

# LA EVALUACIÓN DE UN PROGRAMA DE FORMACIÓN PERMANENTE DEL PROFESORADO EN ESPAÑA: METODOLOGÍA Y DISEÑO

Estela D'Angelo, Gabriel Rusinek\*

**SÍNTESIS:** En este artículo se presenta el diseño metodológico de la evaluación externa de un programa de formación permanente del profesorado en España, organizado al amparo de un acuerdo de colaboración entre administraciones educativas del Estado español, en el que participaron 4.257 docentes de un total de 147 centros educativos. Después de explicitar una conceptualización de la evaluación de programas, se describe la negociación del diseño, la selección de muestras y los problemas de acceso, así como las técnicas de recogida y las estrategias de análisis de los datos cualitativos y cuantitativos. Se discuten luego los procedimientos utilizados para corroborar las interpretaciones del equipo evaluador y para llegar a emitir juicios de valor empíricamente fundados sobre el programa. Finalmente, se considera el rol del informe de evaluación en relación con la toma de decisiones organizativas y formativas por parte de organizadores y ponentes, y con la mejora de los procesos educativos por parte de todos los agentes implicados.

**Palabras clave:** evaluación de programas; formación permanente; profesorado; metodología.

## **A AVALIAÇÃO DE UM PROGRAMA DE FORMAÇÃO PERMANENTE DO PROFESSORADO NA ESPANHA: METODOLOGIA E PLANEJAMENTO**

**SÍNTESE:** Neste artigo, apresenta-se o planejamento metodológico da avaliação externa de um programa de formação permanente do professorado na Espanha, organizado ao amparo de um acordo de colaboração entre administrações educativas do Estado espanhol, no qual participaram 4.257 docentes de um total de 147 centros educativos. Depois de explicar o conceito da avaliação de programas, descreve-se a negociação do planejamento, a seleção de amostras e os problemas de acesso, assim como as técnicas de coleta e as estratégias de análise dos dados qualitativos e quantitativos. A seguir, discutem-se os procedimentos utilizados para corroborar as interpretações da equipe avaliadora e para chegar a emitir pareceres empíricamente fundados sobre o programa. Finalmente, considera-se o papel do relatório de avaliação em relação com a tomada de decisões organizativas e formativas por parte de organizadores e expo-

\* Profesores titulares en la Universidad Complutense de Madrid, España.

sitores e com a melhoria dos processos educativos por parte de todos os agentes implicados.

*Palavras-chave:* avaliação de programas; formação permanente; profesorado; metodologia.

**THE EVALUATION OF A PERMANENT TRAINING PROGRAM OF TEACHERS IN SPAIN: METHODOLOGY AND DESIGN**

*ABSTRACT:* This article presents the methodological design of the external evaluation of a permanent training program of teachers in Spain, organized under the protection of a collaboration agreement between Spanish educational administrations, organized under the protection of an agreement of collaboration between educational administrations of the Spanish State, in which took part 4,257 teacher of a total of 147 educational centers. After making explicit a conceptualization of the evaluation of programs, the negotiation of the design, the selection of samples and the problems of access are described, as well as the withdrawal technologies and the strategies of analysis of the qualitative and quantitative information. The procedures are then discussed, used to corroborate the interpretations of the assessor equipment and to manage value judgments founded empirically on the program. Finally, it is considered to be the role of the evaluation report in relation to the capture of organizational and formative decisions on the side of organizers and referees, and the improvement of the educational processes on the side of all the agents implied.

*Keywords:* program evaluation; permanent training; teachers; methodology.

## 1. UNA VISIÓN DE LA EVALUACIÓN DE PROGRAMAS EDUCATIVOS

La evaluación externa e independiente de programas educativos es un importante recurso de las sociedades democráticas para disponer de datos contrastados como garantía de imparcialidad en la toma de decisiones organizativas, curriculares y presupuestarias, así como para facilitar los procesos de mejora de los servicios.

En este artículo realizaremos un análisis metodológico de nuestra experiencia como evaluadores externos de un programa de formación permanente del profesorado, organizado en 2011 por una agencia del Ministerio de Educación español, en coordinación con las autoridades educativas de quince comunidades autónomas, y en el que participaron 4.257 profesores<sup>1</sup> de 147 centros<sup>2</sup> de educación primaria y secundaria (AA. VV., 2013). Nuestra intención

<sup>1</sup> Utilizamos la palabra «profesor» –en carácter de término no marcado por el género– para referirnos a profesionales de la educación, tanto de primaria y secundaria como de programas de cualificación profesional inicial y ciclos formativos de formación profesional.

<sup>2</sup> Teniendo en cuenta la comunidad de lectores iberoamericana, utilizamos el término «centros educativos» para referirnos a escuelas o colegios de educación primaria y a colegios o institutos de educación secundaria.

aquí no es plantear cómo han de hacerse las evaluaciones de programas, sino compartir nuestra propia experiencia y pensar metodológicamente con los lectores, como sugieren en sus cursos algunos evaluadores cuyos trabajos valoramos para el desarrollo de esta tarea (KUSHNER, 2002; STAKE, 2006). Para ello utilizamos en este artículo la primera persona del plural, resaltando nuestra presencia en las situaciones de recogida de datos, interpretación y corroboración. Entendemos, de acuerdo con Watzlawick y otros (1967), que la atribución de sentido y valor a los datos objetivos es inherente a la condición humana y que, aunque nuestro análisis de esos datos haya sido sistemáticamente imparcial, afirmar una neutralidad conceptual mediante el uso del reflexivo o de la tercera persona es, como sugiere Eisner (1998), solo una estrategia retórica.

Comenzamos explicitando nuestra visión de la evaluación de programas educativos, para luego describir el enfoque y el diseño que consensuamos con la agencia que encargó la evaluación externa. Luego describimos el proceso de recogida de datos, junto con los problemas de acceso suscitados por el tamaño de la muestra solicitada y su dispersión geográfica. A continuación, detallamos el análisis de los datos cualitativos provenientes de observaciones, entrevistas, grupos de discusión y revisión de materiales escritos, y explicamos cómo utilizamos las categorías emergentes para confeccionar unos cuestionarios que nos permitieron una cuantificación descriptiva. Por último, detallamos cómo estos datos nos permitieron realizar atribuciones de valor y mérito en relación con los criterios de evaluación requeridos por quienes encargaron la evaluación.

La evaluación está presente a diario en nuestras vidas, aunque de manera informal: todas las personas valoran la calidad de las comidas que comen, los espectáculos que ven o los coches que conducen. Sin embargo, parece ser que la evaluación formal de los programas no es tan frecuente en nuestras sociedades. La evaluación formal, externa e independiente, constituye un mecanismo crucial para mejorar productos y servicios. En relación con los servicios públicos, en particular, es una demanda ineludible para mejorar la vida de los ciudadanos y, por supuesto, para visibilizar si los impuestos que pagan son utilizados de manera responsable por sus gobiernos.

Conceptualizamos la evaluación formal como un procedimiento sistemático, metódico y neutral para obtener información sobre el valor y el mérito de algo que se evalúa. En términos generales, el valor se refiere a si lo evaluado da respuesta a las necesidades de sus usuarios, y el mérito se refiere a cómo lo hace, es decir, a los atributos del servicio que brinda. Sin embargo, como advierte Stake (2006), existen diferencias sustanciales entre las evaluaciones de objetos, las de personas y las de programas. Los objetos son comparables con otros de su categoría, según determinados aspectos

de esos objetos denominados «criterios», que pueden ser cuantificados en estándares. Por ejemplo, los coches pueden ser comparados en cuanto al criterio de resistencia a los choques frontales de acuerdo con los estándares de la agencia Euro NCAP. Las personas también son comparables, aunque como algunos criterios no pueden ser fácilmente cuantificados, en vez de estándares se utilizan «indicadores» de un grado de desarrollo de ese criterio. Por ejemplo, los profesores pueden ser comparados en cuanto al criterio de formación académica, y por ello en determinados procedimientos administrativos existen unos estándares para ese criterio.

Por el contrario, en la evaluación de programas la comparación resulta problemática, ya que a menudo estos presentan características únicas que difícilmente permiten parangonarlos con otros programas, anteriores o simultáneos, y menos con «estándares» que tengan validez intercultural. La complejidad determinada por la gran cantidad de agentes implicados, en general con objetivos diferentes, exige comprender las perspectivas de todos (por ejemplo, la administración educativa, los organizadores del programa, el personal que colabora, los ponentes, el profesorado participante en la formación y, en general, las comunidades educativas a las que estos sirven), a fin de poder realizar una síntesis que proporcione criterios válidos para cada programa en particular.

58

Nuestra aproximación a la evaluación externa de programas se basa en el enfoque respondiente de Stake (2006)<sup>3</sup>, cuya principal pretensión es obtener información útil no solo para que los organizadores e instituciones financiadoras de un programa estén en condiciones de tomar decisiones, sino también para que todos los agentes implicados en él puedan mejorarlo. Considerando que, además de la efectividad en la realización de lo planificado, el éxito o el fracaso de un programa depende de la experiencia que tengan los agentes implicados en él, el mencionado enfoque exige la utilización extensiva de técnicas cualitativas de recogida de datos que añadan la comprensión de esa experiencia a los datos obtenidos por técnicas cuantitativas.

Esa es también la visión personalizadora de Kushner (2002), para quien el valor no está ligado exclusivamente a un conjunto de indicadores definidos con sentido político sino también a cómo incide el propio programa en la vida de las personas que participan en él, como suministradores o usuarios, y cuyas ideas nos sugieren la obligación ética de analizar el valor de todo programa educativo en relación a cómo influye en las vidas del profesorado participante, del alumnado y de las comunidades educativas en su conjunto.

---

<sup>3</sup> La obra de Stake (2004), titulada originalmente en inglés *Standards-based and responsive evaluation*, fue traducida al español como *Evaluación comprensiva y evaluación basada en estándares* (STAKE, 2005).

Finalmente, nuestra aproximación también se nutre de las ideas sobre evaluación *democrática* de MACDONALD (1993), quien entiende la evaluación como un servicio a la comunidad, oponiéndose a modelos burocráticos o autocráticos.

## 2. LA NEGOCIACIÓN DEL DISEÑO, EL MUESTREO Y EL ACCESO

El programa a evaluar presentaba gran complejidad, no solo por tratarse de una propuesta de formación permanente del profesorado en el contexto de un Estado descentralizado, sino porque consistía en un modelo de formación en cascada, en el que la figura de un «docente-coordinador» por centro educativo viajaba mensualmente a Madrid para recibir formación y, a su regreso, debía transmitirla a sus colegas, fomentar la participación y la reflexión en el contexto de reuniones de ciclo o de departamento, y coordinar la realización práctica de las diferentes propuestas formativas.

Tanto los responsables de encargar la evaluación externa del programa como nosotros, en tanto equipo evaluador, compartíamos con diferentes expertos la idea de que la misión primordial de la evaluación es contribuir a la mejora de los programas. En lo que inicialmente no coincidíamos era en su propuesta de tomar como única referencia el modelo CIPP (contexto / *input* / proceso / producto) de Stufflebeam (2002), más orientado a la toma de decisiones a partir del análisis de las metas, la planificación, la realización y el impacto de un programa.

Tras revisar nuestra aproximación inicial para posibilitar un acuerdo, vimos que la propuesta del modelo CIPP, de desentrañar el entretejido y las interrelaciones de la totalidad de los componentes de un programa, nos permitía enlazar con nuestra visión sistémica de los grupos humanos en educación (STIERLIN, 1997), que había sido provechosa anteriormente en la evaluación de otro programa (D'ANGELO, 2002).

Un programa educativo, al igual que un centro educativo, constituye un sistema que se organiza en torno a elementos estructurales como interacciones, metas, sentimientos, tareas compartidas y cambios generados por la formación, que generan conductas grupales singulares con articulación de comportamientos individuales.

Así, finalmente, llegamos a consensuar un diseño que combinaba aspectos de ambos enfoques y que se apoyaría tanto en datos cualitativos como cuantitativos. De este modo, en la evaluación del contexto se buscaron datos globales para comprender los objetivos del programa; en la evaluación

del *input* (entrada) se identificaron datos que proporcionarían el fundamento para la elección del modo de utilización de los recursos disponibles para lograr esos objetivos; en la evaluación del proceso se analizaron datos sobre la marcha del programa, y en la evaluación del producto, el interés se centró en los datos sobre logros reales y su relación con dichos objetivos.

Desde ese marco referencial y organizacional, también acordamos que la evaluación fuera formativa más que sumativa (SCRIVEN, 1967), es decir, que persiguiera la mejora del programa en vez de la responsabilización por sus eventuales virtudes o defectos.

El siguiente paso en la negociación del diseño fue la definición de la muestra de centros. Para la recolección de datos cualitativos suele usarse un muestreo teórico (CORBIN y STRAUSS, 1990), en el sentido de que la selección responda a una representación con riqueza conceptual en cuanto a las variables, más que una representación estadística de la población analizada. Por ello, el muestreo inicial de los centros educativos tuvo en cuenta variables tales como:

- Que fueran de educación primaria o secundaria (en los de secundaria, que tuvieran o no ciclos de formación profesional o programas de cualificación profesional inicial).
- Que fueran públicos o privados.
- Que fueran urbanos, semirurales o rurales.
- Que estuvieran ubicados en capitales de provincia, ciudades grandes o ciudades pequeñas.
- Que fueran de diferente antigüedad.
- Que tuvieran diferente tamaño.
- Que su alumnado fuera homogéneo o heterogéneo (hijos de familias inmigrantes, alumnos con problemas de aprendizaje o conducta, o necesidades educativas especiales, etcétera).
- Que su profesorado hubiera participado previamente en acciones formativas similares o no.
- Que tuvieran mayor o menor apoyo del claustro a la participación en el programa.
- Que la participación estuviera coordinada por miembros del equipo directivo o por profesores designados por estos.

Sin embargo, quienes encargaron la evaluación insistieron en que se seleccionara un 30% de los centros participantes, añadiendo la representación regional a las variables anteriores por tratarse de una colaboración entre el Ministerio de Educación y distintas administraciones autonómicas con competencia en materia educativa, dando de ese modo visibilidad equitativa a la participación de las distintas administraciones, teniendo en cuenta los matices propios que cada una podría aportar a la concreción curricular.

Aceptamos el desafío por esas connotaciones, aun sabiendo que 51 centros superaban la cantidad necesaria, y efectivamente las categorías analíticas se saturaron (STRAUSS y CORBIN, 1998) mucho antes de realizar todas las visitas pactadas.

El aspecto altamente positivo de ese extenuante reto operativo fue que la recogida de datos cualitativos se convirtió en una enriquecedora experiencia que nos permitió aproximarnos personalmente a las realidades de centros situados en 34 ciudades distribuidas por toda la geografía española. Para los 17 restantes, el contacto se realizó por videoconferencia.

### 3. LA RECOGIDA Y EL ANÁLISIS DE LOS DATOS CUALITATIVOS Y CUANTITATIVOS

61

Nuestro diseño de evaluación multimetódico no podía consistir en administrar unos cuestionarios y luego ilustrarlos con datos de observaciones y entrevistas, ya que esa habitual secuencia metodológica suele incurrir en el error de cuantificar solo temas planteados por los organizadores o que están en la mente de los evaluadores, dejando fuera muchos de los problemas reales que preocupan a los participantes. Así pues, la secuencia diseñada arrancó con la recolección de datos cualitativos para que emergiera inductivamente una gran riqueza de categorías analíticas, y siguió con la recogida de datos cuantitativos a través de unos cuestionarios que usaron esas categorías emergentes.

Para obtener los datos cualitativos utilizamos técnicas como la observación, la entrevista individual semiestructurada y abierta (FONTANA y FREY, 1994), el grupo de discusión –entendido en el sentido inclusivo que establece Morgan (1996, 1997) de técnica de recolección de datos mediante la interacción grupal en un tema determinado por el investigador– y la revisión de materiales escritos. Las observaciones de reuniones organizativas y formativas a nivel general en Madrid fueron realizadas desde el comienzo al final del programa, mientras que las visitas a los centros con observaciones, entrevistas y grupos de discusión demandaron un trabajo muy intensivo que

incluyó viajar a 34 poblaciones distribuidas por toda España a lo largo de tres meses.

Con el fin de comprender el modelo organizacional del programa y su enfoque formativo, realizamos observaciones de todos los encuentros de coordinación entre el Ministerio de Educación y los representantes de las comunidades autónomas, y de todos los encuentros formativos mensuales en Madrid a los que asistían los coordinadores de centro.

En los 34 centros escolares visitados, por otro lado, realizamos observaciones orientadas a comprender sus características y el contexto en el que se desenvolvían los profesores participantes (por ejemplo, barrio circundante, contexto social, tipo de alumnado, clima escolar). En dos centros de primaria y ocho de secundaria también observamos clases en las que se estaban desarrollando unidades didácticas integradas que habían sido programadas como parte de las actividades del programa. En dos centros también observamos clases con metodologías consideradas innovadoras por los propios profesores que, si bien eran consecuencia de programas formativos anteriores, nos permitían comprender los diferentes puntos de partida de los centros participantes para evaluar la coherencia en su adscripción a los diferentes niveles formativos del programa.

62

Además de las observaciones, la técnica principal para obtener información de los agentes implicados en el programa fue la entrevista, tanto individual como grupal, en su formato de grupo de discusión (MORGAN, 1996, 1997). Las primeras entrevistas fueron semiestructuradas, siguiendo un guion de preguntas de interés. Al irse saturando las categorías de análisis, las restantes se realizaron como entrevistas abiertas, es decir, dejando de lado un guión rígido para relanzar el hilo de la conversación en un proceso continuo de búsqueda de los temas emergentes que preocupaban a los informantes.

El total de entrevistas individuales realizadas fue de 197, mientras que el total de grupos de discusión coordinados fue de 60, en los que participaron más de 300 profesores. En los 34 centros visitados entrevistamos de manera individual al profesorado que ejercía la coordinación de las actividades del programa en cada centro educativo, miembros de equipos directivos, profesorado participante en el programa, profesorado no participante, orientadores y otros miembros de equipos de orientación.

En todos ellos también coordinamos grupos de discusión de entre cinco y seis profesores, cuya composición variaba según la disponibilidad del profesorado, pero generalmente incluyendo al docente-coordinador, algún miembro del equipo directivo y algunos profesores participantes.

Aunque el programa no exigía al profesorado participante llevar a cabo la aplicación práctica en sus aulas de las programaciones realizadas durante el proceso de formación, teniendo en cuenta los distintos recorridos de experiencia en el ámbito de actuación del programa, suponíamos que esto sí estaría sucediendo en algunos centros con desarrollos metodológicos y curriculares más avanzados. Al comprobar esta suposición con nuestras observaciones en las visitas a cinco centros, solicitamos y obtuvimos permiso para realizar entrevistas grupales al alumnado que había participado en dichas aplicaciones prácticas. Con los 17 centros restantes de la muestra utilizamos la videoconferencia para realizar las entrevistas individuales a los coordinadores y mantener los grupos de discusión.

Además de entrevistas a organizadores y ponentes del programa, también entrevistamos a 13 responsables de la coordinación en las comunidades autónomas y a otros siete miembros de sus equipos de coordinación técnica, asesoramiento o apoyo. En dos ciudades tuvimos la suerte de que los responsables regionales propusieran la organización de grupos de discusión con los coordinadores de todos los centros educativos participantes en su comunidad autónoma. Coincidiendo con las visitas a los centros, pudimos entrevistar a tres inspectores que no participaban específicamente en el programa.

Finalmente, también procedimos a la revisión de materiales escritos. En primer lugar, revisamos los documentos oficiales del programa como punto de partida para contrastar las intenciones con lo sucedido en la práctica. En segundo lugar, revisamos los materiales formativos del programa a fin de contrastar las ideas propuestas con las distintas maneras en que el profesorado participante las comprendía o aplicaba. En tercer lugar, revisamos, a través del portal web del programa, los trabajos realizados por los equipos docentes de los centros visitados o entrevistados vía videoconferencia, a fin de comentarlos en las entrevistas y grupos de discusión y, en el caso de algunos centros, solicitar autorización para observar algunas aulas en las que las concreciones curriculares se desarrollaran de acuerdo con los planteamientos que asumía el programa. Por último, revisamos las aportaciones del profesorado en el foro de debate proporcionado por el portal web y las que los participantes consideraron conveniente compartir con el equipo de evaluación durante las visitas a los centros.

Las entrevistas individuales y los grupos de discusión fueron registrados con grabadoras digitales de audio y posteriormente transcritos. Esos datos de texto, al igual que las observaciones, la revisión de los materiales escritos y, posteriormente, las respuestas a preguntas abiertas de los cuestionarios a coordinadores de centro y demás profesorado, se fueron volcando al programa de análisis de datos cualitativos Atlas.ti (MUÑOZ JUSTICIA y

SAHAGÚN-PADILLA, 2011). Este programa fue de gran utilidad para manejar un enorme volumen de información y proceder a su interpretación mediante un proceso de codificación conjunta (MACQUEEN y OTROS, 1998; SALDAÑA, 2009), comparación constante entre códigos y establecimiento de categorías centrales (CORBIN y STRAUSS, 1990).

El análisis de los datos recogidos y, fundamentalmente, el hallazgo de categorías emergentes que excedían las previsiones originales sobre los temas objeto de la evaluación, permitieron elaborar, con posterioridad, unos cuestionarios. Los primeros ítems de los mismos surgieron en primer lugar de las cuestiones que interesaban a los organizadores y gestores institucionales para fundamentar tomas de decisión en relación con el mejor cumplimiento de los objetivos que se habían planteado. Sin embargo, la mayor riqueza de esos cuestionarios provino de la gran cantidad de aspectos que emergieron en las entrevistas individuales y grupos de discusión, ya que ampliaban notablemente los temas iniciales, dándonos la posibilidad de aportar a los organizadores una cuantificación de las cuestiones que preocupaban al profesorado, es decir, de su representatividad con respecto a la población participante en el programa.

64

Diseñamos dos cuestionarios destinados al profesorado para su administración a través del portal web del programa y su respuesta anónima. Los dos cuestionarios incluyeron tanto preguntas de opción múltiple como preguntas de respuesta abierta, y estuvieron dirigidos respectivamente a los docentes-coordinadores de centro y al resto del profesorado participante.

El primero consistió en 60 preguntas y fue respondido por 144 coordinadores. Dado que participaron hasta el final del programa 147 centros, su índice de respuesta del 98% permitió adscribir mucha credibilidad a sus resultados. Al entender que la habitual sobrecarga de obligaciones burocráticas de los docentes podía reducir el índice de respuesta, se confeccionó un segundo cuestionario más reducido, con 38 preguntas, que fue finalmente respondido por 2.791 profesores. Teniendo en cuenta que el total de participantes en 144 centros –según fue informado por los docentes-coordinadores de centro– ascendía a 4.257 profesores, el índice de respuesta del 65% nos exigió precaución a la hora de analizar algunos de los resultados, en el sentido de contrastarlos con otros datos para su validación.

Para todo el proceso de interpretación de los datos obtenidos utilizamos algún tipo de corroboración, como es habitual en cualquier diseño exigente, ya que en algunos casos aparecían informaciones contradictorias. Por cautela metodológica, antes de dar por apropiadas determinadas interpretaciones, buscamos sistemáticamente casos negativos o datos que pudieran contradecirlas. Por dar un ejemplo, el 57% de los respondientes a los

cuestionarios para profesorado afirmó haber «leído» de tres a seis de los seis módulos en los que se organizaban los materiales formativos. Sin embargo, al triangular esa información con la registrada en las entrevistas y grupos de discusión, observamos que, de alrededor de 400 profesores que no eran docentes-coordinadores de centro, fueron escasos quienes afirmaron haberlos leído o estaban en condiciones de demostrarlo comentando sus contenidos. La contradicción que ejemplificamos implica que o bien los entrevistados y participantes en grupos de discusión no eran del todo sinceros y contestaban lo que suponían que se esperaba de ellos, o bien todos ellos podían asimilarse a una muestra casual correspondiente al 23% de los respondientes del cuestionario que afirmaron no haber leído ningún módulo. En este ejemplo también consideramos que los 1.466 profesores que no respondieron los cuestionarios podían haberlos leído o no, pero que la propia falta de interés por responder el cuestionario podía sugerir que tampoco lo habían hecho y que, en consecuencia, el porcentaje real de no-lectores podría haber sido mayor.

Estas disquisiciones metodológicas para corroborar las interpretaciones se aplicaron sistemáticamente junto con procedimientos de triangulación (DENZIN, 1978) de observadores (entre nosotros como evaluadores externos, y a veces con la ayuda de agentes externos al contexto observado, como directores de centros de profesores, inspectores o funcionarios de consejerías de Educación), de técnicas de recogida de datos (entrevistas, observaciones, revisión de materiales escritos, cuestionarios), de teorías (al contrastar técnicas cualitativas y cuantitativas) y de fuentes de datos (al recoger datos tanto de los organizadores, colaboradores y ponentes de la acción formativa como del profesorado participante y de algunos alumnos).

Por una consideración ética de respeto a la privacidad de los participantes (BERA, 2011), se garantizó y se mantuvo el anonimato a lo largo de todo el proceso de evaluación. Al librarlos de ciertos tipos de presiones, esa precaución ética constituía a la vez un apoyo metodológico, ya que ayudaba a obtener una mayor sinceridad en sus informaciones. El anonimato se aplicó a los datos de identificación individual de las personas, de los centros escolares en que estas trabajaban o estudiaban, e incluso de las comunidades o ciudades autónomas donde los centros se situaban. El compromiso de anonimato fue informado a los coordinadores regionales con anterioridad a la negociación del acceso a los centros de su jurisdicción, a los responsables de la coordinación en cada centro durante esas negociaciones y a cada uno de los participantes en entrevistas individuales, grupos de discusión u observaciones, inmediatamente antes de la recogida de datos. Para la transcripción de los datos y su utilización en el informe final, se utilizaron códigos con el fin de ocultar esas identidades.

#### 4. EL INFORME DE EVALUACIÓN: DATOS, JUICIOS DE VALOR Y PROPUESTAS DE MEJORA

El producto final de la evaluación de un programa se plasma en un informe de evaluación. Para la evaluación del programa en cuestión, elaboramos en 2011 un informe (D'ANGELO y RUSINEK, 2013) de cien páginas de extensión, que además de emitir juicios sobre su valor y sus méritos, incorporó una gran cantidad de temas emergentes, meticulosamente documentados con citas de datos textuales y análisis numéricos descriptivos, en el que el profesorado participante pudo ver plasmados sus puntos de vista y en el que los organizadores pudieron encontrar aspectos clave para la mejora de ulteriores ediciones del programa. Así, incorporamos esos temas en la evaluación:

- *Del contexto.* Se analizó el modelo curricular objeto de la acción formativa en su contexto europeo y nacional, las acciones formativas previas que constituyeron la base sobre la que se apoyó el programa, las características de los centros participantes y las características del profesorado.
- *Del input.* Se analizaron las características de la acción formativa, sus materiales y el portal web que servía de repositorio y plataforma comunicativa.
- *Del proceso.* Se analizó cómo accedieron los centros a la acción formativa y cómo se les adscribió a los niveles de que constaba, los encuentros formativos mensuales para los coordinadores de centro, cómo se organizaba el trabajo posteriormente en los propios centros educativos y cómo apoyaban ese trabajo los equipos de las comunidades autónomas.
- *Del producto.* Se analizó cómo el programa constituyó una oportunidad para el debate pedagógico y el desarrollo de un lenguaje curricular compartido, el cambio de mentalidad del profesorado, los cambios en las programaciones de aula y en la programación, el desarrollo práctico de unidades didácticas integradas en las aulas y la estructuración de la innovación y los planes de mejora en los centros.

Finalmente, y después de reflejar los niveles de satisfacción global y la percepción sobre la necesidad de mejorar determinados aspectos que habían expresado en los cuestionarios los coordinadores-docentes de centro educativo y el profesorado participante, procedimos a emitir los juicios de valor, como equipo de evaluación externa, acerca del programa.

Estos juicios se realizaron en relación con unos criterios de evaluación previamente consensuados con quienes habían encargado la evaluación y que fueron explicitados en la propuesta que redactamos y que aprobó el equipo gestor del programa, integrado por representantes ministeriales y de las comunidades autónomas. Dichos criterios se referían: 1) a la pertinencia del programa como acción formativa; 2) a la coherencia de los distintos elementos que lo conformaban; 3) al nivel de compromiso con la mejora del currículo de los agentes implicados, expresado como participación; 4) al impacto de la acción formativa, y 5) a su viabilidad en tanto continuidad en el tiempo de sus efectos, y a su eficacia como programa formativo en el campo de la educación.

Evaluar la pertinencia del programa implicó valorar, en sus aspectos legales, pedagógicos y organizacionales, su contribución a un problema de formación del profesorado para llevar a cabo determinadas exigencias del currículo obligatorio, que compartían todas las administraciones educativas del Estado español.

Por su parte, evaluar la coherencia del programa exigió analizar en primer término la relación entre los tres grupos de elementos que lo conformaban, así como entre estos y su finalidad. En este sentido, se valoró la coherencia entre la administración central y las administraciones educativas de las comunidades autónomas; entre estas y los centros participantes de su jurisdicción, y entre la administración central y los centros. En segundo término, se valoró la coherencia del esquema que constituía el eje central del programa, como era la formación en cascada.

Evaluar la participación requirió valorar el nivel de compromiso con la mejora del currículo que llegaron a alcanzar los distintos agentes responsables del programa, así como el nivel de implicación logrado en los propios centros educativos. Aunque la formación se realizaba en los centros educativos y la participación en el programa debía ser autorizada por el conjunto de profesores del centro, la participación individual de los profesores no era obligatoria. Por ello, además de las cifras descriptivas de porcentaje de profesores participantes sobre el total de profesores que trabajaban en esos centros, valoramos otros aspectos que matizaban las cifras. Uno de esos aspectos fue la distribución del índice de participación, ya que aunque el porcentaje global era del 65%, mientras en algunos centros participaba todo o casi todo el profesorado, en otros lo hacía un grupo pequeño, y estas diferencias tenían consecuencias para la valoración no solo de la implicación sino también del impacto del programa, al relativizar el alcance de sus productos. Otro aspecto que tenía consecuencias sobre la implicación es cómo se decidió esa participación: quién la propuso y quién la apoyó, y cuán obligada fue para los centros o para los profesores.

La evaluación del impacto comenzó identificando los efectos generados por los procesos puestos en marcha. Aunque los productos documentados constituían un efecto importante, el impacto no equivalía a la suma de esos productos, dado que su distribución en los centros era dispar: diferentes centros lograban diferentes cantidades, tipos y calidades de productos, dependiendo de la experiencia previa del profesorado en el tema sobre el que trataba la acción formativa, de los niveles de compromiso de los claustros con la innovación, de las habilidades de dinamización de los coordinadores y de las estrategias organizativas implementadas por estos junto con los equipos directivos.

Este conjunto de variables complicaba la ponderación de la representatividad y extensión de los productos, necesaria para emitir un juicio de valor sobre el impacto. Como ejemplo para ilustrar esta relación, analizaremos aquí dos casos. En un centro de secundaria, las reflexiones sobre la planificación y desarrollo práctico de una unidad didáctica integrada en las aulas realizadas por los profesores participantes –en tanto usuarios– y de sus alumnos –en tanto beneficiarios finales– corroboraban que ese producto constituía una demostración empírica de los principios de programa. Sin embargo, su contribución al impacto global del programa quedaba disminuida por el hecho de que en ese centro solo participaba el 14% del profesorado. El caso contrario es el del 26% de centros en los que la comisión de coordinación pedagógica vertebraba la participación, al reunirse en ella el docente-coordinador del programa con los jefes de departamento o coordinadores de ciclo, para que ellos realizaran luego las actividades con los profesores de sus departamentos o ciclos. Por mínimos que fueran los productos conseguidos en esos centros, consideramos que su contribución al impacto global del programa estaba potenciada por la institucionalización de la innovación curricular en sus estructuras organizativas, uno de los objetivos fundamentales del programa.

La evaluación de la viabilidad tomaba en cuenta la continuidad del marco de colaboración creado, así como de los cambios incorporados en los centros educativos. En nuestro análisis, consideramos como marco de colaboración no solamente la prestada entre la administración central y las comunidades autónomas en el contexto de un programa de cooperación territorial, sino también a todas las redes de personas e instituciones, preexistentes o generadas, que se observaron interactuando en el programa, incluyendo las redes de centros. A su vez, aunque desde el punto de vista formal la continuidad en los centros educativos se refería a la institucionalización de las innovaciones curriculares en los documentos del centro, se consideraron de importancia, desde el punto de vista operativo, los liderazgos consensuados que apostaban por la innovación y por la generación y el mantenimiento de redes de colaboración dentro de los propios centros. También se consideró la viabilidad en relación con el altísimo interés de coordinadores de centro

y profesorado participante por continuar la acción formativa, aunque por las observaciones registradas se interpretó que la continuidad de esos efectos del programa se apoyaba, por un lado, en el aprovechamiento que el mismo hacía de varias redes que antes de su inicio habían comenzado a abordar en los centros educativos la innovación curricular específica, y, por otro, en la generación y potenciación de redes nuevas.

La eficacia de un programa suele definirse como el nivel de consecución de sus objetivos, relacionando las condiciones de entrada con el contexto en el que se ha desarrollado. Para emitir un juicio de valor acerca de su eficacia, se hizo una distinción entre los objetivos iniciales, expuestos en el documento oficial del programa, y su contextualización operativa. Así, se consideró que la eficacia del programa también se relacionaba con los resultados de los objetivos situados, derivados de esa contextualización que producía la retroalimentación de los participantes durante la realización de las propuestas.

Finalmente, además de juicios de valor, para todos los criterios se incorporaron propuestas de mejora que provenían de la sistematización de las opiniones tanto del profesorado participante —expresadas en entrevistas, grupos de discusión y respuestas a las preguntas abiertas de los cuestionarios— como de los demás agentes implicados que fueron entrevistados.

La evaluación cuya metodología y diseño hemos descrito en este artículo ha servido para que los organizadores y ponentes del programa pudieran tomar decisiones organizativas y formativas. Según la retroalimentación recibida después de la distribución del informe, la inclusión de datos emergentes para fundamentar los juicios de valor ha contribuido también a que los demás agentes implicados —fundamentalmente los profesores participantes, como usuarios del programa— pudieran sentirse parte de la mejora de los procesos en los que participaban.

Como conclusión, sostenemos que la difusión pública del informe, como de los informes de evaluación de todos los programas financiados con dinero público, aporta a los ciudadanos el acceso a una información confrontada e imparcial que permite visibilizar la forma en que sus gobiernos gestionan los recursos. Y esta visibilidad, creemos, es crucial para una transparencia democrática.

## BIBLIOGRAFÍA

- AA. VV. (2013). «Guía para la formación en centros sobre las competencias básicas». Madrid: Ministerio de Educación, Cultura y Deporte. Disponible en <https://sede.educacion.gob.es/publventa/detalle.action?cod=16109>.
- BERA (BRITISH EDUCATIONAL RESEARCH ASSOCIATION) (2011). *Ethical guidelines for educational research*. Londres: BERA. Disponible en: [www.bera.ac.uk/system/files/3/BERA-Ethical-Guidelines-2011.pdf](http://www.bera.ac.uk/system/files/3/BERA-Ethical-Guidelines-2011.pdf).
- CORBIN, J. y STRAUSS, A. (1990). «Grounded theory research: Procedures, canons, and evaluative criteria». *Qualitative Sociology*, 13(1), pp. 3-21.
- D'ANGELO, E. (2002). *Evaluación externa de los Programas de Atención al Desarrollo para Niños y Niñas de Educación Infantil* (informe inédito). San Salvador: UNICEF - Ministerio de Educación de El Salvador - USAID.
- RUSINEK, G. (2013). «Proyecto COMBAS: Informe de evaluación externa». En AAVV (eds.), *Guía para la formación en centros sobre las competencias básicas*. Madrid: Ministerio de Educación, Cultura y Deporte. Disponible en <https://sede.educacion.gob.es/publventa/detalle.action?cod=16109>.
- DENZIN, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2.<sup>a</sup> ed.). Nueva York: McGraw-Hill.
- EISNER, E. W. (1998). *El ojo ilustrado: Indagación cualitativa y mejora de la práctica educativa*. Barcelona: Paidós.
- FONTANA, A. y FREY, J. H. (1994). «Interviewing: The art of science». En N. K. Denzin y Y. S. Lincoln (eds.), *Handbook of qualitative research*. Thousand Oaks (California): Sage, pp. 361-376.
- KUSHNER, S. (2002). *Personalizar la evaluación*. Madrid: Morata - Paideia.
- MACDONALD, B. (1993). «A political classification of evaluation studies in education». En M. Hammersley (ed.), *Social research: Philosophy, politics and practice*. Londres: Sage, pp. 105-123.
- MACQUEEN y OTROS (1998). «Codebook development for team-based qualitative analysis». *Cultural Anthropology Methods*, 10(2), pp. 31-36.
- MORGAN, D. L. (1996). «Focus groups». *Annual Review of Sociology*, 22, pp. 129-152.
- (1997). *Focus groups as qualitative research* (2.<sup>a</sup> ed.). Thousand Oaks (California): Sage.
- MUÑOZ JUSTICIA, J. y SAHAGÚN-PADILLA, M. Á. (2011). «Análisis cualitativo asistido por ordenador con ATLAS.ti», en C. Izquierdo y A. Perinat (eds.), *Investigar en psicología de la educación. Nuevas perspectivas conceptuales y metodológicas*. Barcelona: Amentia, pp. 299-363.
- SALDAÑA, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks (California): Sage.
- SCRIVEN, M. (1967). «The methodology of evaluation», en R. W. Tyler, R. M. Gagné y M. Scriven (eds.), *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally, pp. 39-83.
- STAKE, R. (2004). *Standards-based and responsive evaluation*. Thousand Oaks (California): Sage.

- (2006). *Evaluación comprensiva y evaluación basada en estándares*. Barcelona: Graó.
- STIERLIN, H. (1997). *El individuo en el sistema: Psicoterapia en una sociedad cambiante*. Barcelona: Herder.
- STRAUSS, A. L. y CORBIN, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2.<sup>a</sup> ed.). Thousand Oaks (California): Sage.
- STUFFLEBEAM, D. L. (2002). «The CIPP model for evaluation», en D. L. Stufflebeam, G. F. Madaus y T. Kellaghan (eds.), *Evaluation models*. Dordrecht (Netherlands): Springer, pp. 279-317.
- WATZLAWICK, P. y OTROS (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes*. Nueva York: Norton.