribing how the evaluation is conducted and for what purposes. This poses important questions for democracy and for the citizen. This article examines the consequences of this ‘capture’ by the administrative system by looking at two examples: one, taken from the field of international development; the other from a government agency commissioning of an evaluation of one of their programs.

Experiences and cases

Public policy evaluation at strategic level: general grant system, early school-leaving and human capital improvement

M.^a LUISA MERINO CUESTA

Key words: Public policies evaluation at macro level, strategic evaluation, evaluation of large clusters

Abstract: The main aim of this paper is to contribute to the knowledge of public policies evaluation at the macro level. It presents the synthesis of an evaluation performed by the National Agency for Public Policies Evaluation (AEVAL), demanded by the Spanish Cabinet. The report is

available at the Agency website. This evaluation examines the particularities of the evaluation at the strategic and operative levels of government, and presents the difficulties found to evaluate large clusters. Lessons learned are presented following the concept of the Agency about the evaluation process: analysis of the demand, planning, implementing, and results.

Autoevaluation and Public Policies: an experience in Argentina's primary schools

ELENA Duro

OLGA Nirenberg

Key words: Plans for improvement, institutional management, educational quality

Abstract: Is it possible to influence or to produce positive changes in the public policy by means of self-evaluation procedures about the performance of organizations which offer public services? Under the premise that self-evaluation promotes “intelligent organizations” (who learn and change from their own experiences) such way is strategic and possible, although not the only one. This article reports the processes and results of the implementation, during the period 2007-2009, of an Instrument for self-evaluation of educational quality (Instrumento para la Autoevaluación de la Calidad Educativa-IACE) in primary schools of Argentina. It was carried out within the institutional framework of UNICEF Argentina in agreement with the Center for the Support of Local Development (Centro de Apoyo al Desarrollo Local-CEADEL), Argentinean Non-Governmental Organization who works in public policies, and mainly in evaluative interventions. The self-evaluation turned out an effective strategy to influence educative public policies. Therefore the traditional top-down directionality in making decisions and defining policies processes was replaced by a complementary bottom-up alternative, where school's human resources and students' families outlooks and suggestions about future actions are reflected.

Metaevaluation of the system for evaluating Andalusian University Grades

BALTASAR FERNÁNDEZ-RAMÍREZ

ENRIQUE REBOLLOSO PACHECO

PILAR CANTÓN ANDRÉS

Key words: Evaluation in Higher Education, strategic evaluation, Bologna process, quality in Higher Education

Abstract: The Bologna process of European convergence assigns evaluation a key role, as a system for quality assurance of grades, in order to increase the mutual trust among different national education systems.

However, current evaluation models neither have been revised to include such a legitimistic perspective, nor have assumed the criteria for internationalization included in the official goals of Bologna declarations. The paper describes, through the opinion of academic staff, a metaevaluation of the evaluation system of grades used in the Andalusian universities. A questionnaire was administered to 81 academics, chosen because of their previous experience as members of self-evaluation committees. The questionnaire included 82 criteria for judging the quality of evaluation stages and key elements. Results show the valued characteristics that any evaluation system should fit. Academics have a certificationist conventional perspective over evaluation, rejecting a political alternative perspective and the opportunities added by evaluation in supporting the efforts of internationalization and European convergence.

G_aP

AUTORES

PILAR CANTÓN ANDRÉS

pcanton@ual.es

Profesora Titular de Psicología Social de la Universidad de Almería. Miembro del grupo de investigación “Evaluación de programas en contextos de evaluación social” de dicha Universidad, cuyos resultados se han reflejado en diversas publicaciones científicas. Especialista en evaluación de programas educativos.

ELENA DURO

eduro@unicef.org

Profesora de Ciencias de la Educación de la Universidad Nacional de La Plata. Ha cursado estudios de posgrado en metodología de la investigación en políticas educativas y gerencia de programas. Responsable del área de Educación en UNICEF Argentina, desde el año 2000. Directora de Planeamiento Educativo del Gobierno de la Provincia de Buenos Aires y consultora del Instituto Interamericano de Derechos Humanos (IIDH) con sede en Costa Rica. Ha ejercido docencia en temas de psicología educacional, didáctica y planificación en la Universidad Nacional de La Plata. Sus publicaciones se centran en temas de educación, derechos de la infancia, trabajo infantil, evaluación educativa y planificación de políticas públicas.

BALTASAR FERNÁNDEZ-RAMÍREZ

bfernan@ual.es

Profesor Titular de Psicología Social de la Universidad de Almería. Miembro del grupo de investigación “Evaluación de programas en contextos de evaluación social” cuyos resultados se han reflejado en diversas publicaciones científicas. Ha sido Director Técnico de la Unidad de Calidad e Investigación de Evaluación (1998-2003) y Director del Máster en Evaluación de la Universidad de Almería. Especialista en psicología de la ciudad y evaluación de programas de intervención social.

ESTER GARCÍA SÁNCHEZ

estgarsa@polsoc.uc3m.es

Doctora en Ciencias Políticas y Sociología por la Universidad Complutense de Madrid y Profesora Titular Interina de Ciencia Política y de la Administración en la Universidad Carlos III de Madrid. En esta Universidad desempeña los cargos de Secretaria Académica del Instituto de Política y Gobernanza “Fermín Caballero” y Subdirec-

tora del Máster en Gestión y Análisis de Políticas Públicas, que se imparte en colaboración con el INAP. Desde 2005 es Vocal de la Junta Directiva de la Sociedad Española de Evaluación. Especialista en evaluación y metaevaluación de políticas públicas, especialmente en el campo de la educación, temas en los que ha participado en diversos proyectos de investigación. Entre su obra publicada, destacan sus artículos en revistas como *Gestión y Análisis de Políticas Públicas*, *Andamios* y *Revista del CLAD. Reforma y Democracia*.

SAVILLE KUSHNER

saville.kushner@uwe.ac.uk

Doctor en Educación por la University of East Anglia (Norwich) y Catedrático de Evaluación en la University of West England (Bristol). Sus investigaciones giran en torno a la evaluación de programas, tanto en su vertiente teórica como aplicada. Ha sido asesor de la Oficina Regional para América Latina y el Caribe de UNICEF, en materia de seguimiento y evaluación de programas. Entre sus publicaciones destacan la obra coeditada con Nigel Norris, *Dilemmas of Engagement: Evaluation and the New Public Management* (Oxford: Elsevier, 2007), y *Personalising Evaluation* (Londres: Sage, 2000), traducida al castellano con el título de *Personalizar la evaluación* (Madrid: Morata, 2002).

M.^a LUISA MERINO CUESTA

mluisa.merino@aeval.es

Evaluadora especializada de la Agencia Estatal de Políticas Públicas y Calidad de los Servicios (AEVAL). Licenciada en Filosofía y Letras por Universidad Autónoma de Madrid y Máster en Evaluación de Programas y Políticas Públicas por la Universidad Complutense de Madrid. Funcionaria del Cuerpo de Sistemas Informáticos de la Administración General del Estado y Evaluadora EFQM por el Club de Calidad de Madrid. Posee experiencia directiva y gestora en procesos de planificación y gestión de cambio organizacional, tanto en el sector público como en el privado. Ha sido la encargada de la formación en evaluación de políticas públicas en los programas de posgrado y especialización en los que colabora la AEVAL. Autora de diversos artículos, guías y materiales didácticos sobre evaluación de políticas públicas.

OLGA NIRENBERG

info@ceadel.org.ar

Doctora en Ciencias Sociales por la Universidad Nacional de Buenos Aires y Diplomada en Salud Pública por la citada universidad. Preside la ONG Centro de Apoyo al Desarrollo Local (CEADEL), que en alianza con UNICEF Argentina coordina proyectos sobre evaluación de Calidad Educativa. Es consultora de la Fundación

ARCOR en cuestiones de gestión de proyectos y de evaluación. Ha participado como evaluadora en diversos programas sociales nacionales e internacionales. Ejerce docencia y cuenta con diversas publicaciones en materia de diseño y evaluación de programas sociales.

ENRIQUE REBOLLOSO PACHECO

erebollo@ual.es

Catedrático de Psicología Social de la Universidad de Almería. Ha estudiado en las Universidades de Western Michigan y California-Los Angeles (UCLA), en las que se especializó en evaluación de programas educativos. Ha sido Director de la Unidad de Calidad e Investigación de Evaluación de la Universidad de Almería (1994-1996 y 1998-2004). En esta Universidad dirige el grupo de investigación “Evaluación de programas en contextos de evaluación social”, que se ha distinguido por la realización de diversos proyectos relacionados con la evaluación institucional y la evaluación de la calidad ambiental, así como sobre la “Metaevaluación de los procesos de evaluación de las titulaciones de las universidades andaluzas”, temas todos ellos que han dado lugar a diversas publicaciones.

ROBERT E. STAKE

Doctor en Psicología por la Universidad de Princeton y Profesor Emérito de la Facultad de Educación de la Universidad de Illinois (Urbana-Champaign), en la que dirige, desde su creación, en el Center for Instructional Research and Curriculum Evaluation (CIRCE). Es una de las figuras más destacadas y de mayor prestigio internacional en el ámbito de la evaluación. Ya en la década de los setenta desarrolló el llamado *enfoque de evaluación comprensiva* y es uno de los más destacados promotores del método de estudio de caso en evaluación. En 1988, fue distinguido con el *Premio Lazerfeld* de la American Evaluation Association. Es Doctor Honoris Causa por las Universidades de Uppsala (Suecia) y Valladolid. Cuenta en su haber con más de un centenar de publicaciones sobre evaluación educativa.

ALEJANDRO TIANA FERRER

alejandro.tiana@oei.es

Doctor en Filosofía y Letras (Pedagogía) por la Universidad Complutense de Madrid y Catedrático de Teoría e Historia de la Educación de la Universidad Nacional de Educación a Distancia. Es Director General del Centro de Altos Estudios Universitarios de la Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI). Ha sido Secretario General de Educación, así como Director del Centro de Investigación y Documentación Educativa (CIDE), del Instituto Nacional de Calidad y Evaluación (INCE) y de la International Association for the Evaluation of Educational

Achievementy. Es autor de dieciocho libros y más de cien capítulos y artículos de revista sobre temas relativos a los sistemas educativos contemporáneos y evaluación educativa.

*CRITERIOS PARA LA PRESENTACIÓN
Y PUBLICACIÓN DE ORIGINALES*

G^a_pP

CONDICIONES DE PRESENTACIÓN PARA SU PUBLICACIÓN EN LA REVISTA

- Los autores deberán enviar un archivo de cada original a la dirección de correo electrónico de la Secretaría de la revista.
- Los trabajos pueden ser enviados en español o en inglés. La edición en formato impreso será siempre en español y la edición en formato electrónico puede incluir la versión en inglés.
- Los trabajos que no se ajusten a los criterios formales exigidos serán devueltos a su autor.

EVALUACIÓN Y ACEPTACIÓN DE ORIGINALES

- El Consejo de Redacción remitirá al autor acuse de recibo de su trabajo y le informará si el trabajo ha sido admitido al proceso de evaluación, o bien de su rechazo por no cumplir los criterios formales o editoriales.
- Todos los trabajos que cumplan con los criterios formales y editoriales serán sometidos a un proceso de informe de dos evaluadores anónimos. Si existe contradicción entre los informes de los dos evaluadores se procederá a una tercera evaluación que se considerará definitiva.
- Los artículos no deben incluir expresiones que hagan referencia a la identidad del autor o autores dentro del texto.
- Bajo ninguna circunstancia se revelará la identidad de los evaluadores y sus informes motivados serán decisivos a la hora de que el Consejo de Redacción entienda procedente o no la publicación de los trabajos.
- En un plazo máximo de seis meses (que podrá ampliarse en circunstancias extraordinarias) el Consejo de Redacción remitirá al autor los resultados de la evaluación y le informará, en su caso, del tiempo de que dispone para realizar las correcciones pertinentes.

CRITERIOS FORMALES DE PRESENTACIÓN DE ORIGINALES

- Los *artículos* deben ser presentados exclusivamente en formato electrónico en tamaño DIN A4, a un espacio y medio atendiendo a márgenes normalizados a

cada página. No se establece una extensión máxima, pero se sugiere que los trabajos considerados *Estudios* no excedan las 10.000 palabras, incluyendo notas y referencias bibliográficas y que los trabajos considerados *Experiencias* y *Casos* no excedan las 8.000 palabras, incluyendo notas y referencias bibliográficas.

- En archivo aparte el autor o autores deben enviar sus datos completos (institución, dirección postal, dirección electrónica y teléfono) y una breve reseña curricular (máximo 100 palabras) en la que incluyan grado académico, institución a la que pertenecen, líneas de investigación y principales publicaciones.
- Se recomienda que el título no exceda de 100 caracteres (incluyendo espacios).
- Cada trabajo debe acompañarse de un resumen de no más de 200 palabras, tanto en español como en inglés, y de una serie de palabras clave (máximo siete), preferentemente no incluidas en el título e igualmente en español y en inglés.
- Los gráficos, tablas, cuadros y otros dibujos e imágenes deberán integrarse dentro del texto e ir numeradas consecutivamente. Se hará referencia a las fuentes de procedencia y llevarán un título.
- Las citas en el texto deben hacerse entre paréntesis por autor y año, seguido de la página o páginas citadas (Lindblom, 1959: 79). Si se cita el trabajo de dos autores deberá incluirse el apellido de los dos unidos por la conjunción “y”. Si se trata de tres o más autores, es suficiente citar el primer autor seguido de *et al.* Si se citan varios trabajos de un autor o grupo de autores de un mismo año, debe añadirse las letras a, b, c..., después del año, tanto en el texto como en el apartado final de “Referencias bibliográficas y documentales”.
- Las notas deben ir a pie de página y el autor debe asegurarse de que estas notas y las citas incluidas en el texto coincidan con los datos aportados en la bibliografía.
- Las referencias a prensa deberán incluir, cuando sea pertinente, el título del artículo que se cita entre comillas y, en todo caso, el nombre de la publicación en letra cursiva, la fecha de publicación y, si procede, la página o páginas. Por ejemplo:

The Economist, 27 de enero de 2008.

Toner, R., “New Deal Debate for a New Era”, *The New York Times*, 1 de agosto de 2007.

- Al final del texto debe incluirse, por orden alfabético, el listado completo de la bibliografía empleada. Se deben seguir los siguientes formatos:

Libro:

Meny, I. y Thoenig, J.C. (1992), *Las políticas públicas*. Barcelona: Ariel.

Weber, M. (1964 [1922]), *Economía y sociedad*. México D.F.: Fondo de Cultura Económica.

Trabajo en un libro colectivo:

Sabatier, P.A. y Jenkins-Smith, H.C. (1999), “The Advocacy Coalition Framework: An Assessment”, en P.A. Sabatier, ed., *Theories of Policy Process*. Boulder: Westview Press.

Artículo:

Goodsell, Ch. (2006), “A New Vision for Public Administration”, *Public Administration Review*, 66(5): 623-635.

Página web:

Schmitter, P.C. (2004), “On Democracy in Europe and the Democratization of Europe”, *Documentos de Trabajo Política y Gestión 1*. Madrid: Universidad Carlos III de Madrid. En línea: <<http://www.uc3m.es/uc3m/dpto/CPS/schmitter.pdf>> (consulta: 20 diciembre 2008).

- Se recomienda evitar el uso de palabras en idioma distinto al español y de neologismos innecesarios. En caso de que su uso sea necesario por no existir traducción al español, se pondrá en letra cursiva.
- La primera vez que se utilicen siglas o acrónimos deberán escribirse entre paréntesis precedidos por el nombre completo.

DERECHOS Y OBLIGACIONES DE LOS AUTORES

- Los autores ceden al INAP los derechos correspondientes al trabajo publicado en la revista, incluidos, con carácter exclusivo e ilimitado, los derechos a reproducir y distribuir el texto en cualquier forma o soporte (impresión, medios electrónicos o cualquier otra), además de los de traducción a cualquier otra lengua, sin perjuicio de las normas particulares o locales aplicables.
- El autor deberá adjuntar al original del texto que presente una declaración formal en la que haga constar que el contenido esencial del mismo no ha sido publicado ni se va a publicar en ninguna otra obra o revista mientras esté en proceso de evaluación en la revista GAPP, comprometiéndose en todo caso a comunicar de inmediato a la Dirección de la revista GAPP cualquier proyecto de publicación de dicho texto.

SECRETARÍA DE LA REVISTA: “GESTIÓN Y ANÁLISIS DE POLÍTICAS PÚBLICAS”

Instituto Nacional de Administración Pública
Centro de Publicaciones
C/ Atocha, 106. 28012 MADRID
Teléfono: (91) 273 91 19. FAX: (91) 273 92 87
Correo electrónico: revistagapp@inap.es