

EVALUACION DE POLITICAS

Marco conceptual y organización institucional

Por XAVIER BALLART

SUMARIO

I. INTRODUCCIÓN.—II. ¿QUÉ ENTENDEMOS POR EVALUACIÓN?: 1. *Concepto de evaluación*. 2. *Ambito material de la evaluación*. 3. *Evaluación y juicio valorativo*.—III. ¿CÓMO HACERLO? LA CUESTIÓN DE LOS MÉTODOS: 1. *Primer paso: Crear datos sobre eficacia*: a) Tipos de medidas; b) Selección de variables y fuentes de datos. 2. *Segundo paso: la evaluación de las políticas y programas*: a) La evaluación de resultados; b) La evaluación de proceso.—IV. ORGANIZACIÓN INSTITUCIONAL DE LA EVALUACIÓN: 1. *Estados Unidos: el caso del mercado más abierto y plural*. 2. *Reino Unido: de la preocupación por el impacto final a las iniciativas centradas en la productividad*. 3. *Alemania: institucionalización de la evaluación en la administración central*. 4. *Suecia: fragmentación y descentralización de la evaluación*. 5. *Holanda: la reconsideración de las políticas*. 6. *Francia: creación de un dispositivo central de evaluación dependiente del primer ministro*.—V. IMPLICACIONES DE LA EXPERIENCIA COMPARADA EN LA DEFINICIÓN DE UNA POLÍTICA DE EVALUACIÓN PROPIA.—REFERENCIAS BIBLIOGRÁFICAS.

I. INTRODUCCION

Este trabajo presenta una descripción detallada de la organización y uso de la evaluación de la actividad de los poderes públicos en Estados Unidos y en cinco países europeos (Reino Unido, Alemania, Suecia, Holanda y Francia). Desde una perspectiva comparada se presenta información sobre el desarrollo-institucionalización de la evaluación en cada país y sobre la utilización de sus resultados en el proceso político-administrativo.

Si bien las seis democracias consideradas siguen una misma dirección, creemos que de la observación de la experiencia comparada se desprende claramente la idea de que cada país ha de definir su política de evaluación

en función de sus propios condicionamientos: situación fiscal y económica, orientación política del gobierno, características jurídico-políticas de las instituciones, tradición científica y capacidad de oferta por parte de la comunidad universitaria y profesional. Por esta razón, al final del trabajo se incluyen una serie de consideraciones sobre el caso español y sobre posibles líneas de desarrollo y organización de la evaluación en nuestro país.

El análisis de los seis casos individualmente considerados requería una clarificación de conceptos clave. El trabajo se inicia, pues, planteando algunas cuestiones básicas sobre el concepto y los métodos de la evaluación, cuestiones que se han tratado con mayor detalle en otro lugar (Ballart, 1992).

II. ¿QUE ENTENDEMOS POR EVALUACION?

1. *Concepto de evaluación*

El desarrollo de la evaluación, tanto por lo que se refiere a sus métodos como por lo que se refiere al tipo de cuestiones a las que se dirige, ha resultado en una clara expansión de su campo de aplicación.

La primera y principal idea en relación con la evaluación la identifica con un medio para determinar los resultados y efectos de las políticas y programas a través de rigurosos instrumentos metodológicos (preferiblemente, diseños de tipo experimental). El pensamiento más reciente, sin embargo, sugiere que la evaluación puede abarcar todas las esferas del ciclo vital de una política o programa, desde su concepción hasta su impacto pasando por su ejecución. Una definición bastante aceptada de evaluación es, pues, aquella que entiende que la evaluación consiste en *la aplicación de métodos de investigación sistemáticos con el objeto de examinar el diseño, la implementación y la utilidad de las políticas y programas públicos*.

Generalmente, la evaluación tendrá un carácter retrospectivo, o como mínimo, concomitante. A diferencia de otras formas de análisis prospectivo que pretenden informar decisiones que todavía no se han tomado, la evaluación retrospectiva busca producir información sobre lo que ha ocurrido o está ocurriendo.

La adopción de una perspectiva cíclica en relación con las políticas o programas de la Administración supone que la evaluación puede examinar como mínimo tres fases diferentes que se corresponden con las de creación o formulación, implementación, y producción de resultados o impactos de una política o programa. En cada una de estas fases se pueden plantear cuestiones, que si bien variarán en función de las necesidades de información

y de la audiencia (los gestores plantearán cuestiones que serán distintas de las de los políticos, legisladores o hasta medios de comunicación y usuarios), serán todas ellas de carácter retrospectivo.

Así, los estudios que se centran en la fase de formulación y diseño de las políticas persiguen producir información sobre:

1) El problema del que se ocupa la política: ello supone revisar o elaborar, en su caso, los estudios de necesidades correspondientes.

2) El proceso de formulación y diseño de la política: esto supone la investigación de los procesos políticos y sociales que conducen a la formulación de nuevas intervenciones.

3) La racionalidad intrínseca de las mismas intervenciones: ello supone preguntarse hasta qué punto la política, tal y como ha sido llevada a término, sirve a sus objetivos.

Los estudios de implementación de las políticas persiguen producir información sobre:

1) Las características del grupo de población servido, esto es, el nivel de cobertura conseguido, los posibles sesgos en su composición, o las razones por las que determinados grupos de población estiman que «no vale la pena» seguir el programa y optan por abandonar.

2) Los problemas generados por la propia organización administrativa, en términos de información y comunicación, acceso, tiempo, desplazamientos o inadecuación de los procedimientos administrativos establecidos.

Por último, los estudios de eficacia o impacto persiguen producir información sobre los efectos del programa. Para ello presuponen que existen un conjunto de objetivos operacionales e intentan determinar el éxito del programa en función de si se ha producido algún cambio en la dirección deseada hacia la consecución de estos objetivos. La función de los evaluadores consiste en este caso en intentar aislar el impacto de la intervención pública de otros factores que pudieran explicar por sí solos los efectos observados. De forma general, se puede afirmar que para ello será preciso recurrir a técnicas de carácter experimental o estadísticas (1).

2. *Ambito material de la evaluación*

La definición propuesta tiene igualmente un carácter expansivo en la medida que incluye la evaluación retrospectiva tanto de políticas como de programas y servicios. Un programa es un conjunto de recursos y actividades

(1) Esta concepción y tipología de los estudios de evaluación se encuentra en toda la literatura, si bien originalmente se debe a Rossi y Freeman (1985: 39-46).

organizados para obtener ciertos objetivos. Son programas los planes de expansión de la educación preescolar, de ayuda a los alquileres, de reforma integral de barrios degradados, o de formación e inserción de jóvenes limpia-cristales, por poner algunos ejemplos.

Muchas actuaciones públicas no adoptan la forma de un programa en el sentido convencional del término. Se trata de iniciativas políticas que no implican necesariamente la creación de nueva burocracia o la distribución de servicios específicos. Cambios en las leyes impositivas, una nueva regulación de la caza, la prohibición de circular sin el cinturón de seguridad..., son ejemplos de políticas públicas que la Administración podría llevar a término cambiando simplemente las órdenes o los procedimientos ordinarios. También estas políticas tienen, sin embargo, consecuencias que políticos y directivos públicos querrían conocer. Quizá cabría, pues, afirmar que la evaluación de las políticas tiene de forma general mayor relevancia a nivel de instancias de decisión política que no de gestión operacional e intenta responder a cuestiones relacionadas con la generalización de una determinada aproximación a un problema.

A todo ello aún cabría añadir los servicios de transporte urbano o interurbano, de policía, bomberos, ambulancias, recogida de basuras, limpieza viaria, bibliotecas, museos..., actuaciones públicas que no se articulan formalmente como un conjunto de programas dada la estructura administrativa existente y la forma en que se toman las decisiones que conciernen al presupuesto. La evaluación de este tipo de servicios intentaría identificar y especificar objetivos y establecer medidas de los servicios prestados, con el propósito de construir las bases de datos necesarias para ir ajustando mejor los mismos objetivos operacionales, e ir obteniendo resultados respecto a aquellas dimensiones que se consideren críticas, para una misma Administración o para distintas Administraciones que sean susceptibles de comparación.

3. *Evaluación y juicio valorativo*

La definición de evaluación propuesta ha sido considerada excesivamente tecnocrática por una parte de la comunidad científica, que ha propuesto definiciones alternativas basándose en el principio de que *evaluar una política consiste en valorar-emitir un juicio de valor sobre ésta*. De acuerdo con este concepto, habría que distinguir entre los expertos encargados de hacer el análisis —técnicos— y aquellos otros legitimados para emitir un juicio valorativo —las autoridades evaluadoras—, esto es, Parlamento, Gobierno, Tribunal de Cuentas, Pleno municipal... u otros órganos que se puedan constituir. Desde esta perspectiva se considera que sólo estas instituciones tienen la

legitimidad democrática para hacer juicios sobre el mérito u oportunidad de las políticas públicas (Guba y Lincoln, 1987: 210-215; Monnier, 1991-1992: 148-142; Viveret, 1989: 25).

Si bien el primer concepto de evaluación indica el requisito de una aproximación metodológica rigurosa, este segundo concepto pone de relieve el inevitable *input* que suponen los factores políticos en el proceso de evaluación. Ciertamente, la evaluación de la eficacia de las políticas y programas públicos busca establecer relaciones de causalidad entre medios y efectos y comparar estos últimos a los objetivos iniciales de la política. Este esfuerzo precisa, sin embargo, de un conocimiento profundo de la política. *La evaluación tiene que situarse dentro de la política*, sobre todo si adoptamos una visión dinámica que tenga en cuenta, primero, la interacción continua entre medios, efectos y objetivos, y segundo, la importancia de los juicios de los principales actores, incluyendo entre éstos hasta los mismos beneficiarios de la política. Su implicación, como la de todos aquellos que tengan un interés en la política, es importante, como condición para un cierto pluralismo, y también para la integración de todas las informaciones, interpretaciones y juicios pronunciados en el marco del proceso evaluador.

III. ¿COMO HACERLO? LA CUESTION DE LOS METODOS

Si bien la discusión en profundidad de los métodos de evaluación supera claramente el alcance de este trabajo, se incluyen a continuación algunas consideraciones sobre: primero, la necesidad de concentrar esfuerzos en la medida de la eficacia y en la construcción de sistemas de indicadores internos, simples y permanentes como primer paso para poder llevar a cabo evaluaciones en el futuro, y segundo, la posible utilización de los datos obtenidos para la evaluación de las políticas y programas públicos y, por tanto, para aislar el impacto de la intervención pública de otros factores causales. De acuerdo con el sentido amplio de evaluación adoptado, se hace igualmente referencia a aquellos métodos que se centran en el análisis de los procesos más que en el análisis de los resultados finales.

1. *Primer paso: crear datos sobre eficacia*

Si bien, como se ha indicado anteriormente, la discusión sobre la evaluación de la actividad de la Administración pública ha ido ampliando su objeto, se puede afirmar que es una discusión que ha girado en torno a la idea de resultados, efectos o impactos.

a) *Tipos de medidas*

Generalmente, se entiende por medidas de eficacia aquellas categorías de indicadores que aportan información sobre el grado en que los objetivos de una política o programa se han conseguido. A nuestro entender, sin embargo, las medidas de eficacia también deberían aportar información sobre los siguientes aspectos:

1) El grado en que se producen efectos no esperados en la población a la que van dirigidos los servicios, ya sean estos efectos positivos o negativos (Shortell y Richardson, 1978).

2) El grado en que la cantidad de servicios producidos se adecúa a las necesidades y expectativas de la población destinataria (Scriven, 1973).

3) La «forma» en que se produce la respuesta por parte de los servicios públicos a las demandas de servicio, lo que nos acerca a la idea de servicios implícitos o de servicio como «paquete» de servicios.

4) El grado de satisfacción de los ciudadanos con los servicios prestados, independientemente de que las percepciones subjetivas sean dispares de las medidas objetivas de la Administración (Fitzsimmons y Sullivan, 1982; 16-28).

Desde esta perspectiva, que combina indicadores objetivos y subjetivos, la evaluación conecta con la preocupación por determinar la medida en que las Administraciones públicas responden a las demandas de los ciudadanos y adopta, por tanto, la óptica del cliente-destinatario de los servicios y de la utilidad concreta que obtienen éstos de la intervención pública.

En sustitución de indicadores de eficacia a menudo se utilizan meras medidas de actividad o de carga de trabajo, dado que son más fáciles de construir a partir de contabilizar los recursos utilizados y/o los *output* o productos que resultan de la transformación de aquéllos en servicios específicos. Si bien el cómputo de las horas de trabajo invertidas, de los clientes atendidos, de las visitas, entrevistas o salidas realizadas, por poner algunos ejemplos, pueden ser de utilidad para propósitos operacionales o para justificar gastos, no nos parece que este tipo de medidas aporten excesiva información sobre los servicios en términos de calidad.

Como complemento de los indicadores de eficacia sí tiene mayor interés aportar información en términos de eficiencia, esto es, sobre la relación entre la cantidad de servicios producidos y la cantidad de *input* o recursos utilizados en su elaboración y distribución, puesto que un servicio eficaz se puede producir de forma intolerablemente ineficiente. Ahora bien, concentrar todos los esfuerzos en el análisis de la eficiencia y, por tanto, en la determinación del coste unitario de los servicios públicos no tiene demasiado sentido en un

contexto en que la mayor parte de los recursos tienden a ser bastante rígidos y en el que esta información podría, en último término, interesar únicamente a aquellos que precisan argumentos para denunciar la situación de ineficacia en la que se presupone operan las Administraciones públicas.

b) *Selección de variables y fuentes de datos*

La evaluación debe, pues, servir esencialmente para mejorar la calidad de los servicios finales, en términos del binomio eficacia-eficiencia, y para introducir elementos correctores en relación con su gestión y hasta planificación.

El primer paso en este sentido pasa por producir datos sobre la eficacia de los servicios, internamente y de forma continuada, de forma que sea posible elaborar periódicamente informes exhaustivos de interés para administradores-gestores, o más sintéticos de interés para los niveles políticos o hasta la prensa y la población en general. Para que estos datos tengan alguna posibilidad de ser utilizados es importante: primero, que los sistemas de mediación sean asumidos por sus propios usuarios (Patton, 1986), y segundo, que sean relativamente simples, ya que los análisis excesivamente sofisticados no se pueden realizar de forma habitual y tienden a producir información poco clara para los decisores o para la población.

Para la selección de los indicadores y de las fuentes de datos es preciso elaborar antes un mapa completo de los objetivos que definen eficacia en relación con un servicios específico. Para ello será conveniente recurrir a una pluralidad de instrumentos, entre los que cabe incluir:

1) Las declaraciones oficiales de objetivos (que se encuentran en los documentos de planificación, presupuestación o hasta memorias de actividades de la misma Administración).

2) Las discusiones generadas a través de las técnicas adecuadas entre los profesionales responsables o entre las organizaciones corporativas y asociaciones ciudadanas relacionadas con los servicios estudiados.

3) Los análisis de las quejas de los ciudadanos y de las preocupaciones expresadas por éstos individualmente en los medios de comunicación.

4) La literatura científica especializada.

A su vez, los indicadores y las fuentes de datos deberán reunir determinadas características de validez aparente y fiabilidad, además de cubrir la totalidad de los objetivos relevantes. Desde una perspectiva operativa, es igualmente importante que aquellas variables que se decida controlar dependan realmente de la actividad de la Administración, puesto que en el caso contrario poco se podrá hacer para incidir sobre su cambio en la dirección deseada.

Finalmente, los costes de recoger la información necesaria deberán ser razonables, ya que en caso contrario dejará de tener sentido la evaluación.

Esta última cuestión nos llevará a plantear la oportunidad de hacer una utilización intensiva de los datos que ya recoge de forma ordinaria y continuada la Administración. A menudo serán, sin embargo, necesarios otros procedimientos de recogida de información, como, por ejemplo, los estudios de opinión entre los usuarios del servicio o entre la población en general (y que incluye, por tanto, aquellos que siendo beneficiarios potenciales del servicio no lo utilizan). Otro procedimiento de recogida de datos es el que se fundamenta en la observación directa, si bien para esto es preciso proceder de una forma lo bastante sistemática como para que diferentes observadores otorguen una misma calificación a una misma condición (2).

La producción periódica y continuada de datos sobre la eficacia de un servicio debería permitir la comparación interanual de los resultados globales obtenidos por una misma Administración. Este proceso debería a su vez permitir elaborar bases de datos que sirvan para formular objetivos más racionales.

Una segunda posibilidad para la interpretación de los datos consiste en comparar los resultados para distintos grupos de población en función de criterios de equidad. Para poder garantizar que ningún grupo de población no recibe peores servicios que la media, la Administración precisa datos sobre la eficacia para distintos grupos de población. En este sentido, los datos se pueden separar en función, por ejemplo, del lugar de residencia, de la edad, el sexo, la renta o el lugar de origen.

La comparación con otras Administraciones ofrece otra fuente de interpretación, si bien en este sentido a menudo nos encontraremos con problemas derivados de las diferencias en las problemáticas de distintas comunidades.

2. Segundo paso: *la evaluación de las políticas y programas*

Los datos producidos por las Administraciones en su actividad de medida periódica y continuada constituyen una buena base de partida para la evaluación de la eficacia de las políticas y programas que, como se ha indicado más arriba, en puridad, implica la idea de aislar el impacto de la intervención pública de otros factores externos que puedan encontrarse en la causa de los efectos observados. ¿Cómo se consigue esto?

(2) Sobre selección de variables y fuentes de datos, *vid.* la literatura sobre métodos de investigación empírica. En general, para ciencias políticas, Buttolph Johnson y Joslyn (1986).

a) *La evaluación de resultados*

Siguiendo con el procedimiento hasta aquí explicado y partiendo de que existen unos objetivos específicos y unos indicadores de estos objetivos susceptibles de medición, todo lo que tendría que hacerse es recoger datos sobre los indicadores para aquellos individuos que se beneficiaron del programa y para un grupo de control equivalente que no hubiera obtenido estos beneficios, y analizar los datos sobre los participantes y los no participantes, en función de los objetivos de partida (Weiss, 1972: 24).

Concretamente, en los diseños experimentales la asignación aleatoria de los individuos a grupos experimentales y de control distribuye los posibles efectos externos a la intervención de forma aleatoria entre los dos grupos; mientras que en los diseños cuasi-experimentales es el investigador quien procura identificar las características del grupo que recibe el programa, para después intentar formar un grupo de comparación no equivalente, pero sí similar en todas las características relevantes al grupo tratado. Los diseños no experimentales, en cambio, analizan los resultados para un único grupo de sujetos que pueden presentar variaciones tanto respecto a su participación en el programa como a otros factores, para, a través de métodos estadísticos, eliminar las posibles diferencias entre los sujetos que no sean debidas a la intervención (3).

En la práctica, sin embargo, lo que parece fácil puede complicarse enormemente. Esta aproximación, que hemos llamado en otro lugar «científico-racional», es adecuada cuando se trata de programas bien definidos, donde una serie de individuos reciben una misma intervención, cuyos efectos pueden ser anticipados y medidos. No lo es tanto, en cambio, cuando los programas no reúnen estas condiciones, tanto por los problemas que presentan la identificación y operacionalización de los objetivos como por la complejidad intrínseca de los mismos programas o la dificultad técnica de aplicar los diseños que permiten aislar el impacto neto de la intervención pública.

b) *La evaluación de proceso*

Ciertamente, los objetivos de las políticas y programas más innovadores acostumbran a ser generales, abstractos y confusos. Esto es así porque el proceso de decisión pública se fundamenta en acuerdos sobre aspectos generales. Se podría afirmar que en la negociación política se hace sobre programas y no sobre objetivos. Una vez existe un acuerdo de principio sobre la necesidad de hacer «algo» en relación con los problemas que se hacen un hueco en la

(3) Para una introducción en castellano a los diseños de tipo experimental, desde una perspectiva de políticas públicas, *vid.* Alvira Martín (1991) o Ballart (1992).

agenda política, «lo que se tenga que hacer» escapa al acuerdo, de lo que resulta la aprobación de programas generales sin objetivos específicos.

A menudo las políticas y los programas son extremadamente complejos, disponen de distintos medios alternativos para alcanzar objetivos múltiples y a menudo contradictorios. Esto dificulta la definición de sus componentes, la reconstrucción de la cadena de efectos buscados y, por tanto, su reducción a un modelo simplificado.

Los diseños de investigación que parten de la idea de establecer controles experimentales y estadísticos presentan igualmente dificultades relacionadas no tanto con los objetivos de las políticas como con la aplicación del mismo dispositivo de evaluación. Muchas veces, simplemente no será posible construir los controles que requieren las técnicas experimentales o estadísticas.

Estas dificultades y la demanda de evaluaciones que puedan ser realizadas en períodos de tiempo relativamente cortos, sin comprometer grandes cantidades de recursos, y que respondan a las necesidades de sus usuarios, ha dado lugar a la formulación de una serie de modelos que pueden caracterizarse de forma general por el hecho de recurrir a aproximaciones cualitativas.

Desde esta perspectiva se podría hablar de un modelo de evaluación de proceso (Monnier, 1990-1991: 136-140), que se caracterizaría por el intento de adaptarse a la concepción de las políticas públicas como entidades dinámicas que van transformándose en la medida que son llevadas a la práctica. Si efectivamente los objetivos iniciales de una política son objeto de variaciones importantes, los recursos utilizados no coinciden con los recursos previstos y, en general, se inicia la política como una experiencia abierta de la que se espera aprender de cara al futuro, entonces la evaluación deberá centrarse en la identificación y comprensión de los hechos: en el proceso. La evaluación se convierte entonces en una investigación sobre qué hechos se encuentran en el origen de los éxitos y de los fracasos observados.

Para ello puede ser útil adoptar instrumentos de participación pluralista en el proceso evaluador, con el objeto de incorporar los valores y las opiniones del conjunto de actores que tengan relación con la política o programa evaluado. En este sentido puede ser importante no perderse la diversidad de percepciones sobre los criterios de evaluación que deben aplicarse en cada situación específica (Majone, 1989: 168).

Esta aproximación responde igualmente a la preocupación por la utilidad efectiva de los resultados de las investigaciones. En este sentido se entiende que al ofrecer al conjunto de actores implicados la posibilidad de incidir en el proceso evaluador y hasta de decidir las cuestiones que deban examinarse, se incrementarán las posibilidades de que los resultados de la evaluación sean eventualmente de alguna utilidad (Patton, 1986: 330).

Finalmente, creemos que se puede afirmar que la evaluación de los procesos —y más si es pluralista o participativa, dado que precisa un elevado nivel de familiaridad con las intervenciones y con sus principales actores (su experiencia, su posición en relación con la organización, sus opiniones y argumentos)— supone no obtener ningún tipo de medida sobre los efectos concretos de los programas en poblaciones específicas.

IV. ORGANIZACION INSTITUCIONAL DE LA EVALUACION

La evaluación de la acción pública se produce principalmente en aquellos países donde existe un desarrollo importante de las políticas de bienestar social y en las organizaciones de la ONU dedicadas a programas de ayuda internacional. De una parte, Estados Unidos, Canadá, Alemania, Suecia y Gran Bretaña cuentan con iniciativas de este tipo desde finales de los años sesenta, mientras que Noruega, Dinamarca, Holanda y Francia intensifican la demanda de evaluación en los años ochenta (Derlien, 1990: 148-149). De otra parte, la evaluación de políticas públicas se encuentra formalmente institucionalizada en la ONU, y de forma singular en el Banco Mundial, OMS, FAO, UNESCO y UNIDO (Hellstern, 196: 282).

Empezamos el análisis comparado con el caso de los Estados Unidos (EEUU en adelante), indudablemente el país con un sistema de evaluación más avanzado, tanto por su elevado nivel de institucionalización como por el grado de sofisticación metodológica alcanzado. Generalmente, se afirma que únicamente en los EEUU, en el Canadá y en algunas organizaciones internacionales existe una política de evaluación sistemática de las políticas y programas públicos, mientras que en los demás países de la Europa occidental ésta se produce *ad-hoc*, es decir, en relación con demandas específicas.

Ciertamente, la experiencia europea en este campo es mucho más limitada y no alcanza todavía el carácter de una función asumida como propia por las instituciones. Sin embargo, la presentación de algunos datos sobre la experiencia de cinco países europeos (Reino Unido, Alemania, Suecia, Holanda y Francia) nos permite sacar algunas conclusiones sobre los factores que conducen a un mayor desarrollo de la evaluación en unos determinados países y sobre la política de evaluación que puede tener mayor sentido en nuestro país, partiendo de la experiencia comparada en esta materia (4).

(4) La información que se presenta a continuación se basa en un limitado número de publicaciones para EEUU y en los trabajos del «Working Group on Policy Evaluation» —coordinado por R. Rist en el marco de las reuniones organizadas por el Instituto de Ciencias Admi-

1. *Estados Unidos: el caso del mercado más abierto y plural*

La evaluación de los programas públicos emerge en los EEUU sobre la base de una tradición consolidada de aplicación de las ciencias sociales al análisis de las políticas de las instituciones de gobierno. Esta tradición tiene su origen en las políticas federales del *New Deal*, y se desarrolla de forma intensiva durante la década de los sesenta para dar respuesta a la preocupación creciente sobre la utilidad de los programas de la «Great Society» y de la «Guerra contra la pobreza» del presidente Johnson. Desde entonces ha producido numerosos estudios empíricos sobre todo tipo de programas sociales, que se caracterizan por incorporar las técnicas de investigación y los métodos estadísticos propios de las ciencias sociales, lo que hace de la misma evaluación un área de conocimiento interdisciplinar.

La década de los sesenta se corresponde asimismo con la extensión de la reforma presupuestaria conocida como «Planning, Programming and Budgeting System» (en adelante PPBS) a toda la Administración federal norteamericana. El PPBS proponía organizar el gasto en función del producto final, y de su contribución por tanto a los objetivos de las diversas unidades y servicios públicos, lo que permitiría agrupar a aquellos programas que tuvieran una forma similar de tratar un determinado problema y establecer comparaciones entre programas alternativos. Si bien esta aproximación racionalizadora no pudo superar las dificultades encontradas para llevar a cabo estimaciones válidas y fiables de los resultados de la intervención pública *a priori*, sí puso de manifiesto las posibilidades del análisis retrospectivo, sobre la base de mejorar las bases de datos y el conocimiento sobre la forma de operar la intervención pública sobre la realidad.

Estas dos corrientes, que según todos los autores (Rossi y Freeman, 1985: 28; Chelimsky, 1985: 2; Nioche, 1988: 136) se encuentran en el origen de la evaluación de programas en los EEUU, tienen sin embargo un peso desigual. Efectivamente, como ha señalado Nioche:

El lento y progresivo desarrollo de la evaluación empírica de las políticas sobre la base de la investigación aplicada y de la acumulación de conocimientos y *know-how* demostró ser más eficiente y útil para los decisores públicos que no la aproximación tecnocrática del PPBS (1988: 59).

nistrativas de Bruselas— para las experiencias europeas. El autor quiere expresar su agradecimiento a los profesores J. Leca, presidente del Conseil Scientifique de l'Evaluation de Francia; J. P. Nioche, profesor de Hautes Etudes Commerciales» en París, y E. Monnier, director del Laboratoire d'Evaluation pour les Organismes et les Politiques Publiques de Lyon, por los materiales facilitados y por sus explicaciones sobre el caso francés.

La segunda mitad de los años sesenta y los setenta experimentaron el crecimiento más importante de las prácticas de evaluación, tanto a nivel de demanda como de oferta tanto pública como privada. Entre las evaluaciones llevadas a cabo, algunas fueron especialmente complejas, tanto por sus dimensiones (en cuanto a número de personas afectadas y presupuesto) como por los diseños de investigación aplicados. En este sentido, las más espectaculares fueron los experimentos que intentaban determinar los efectos de un impuesto negativo asociado a la garantía de un nivel mínimo de renta, los experimentos con los llamados «bonos de vivienda» y los desarrollados en el ámbito de la seguridad ciudadana y la criminología (5).

Este enorme esfuerzo para determinar los efectos de la intervención pública ha dado lugar a una verdadera «profesión», como demuestran, primero, el número de publicaciones y revistas aparecidas (*Evaluation Review*, *Evaluation and Program Planning*, *Evaluation Studies Review Annual*); segundo, la creación de asociaciones nacionales y comités dedicados al control de la calidad de los trabajos («Evaluation Research Society»), y tercero, el crecimiento experimentado por las empresas e institutos de investigación que se crearon originariamente para optar a los contratos ofrecidos por las Administraciones públicas (Abt. Ass., Systems Development Corp., Mitre, Rand, Stanford International).

La normalización de la función de evaluación en el sistema político-administrativo tuvo su traducción en la creación de unidades especializadas tanto dentro de la Administración como dependientes del Congreso.

- A nivel del ejecutivo, los departamentos y agencias que gestionan la mayor parte de los programas sociales (trabajo, sanidad, educación y bienestar) adoptan estructuras de evaluación, apoyadas por la Oficina del Presupuesto, que se encarga de llevar a cabo una labor de estímulo y divulgación de metodologías.
- A nivel del legislativo, el principal instrumento para la evaluación de los programas de la Administración es la «General Accounting Office» (en adelante, GAO). Esta agencia, que a principios de los ochenta contaba con más de 3.000 empleados (Walker, 1985: 272), puede actuar por iniciativa propia, a requerimiento de un miembro del Congreso o de un departamento ejecutivo, o simplemente porque así lo establece la legislación que autoriza la intervención. Además de la GAO, el «Congressional Research Service», el «Congressional Budget Office» y la «Office of Technological Assesment» pueden trabajar

(5) Vid. Ballart (1992), para una descripción más detallada de estas experiencias.

para el Congreso en la evaluación de opciones programáticas alternativas.

Resumiendo, pues, la evaluación constituye hoy en los EEUU una función perfectamente consolidada y que ha dado lugar a un mercado plural en el que participan tanto instituciones públicas y universidades como institutos de investigación, fundaciones y empresas privadas.

2. Reino Unido: *de la preocupación por el impacto final a las iniciativas centradas en la productividad*

A finales de los años sesenta y principios de los setenta el interés por el análisis de las políticas y la evaluación retrospectiva de sus resultados encontró dos puntos de apoyo, según Jenkins y Grey (1990: 56), en las iniciativas del Tesoro con el propósito de poner en funcionamiento sistemas efectivos de control del gasto, y en el movimiento de reforma que reclamaba más gestión en el sector público y más racionalidad en los procesos de decisión a nivel central.

Ambas preocupaciones quedaron reflejadas en el libro blanco sobre «La Reorganización del Gobierno central» de 1970 y se traducirían en la creación del *Central Policy Review Staff (CPRS)*, a nivel de gabinete y del *Program Analysis and Review (PAR)*, a nivel de departamentos ministeriales. El CPRS adoptó la forma de un *think tank* integrado por altos funcionarios de la Administración, expertos académicos y dirigentes de empresas privadas, que tenían por misión generar propuestas de evaluación a nivel de gabinete. Mientras que el PAR se estructuró como un programa que tenía por objetivo forzar a los departamentos ministeriales a entrar en el análisis y en la evaluación de sus servicios.

Ambas iniciativas acabarían por desaparecer sin haber conseguido sus propósitos. Entre las razones del fracaso del PAR, Jenkins y Grey (1990: 57) han señalado, primero, dificultades técnicas derivadas de la pobreza de los datos existentes y de la falta de personal con la preparación adecuada; segundo, dificultades a nivel de organización, en la medida que los departamentos no contaban con unidades especializadas, y tercero, dificultades políticas, dado que la evaluación presentaba algunos riesgos.

La llegada al poder de M. Thatcher en 1979 supuso un cambio de orientación: la preocupación por el control de los recursos utilizados sustituiría la preocupación por cuantificar el *output* público o por delimitar los efectos netos de la intervención. En lugar del CPRS y del PAR aparecieron, primero,

la *Efficiency Strategy (ES)* y después la llamada *Financial Management initiative (FMI)*.

La «ES» fue diseñada por D. Rayner, un conocido hombre de negocios, a quien M. Thatcher encargó el programa de eficiencia de su gobierno. Su estrategia para conseguir mayor eficiencia en la Administración central consistió en organizar una oficina de pequeñas dimensiones con apoyo directo de la Primera Ministra, y con la autoridad y capacidad necesarias para llevar a cabo estudios rápidos de actividades específicas de los departamentos y agencias administrativas. El programa perseguía objetivos relacionados no únicamente con la idea de productividad —sacar más provecho de los recursos invertidos y eliminar obstáculos burocráticos a una mejor gestión—, sino también con la idea de implementar los cambios propuestos, de una forma rápida y efectiva.

Según Jenkins y Gray (1990: 60), si bien los resultados de esta estrategia en términos de evaluación fueron bastante limitados, el apoyo continuado que obtuvo de la Primera Ministra permitió desarrollar una tarea de pequeñas reformas que ha sido muy bien recibida por el Parlamento y la opinión pública.

La «FMI» fue la respuesta del gobierno conservador a la demanda de un sistema más amplio de revisión de políticas por parte del Parlamento. Sin embargo, a diferencia de las iniciativas de los años setenta, la «FMI» reflejaba una mayor preocupación por la contención del gasto y por el desarrollo de sistemas descentralizados de gestión responsable. Coordinada centralmente por un equipo de reducidas dimensiones, la «FMI» encargaba a los departamentos ministeriales la elaboración de sus estrategias, la descripción de las actividades que estaban llevando a cabo para gestionar sus objetivos, y el establecimiento de mecanismos de asignación de responsabilidades por los recursos utilizados.

Como en el caso de la «ES», la «FMI» contó con el apoyo incondicional de la Primera Ministra y ha dado lugar a sistemas de control del gasto y de medida de *outputs* en relación con los objetivos iniciales. Jenkins y Grey (1990: 61) señalan, sin embargo, cómo la excesiva preocupación por los recursos y por el establecimiento de sistemas de control de gestión internos ha dejado poco margen a la evaluación de la eficacia final de las políticas.

A nivel parlamentario, la «Comisión de Contabilidad Pública» ha visto reforzada su capacidad de control desde la aprobación de una ley en 1983, por la que se amplían las funciones tradicionales del «Comptroller and Auditor General» y de su personal en la «National Audit Office» (equivalente a nuestro Tribunal de Cuentas) con las de investigar si los departamentos del ejecutivo utilizan los recursos de que disponen de una forma eficiente y eficaz. Si bien sus investigaciones difícilmente podrían considerarse evaluaciones en sentido

estricto, la selección de nuevos profesionales y la extensión de su ámbito de actuación puede tener consecuencias importantes, a pesar de las limitaciones que impone el sistema mayoritario de gobierno.

3. *Alemania: institucionalización de la evaluación en la Administración central*

En Alemania, al igual que en los EEUU y en el Reino Unido, el desarrollo de la evaluación se asocia a la progresiva adopción de nuevas políticas y programas en las áreas de educación, sanidad, vivienda y asuntos sociales a finales de los años sesenta y principio de los setenta, y a la necesidad política de justificar el incremento del gasto público en nuevas intervenciones cuyos efectos no habían sido demostrados de forma empírica.

Según Derlien (1990: 38), después de la formación del gobierno de coalición social-liberal de 1969 y de la primera ola de reformas administrativas (unidades de planificación, modelos de previsión, análisis coste-beneficios), la evaluación sería practicada e institucionalizada de forma gradual, tanto a nivel federal como a nivel de los estados y ciudades importantes. Sin embargo, este mismo autor señala que hay que buscar las causas concretas del desarrollo inicial de la evaluación en la presiones políticas del Parlamento, en el papel otorgado a Hacienda por las leyes de reforma presupuestaria y en la predisposición de la Administración a mejorar sus sistemas de información para crear argumentos con los que hacer frente a las críticas externas.

De forma general, se considera sin embargo que la evaluación se desarrolla básicamente a nivel ministerial, lo que tuvo su traducción a nivel institucional en la creación de unidades especializadas en algunos ministerios centrales. Concretamente se crearon unidades de evaluación en los Ministerios de Agricultura, Transportes, Ayuda al Desarrollo y Tecnología. A su vez, las Agencias Federales para las Relaciones con la Prensa y para la Divulgación de la Información sobre la Salud, establecieron organismos de evaluación de sus políticas de comunicación. En relación con estos organismos, Derlien (1990: 39) señala que no se trataba tanto de que las evaluaciones fueran efectivamente realizadas por el personal administrativo-funcionario como de que los ministerios contaran con especialistas que pudieran mantener contactos con equipos de investigadores externos, contratar los que fueran más adecuados para cada caso y discutir los diseños de investigación y los trabajos de recolección de datos.

Otra fuente de evaluaciones fueron los experimentos en materia de vivienda, educación y sanidad que, como en EEUU, se llevaron a cabo en Alemania a principios de los años setenta con el propósito de determinar los resultados

de una intervención antes de generalizar su aplicación a toda la población. La supervisión de estos experimentos y en general de las actividades de evaluación de los ministerios centrales ha sido desempeñada en Alemania, según Derlien (1990: 40), por una sección especializada en técnicas de planificación y evaluación que se creó dentro de la Oficina del Canciller en 1969.

El Parlamento federal también ha jugado un importante papel en el establecimiento de mecanismos de control y evaluación. Al igual que en los EEUU, desde finales de los años sesenta el Parlamento ha legislado para numerosos programas la obligación de aportar información sobre su implementación y efectos. Ejemplos de esta política legislativa han sido importantes leyes como, por ejemplo, la Ley de Mejora de la Estructura Económica Regional, la Ley del Mercado y de la Ocupación, la Ley de Subsidios Sociales, la Ley de Inversión en Hospitales, la Ley de Becas y Ayudas, la Ley del Aborto, la Ley de Protección de los Arrendatarios y varias leyes impositivas.

Otra forma de forzar la evaluación de programas adoptada por el legislativo alemán ha consistido en estipular la terminación automática de un programa en un determinado período de tiempo a no ser que demuestre su eficacia (la llamada «Sunset Legislation» norteamericana). Esta técnica legislativa se aplicó, por ejemplo, en la citada Ley de Protección de los Arrendatarios.

El Tribunal de Cuentas Federal, por el contrario, no ha entrado en la valoración de las políticas y programas públicos en términos de eficacia y eficiencia, si bien ésta es una tarea que le encomienda la legislación. Según Derlien (1990: 40), la posición políticamente neutral de esta institución, su misión tradicional de auditoría de las cuentas públicas y la formación de su personal se hallan entre las posibles causas de este abstencionismo. Desde 1985, sin embargo, el Tribunal de Cuentas recoge en su informe anual los resultados de las evaluaciones llevadas a cabo por los mismos Ministerios.

4. Suecia:

fragmentación y descentralización de la evaluación

Suecia ha sido uno de los primeros países en desarrollar estudios de evaluación de las políticas públicas, y son muchos los estudios realizados desde mediados de los años sesenta en relación con políticas de asistencia social y sanitaria, educación, trabajo y conservación de la energía (Vedung, 1990: 3). Ello no debería sorprender tratándose de un país en el que, según Ysander (1984: 53), más del 70 por 1000 del montante total del producto interior pasa por los presupuestos públicos.

Otros autores (Deleau, *et al.*, 1986: 146) han señalado, sin embargo, otros factores que podrían explicar una menor presión política en el caso sueco

para realizar análisis *ex post* de los programas públicos. Primero, hasta hace relativamente pocos años, Suecia ha mantenido un fuerte consenso sobre la justificación de los programas públicos de intervención, sobre todo en el área de bienestar social. Segundo, existe una percepción generalizada de que la Administración funciona bien. Tercero, la implementación de un sistema de planificación y presupuestación similar al PPBBS norteamericano sirvió en Suecia para descentralizar de forma significativa la Administración central, hasta el punto que dos tercios del presupuesto son administrados por municipios y corporaciones de carácter supramunicipal.

Hasta las reducidas dimensiones del país y la proximidad de la Administración a los usuarios finales de los servicios podrían haber funcionado como un elemento sustitutorio de la evaluación. Sin embargo, Suecia ha desarrollado su propio sistema, que como demuestra Vedung (1990: 16), no se adapta a ninguno de los modelos de organización de la evaluación discutidos por la literatura.

El modelo sueco se caracteriza por la fragmentación de la evaluación que se desarrolla de forma independiente bajo el control de las múltiples comisiones creadas *ad-hoc* para la formulación o enmienda de las políticas de cada uno de los distintos niveles de gobierno.

A nivel central, las leyes más importantes son elaboradas y discutidas por una comisión que nombra el gobierno y que incluye a altos funcionarios, a representantes de los partidos políticos y de los grupos de interés, y a expertos especialistas del mundo académico. Corresponde a estas comisiones el poder de decisión sobre los estudios previos que deban realizarse, y sobre la información que sea preciso recoger sobre los resultados de políticas pasadas. Además, los Ministerios centrales suecos generalmente cuentan con una comisión de investigación y desarrollo, de composición igualmente plural, encargada de desarrollar la investigación de aquellos aspectos que precise la política del Ministerio. Esta constituye una segunda fuente importante de evaluaciones a nivel central.

A nivel local, dada la importancia de este nivel de gobierno en la prestación de los servicios, tanto las asociaciones nacionales de municipios como las mismas instituciones centrales de gobierno han impulsado el desarrollo de medidas comparativas y de evaluaciones de políticas sectoriales.

Una tercera fuente de evaluaciones ha sido en el caso de Suecia la Oficina Nacional de Control o Tribunal de Cuentas que, por imperativo legal desde 1967 se ocupa de evaluar la actividad y la eficiencia con la que actúan las agencias nacionales en las que se organizan los ministerios centrales en Suecia. A diferencia de otros casos estudiados, la Oficina Nacional de Control sueca ha ido ampliando progresivamente su papel desde una actividad de control

externo a una actividad de control más directo: su estrategia consiste en examinar, primero, la misión y objetivos de la agencia; segundo, su actividad, y tercero, sus mecanismos de control interno de gestión. A su vez, la Oficina Nacional de Control estimula a las agencias administrativas para que analicen sus actividades por sí mismas y les ofrece información y sugerencias que puedan interesarles (Vedung, 1990: 9).

Este modelo ha sido seguido asimismo por la Oficina del Auditor General noruego, que incluye desde 1986 una sección de Auditoría de Rendimientos que tiene como misión la evaluación de la utilización de recursos y bienes públicos en base al análisis crítico de objetivos y resultados (Eriksen, 1990: 127). En este sentido, Noruega sigue, con retraso, los pasos de Suecia en la adaptación efectiva de esta institución de control a sus nuevas obligaciones legales.

5. *Holanda: la reconsideración de las políticas*

La introducción del análisis de las políticas y programas públicos tuvo lugar en el caso de Holanda a principios de los años setenta, con la creación por el Ministerio de Hacienda de un Comité Interdepartamental para el Desarrollo del Análisis de las Políticas. El llamado «COBA» estaba integrado por representantes de todos los ministerios que contaban con el soporte de personal especializado del Ministerio de Hacienda y tenían por misión promover la aplicación de análisis científicos de las políticas ministeriales. En sus años de existencia, según Bemelmans-Videc, *et al.* (1990: 110), las actividades del COBA se centraron en la clarificación de los objetivos de los departamentos ministeriales, en base al desarrollo de estructuras que iban desde los propósitos más generales y abstractos hasta los objetivos operacionales concretos que debían tener indicadores claros de rendimiento.

Estos mismos autores holandeses señalan, sin embargo, como las dificultades encontradas tanto a nivel político como administrativo, llevaron en último término a la desaparición del COBA. Ni gestores ni políticos estaban excesivamente interesados en fijar objetivos claros y específicos, puesto que ello les hacía más vulnerables, mostraba sus propios errores y, en último término, su ineficacia a la hora de alcanzar sus propósitos.

Una comisión formada en el año 1980 para estudiar la estructura organizativa de la Administración Pública holandesa recomendó en su informe final la sistematización de la evaluación de las políticas en las instituciones centrales de gobierno por la vía de incorporar la técnica de la llamada «Sunset Legislation», esto es, programas que tienen una fecha límite cuya continuación depende de que demuestren su eficacia. En esta dirección, el impulso del Parlamento, al estipular la obligatoriedad de las evaluaciones de determinadas

políticas en las mismas leyes que las aprobaron, ha sido importante para la institucionalización de la evaluación en Holanda.

Al mismo tiempo, la recesión económica de principios de los ochenta y la necesidad de encontrar formas de reducir el gasto público crearon las condiciones necesarias para que esta técnica legislativa y, en general, la idea de evaluar retrospectivamente las políticas y programas encontrarán un apoyo que antes no tenían de la parte de políticos y altos funcionarios. Según Bemelmans-Videc *et al.* (1990:112), la Administración central holandesa ha dado un fuerte empuje a la demanda de evaluaciones en esta última década, en todas las materias relacionadas con las políticas de bienestar social (ocupación, seguridad social, participación ciudadana a nivel local), lo que ha contribuido de forma notable a crear un mercado plural en el que participan, además de las administraciones públicas, instituciones relacionadas con el mundo académico y empresas de consultoría privadas.

De forma general se considera, sin embargo, que el ejemplo más importante de evaluación sistemática de las políticas es, sin embargo, el llamado *Procedimiento de Reconsideración de las Políticas (PRP)*. Este procedimiento sigue el año fiscal y consiste en la reconsideración de los programas y servicios estudiados, con el propósito de proponer alternativas a su modo de operar que redunden en un menor gasto público.

La selección de los programas que son objeto de reconsideración cada ejercicio fiscal corresponde al Consejo de Ministros y los resultados de las investigaciones son igualmente discutidos por el Consejo en pleno. Los trabajos son realizados por equipos mixtos, dirigidos por un representante del ministerio más directamente afectado, que además suele contar con la asistencia de un comité de altos funcionarios y representantes del Ministerio de Hacienda.

El procedimiento de reconsideración implica tres tareas: primero, describir la política o programa; segundo, evaluar el desarrollo del proceso de implementación, eficacia y eficiencia, y tercero, proponer un mínimo de cuatro alternativas que requieran menos esfuerzo presupuestario, además de explorar las posibilidades de privatización, descentralización y o aplicación del principio de maximización de los beneficios.

Esta iniciativa ha sido juzgada con un éxito relativo debido, sobre todo, al carácter obligatorio del procedimiento, a la especificidad de las directrices y a los estrechos márgenes temporales en que deben completarse las investigaciones. El principal resultado ha sido, sin embargo, los recortes presupuestarios y las operaciones de reorganización administrativa, descentralización o hasta privatización (Bemelmans-Videc, *et al.*, 1990: 115).

6. *Francia:*
creación de un dispositivo central de evaluación
dependiente del primer ministro

El gobierno socialista francés sancionó en enero de 1990 la evaluación como procedimiento institucional de revisión de las políticas públicas dentro de la Administración francesa. La reforma fue el resultado final de un largo debate que empezó con una conferencia internacional en 1983 y con la presentación de dos informes: el primero en 1986 (Informe Deleau) y el segundo en 1989 (Informe Viveret).

Los esfuerzos para desarrollar una cierta capacidad de evaluación se remontan a los años sesenta, en que se adoptó el programa de la «Rationalisation des Choix Budgétaires» (RCB). Si bien la RCB se centraba en el análisis *ex ante* más que en la evaluación *ex post* de las políticas, la creación de unidades dentro de la Administración francesa especializadas en las nuevas técnicas de programación y presupuestación, generó una importante demanda de estudios que contribuiría a producir un cierto desarrollo de la evaluación en Francia.

En 1987, escribiría Nioche, «la evaluación de las políticas en Francia es una actividad que avanza muy lentamente y que no está organizada ni institucionalizada en el sistema político-administrativo» (1988: 16).

A su vez, este autor señalaba en el mismo texto cómo la preocupación por la evaluación se extiende en Francia a las instituciones de control tradicional —Tribunal de Cuentas e inspecciones ministeriales—, si bien este interés podría, sin embargo, redundar en un menor desarrollo de evaluaciones «científicas» y completas, en la medida que se confunde evaluación y control y se intenta mantener la primera como una función estrictamente interna de la Administración.

La organización e institucionalización de la evaluación en Francia se produce con la reforma Rocard de 1990 (Decreto 90-82, de 22 de enero de 1990). Para ello se crea una estructura administrativa nueva con la que se intenta ampliar a nivel interministerial las iniciativas de evaluación adoptadas por los ministerios individualmente considerados y dotar al conjunto de las administraciones de una política de evaluación coordinada.

Concretamente, el nuevo dispositivo de evaluación francés está constituido por: el Comité Interministerial de la Evaluación (CIME), el Fondo Nacional para el Desarrollo de la Evaluación (FNE) y por el Consejo Científico de la Evaluación (CSE).

El CIME está presidido por el primer ministro y reúne a los principales ministerios del gobierno. Actúa como instancia de coordinación de las evaluaciones del ejecutivo, y en este sentido recibe las proposiciones de evaluación de los ministerios. También pueden presentarle propuestas de evaluación

el Consejo de Estado, el Tribunal de Cuentas y el Defensor del Pueblo. Es, sin embargo, el CIME el órgano que tiene la autoridad formal para tomar una decisión final, y en este sentido afectar los recursos previstos en el Fondo Nacional para el Desarrollo de la Evaluación.

El FNE es, pues, controlado por el primer ministro a través del CIME. La legislación establece, sin embargo, una reserva de hasta un 20 por 100 del Fondo para evaluaciones cuya iniciativa sea adoptada por el Consejo Económico y Social. La utilización de esta reserva por parte de este organismo requiere a su vez la autorización previa del Consejo Científico de la Evaluación.

El CSE se compone de 11 miembros, nombrados por el presidente de la República, entre profesores, altos funcionarios de reconocido prestigio científico y/o administrativo. Su mandato es por un período de seis años, no renovable, y su función consiste en proporcionar asesoramiento científico al CIME. Concretamente, el CSE emite dos dictámenes sobre cada proyecto de evaluación: uno, antes que el primer ministro y el CIME decidan sobre la aprobación de un proyecto, y otro, después que la evaluación haya sido realizada. En este segundo dictamen, que se publica al mismo tiempo que el contenido de la evaluación, el CSE expresa su opinión sobre la calidad de los trabajos realizados.

Si bien este dispositivo de evaluación está llamado a ser el núcleo central de la evaluación en Francia, existen otras instancias públicas independientes cuya misión consiste en evaluar un sector preciso de actividad. Algunos ejemplos de instancias de evaluación dependientes del ejecutivo son el Comité Nacional de la Evaluación de las Universidades, el Comité Nacional de Evaluación de la Investigación, el Comité de Evaluación y de Prospectiva de Nuevas Profesiones, el Comité de Evaluación de la Política de las Ciudades y la Agencia Nacional para el Desarrollo de la Evaluación Médica. Mientras que dependen de la Asamblea Nacional, la Oficina Parlamentaria de Evaluación de las Opciones Científicas y Tecnológicas y la Comisión Nacional de Evaluación de la Renta Mínima de Inserción (CEOPS/ENTP, 1991).

La adopción de dispositivos similares al adoptado por el gobierno central a nivel de las regiones podría contribuir a un desarrollo más plural de la evaluación en Francia. En esta línea, el Consejo Regional de la Región Rhone-Alpes decidió en enero de 1991 crear un «Comité Regional de la Evaluación», equivalente al CSE nacional.

La legislación autorizante otorga al CRE funciones de definición previa de los programas que se someten a evaluación, de control tanto de las propuestas de evaluación como de los equipos encargados de llevarlas a cabo, y de supervisión *a posteriori* de los trabajos realizados. Como en el caso de las

instituciones centrales, la iniciativa queda en manos del presidente regional, si bien en este caso, de una parte, se incluye en el debate anual del presupuesto la deliberación sobre las políticas susceptibles de evaluación, y de otra parte se deja en manos de las comisiones correspondientes del Consejo Regional y del Comité Económico y Social Regional, la discusión sobre los resultados de las evaluaciones (Quermonne, *et al.*, 1991).

V. IMPLICACIONES DE LA EXPERIENCIA COMPARADA EN LA DEFINICION DE UNA POLITICA DE EVALUACION PROPIA

La observación de las experiencias comparadas pone de relieve algunas tendencias que es preciso tener en cuenta a la hora de definir una política de evaluación. Sin embargo, las diferencias económicas, políticas, institucionales y hasta de tradición en las ciencias sociales determinan unos condicionantes propios de cada país.

En primer lugar, la evaluación tiene su «edad dorada» en una época de expansión económica, en la que la ampliación automática de las bases imponibles financia casi sin costes políticos las nuevas políticas de intervención social. Cuando nuestro país se suma a estos procesos nos encontramos con una situación radicalmente distinta que determina que aquellos que antes se preocupaban por el impacto de las políticas cambien su orientación a favor de los análisis que sirvan para incrementar la productividad, racionalizar la gestión de unos recursos cada vez más limitados y hasta, en último término, justificar los recortes presupuestarios.

En segundo lugar, el sistema parlamentario español, tanto a nivel central como autonómico, se basa en una fuerte disciplina de partido que tiende a limitar la función de control del Parlamento. El partido o coalición mayoritaria hará todo lo que sea posible para evitar una discusión política de la que pueda salir perjudicado su gobierno, mientras que la oposición, más interesada en ejercer sus funciones de control, se ve fuertemente limitada por su condición minoritaria.

En tercer lugar, los Tribunales de Cuentas, central y autonómicos, a pesar de estar obligados legalmente a examinar la eficacia y la eficiencia con que actúan los organismos públicos de su jurisdicción, limitan su papel al de supervisión de las cuentas públicas de acuerdo con lo que establecen la ley del presupuesto y las demás leyes que sean aplicables.

En cuarto lugar, tanto por el lado de la oferta como por el lado de la demanda el número de profesionales con conocimiento de métodos susceptibles de ser aplicados a la evaluación de las políticas sociales es bastante

reducido. En el mismo sentido, los sistemas de información internos y en general los bancos de datos estadísticos, a pesar de los avances registrados en los últimos años, presentan importantes lagunas y/o no han sido construidos desde una perspectiva de medición de resultados y evaluación.

Estos factores podrían explicar el retraso de nuestro país en materia de evaluación a la vez que limitan una posible política de revisión de la actuación de los poderes públicos. En este sentido creemos que *lo más realista a corto plazo sería generar una demanda mínima de evaluaciones en distintos campos, realizar dos o tres estudios en una perspectiva temporal concreta e impulsar a los Parlamentos central y autonómicos y a los Tribunales de Cuentas a desarrollar su propia oferta de evaluación o como mínimo a que controlen la actividad desarrollada por las Administraciones.*

En este sentido constatamos como *la mayor parte de las evaluaciones tienen su origen en el ejecutivo.* Cuando han sido los ejecutivos centrales los que han tomado la iniciativa, generalmente lo han hecho para ocuparse de políticas de carácter interdepartamental. Un ejemplo de ello sería el CPRS británico o el mismo dispositivo creado en Francia por el gobierno Rocard. Como en el caso de la «Efficiency Strategy» y de la «Financial Management Initiative» de la época Thatcher o del «Procedimiento de Reconsideración» holandés cabe, sin embargo, que los gobiernos centrales decidan examinar programas más específicos y no sólo desde una perspectiva de eficacia, sino también con una clara orientación a problemas organizativos o de gestión con propósitos de racionalización y ahorro. A nivel de ministerios, en los países con más capacidad se crearon numerosas unidades de evaluación dentro de la Administración central (EEUU, Alemania), mientras que en los países más pequeños, grupos de expertos participan activamente en las actividades de evaluación de las agencias administrativas (Suecia, Holanda).

El *Parlamento* tiene, sin embargo, en la legislación la posibilidad de obligar al ejecutivo a presentar informes específicos en una perspectiva temporal concreta sobre el desarrollo y los efectos de las políticas aprobadas. Otra forma de institucionalizar la evaluación pasa por la adopción de estrategias parlamentarias del tipo «Sunset Legislation», que al condicionar la continuación de un programa a la demostración de su eficacia, obliga de hecho a realizar su evaluación.

Por fin, los *Tribunales de Cuentas* pueden ir cambiando la concepción de su propio rol en la medida que la preocupación por el análisis de la eficacia y de la eficiencia de la actividad pública vaya calando en otros sectores institucionales. Mientras tanto, los Tribunales de Cuentas podrían ir desarrollando un rol más activo en dos ámbitos relacionados: primero, la supervisión de las actividades de evaluación realizadas por el ejecutivo o como consecuen-

cia de eventuales obligaciones legales, y segundo, la selección y formación de personal especializado con la perspectiva de desarrollar una división de metodología que sirva de soporte al conjunto de las Administraciones.

REFERENCIAS BIBLIOGRAFICAS

- ALVIRA MARTIN, F.: *Metodología de la evaluación de programas*, Madrid, CIS, 1991.
- BALLART, X.: *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudios de caso*, Madrid, Instituto Nacional de Administración Pública, 1992.
- BEMELMANS-VIDEC, M. L.; ELTE, R.; KOOLHAAS, E.: «Policy Evaluation in the Netherlands: Institutional Context and State of Affairs», en R. C. RIST: *Program Evaluation and the Management of Government*, New Brunswick, Transaction Publishers, 1990.
- BUTTOLPH JOHNSON, J.; JOSLYN, R. A.: *Political Science Research Methods*, Washington, D. C., Congressional Quarterly Press, 1986.
- CEOPS-ENTPE: *Les Instances Nationales d'Evaluation en France*, Vaulx en Velin, documento de trabajo, 1991.
- CHELIMSKY, E.: *Program Evaluation: Patterns and Directions*, Washington, D. C., American Society for Public Administration, 1985.
- DELEAU, M.; NIOCHE, J. P.; PENZ, PH.; POINSARD, R.: «Les enseignements de quelques expériences nationales», *Documentation Française*, 1986, págs. 135-157.
- DERLIEN, H. U.: «Program Evaluation in the Federal Republic of Germany» y «Genesis and Structure of Evaluation Efforts in Comparative Perspective», en R. C. RIST: *Program Evaluation and the Management of Government*, New Brunswick, Transaction Publishers, 1990.
- ERICKSEN, B.: «Institucional Aspects of Evaluation in Norway», en R. C. RIST: *Program Evaluation and the Management of Government*, New Brunswick, Transaction Publishers, 1990.
- GUBA, E. G., y LINCOLN, Y. S.: «The countenances of Fourth-Generation Evaluation: Description, Judgement and Negotiation», en D. J. PALUMBO: *The Politics of Program Evaluation*, Beverly Hills, CA, Sage, 1987.
- HELLSTERN, G. M.: «Assessing Evaluation Research», en F. X. KAUFMANN, G. MAJONE y V. OSTROM: *Guidance, Control and Evaluation in the Public Sector*, Berlín, De Gruyter, 1986.
- JENKINS, B., y GREY, A.: «Policy Evaluation in British Government: From Idealism to Realism», en R. C. RIST: *Program Evaluation and the Management of Government*, New Brunswick, Transaction Publishers, 1990.
- MAJONE, E.: *Evidence, Argument and Persuasion in the Policy Process*, New Haven, CN, Yale University Press, 1989.
- MONNIER, E.: «Objetivos y destinatarios de las evaluaciones», *Documentación Administrativa*, 224-225, 1991-1992, págs. 131-152.
- NIOCHE, J. P.: «Accountable Public Management», en A. KAKABADSE, P. RONDO

- BROVETO y R. HOLZER (eds.): *Management Development and the Public Sector*, Avebury, Gower Publishers, 1988.
- PATTON, M. Q.: *Utilization-Focussed Evaluation*, Beverly Hills, CA, Sage, 1986.
- QUERMONNE, LARGER y MONNIER: *Un Dispositif d'Evaluation des Politiques Publiques pour la region Rhone-Alpes*, Vaulx en Verlin, documento de trabajo, 1991.
- ROSSI, P. H., y FREEMAN, H. E.: *Evaluation: A Systematic Approach*, Beverly Hills, CA, Sage, 1985.
- SHORTELL, S. M., y RICHARDSON, W. C.: *Health Program Evaluation*, Sant Louis, MI, Molby, 1978.
- SCRIVEN, M.: «Goal-Free Evaluation», en E. R. HOUSE (ed.): *School Evaluation, The politics and Process*, Berkeley, CA, McCutchan, 1973.
- FITZSIMMONS, J. A., y SULLIVAN, R. S.: *Service Operations Management*, Nueva York, McGraw-Hill, 1982.
- VEDUNG, E.: «Development of Evaluation in the Swedish Political System», trabajo presentado en la Conferencia Internacional de La Haya sobre Evaluación de Políticas y Programas, 1990.
- VIVERET, P.: «L'Evaluation des Politiques et des Actions Publiques: Rapport au Premier Ministre», informe no publicado, 1989.
- WALKER, W. E.: «The Conduct of Program Evaluations on Government», *Evaluation and Program Planning*, 8, 1985, págs. 271-280.
- WEISS, C. H.: *Evaluation Research*, Englewood Cliffs, NJ, Prentice Hall, 1972.
- YSANDER, B. C.: «L'Evaluation des Politiques Publiques en Suède», en J. P. NIOCHE y R. POINSARD (eds.): *L'Evaluation des Politiques Publiques*, París, Economica, 1984.