

Gestión y Análisis de Políticas Públicas, número 32, julio de 2023  
Sección: SECCIÓN MONOGRÁFICA. LA EXPERIENCIA AEVAL  
Recibido: 27-10-2021  
Aceptado: 27-09-2022  
Prepublicado: 01-12-2022  
Publicado: 13-07-2023  
ISSN: 1989-8991 – DOI: <https://doi.org/10.24965/gapp.11011>  
Páginas: 126-141



Referencia: Rico Callado, J. (2023). La metodología de evaluación en AEVAL. *Gestión y Análisis de Políticas Públicas*, (32), 126-141. <https://doi.org/10.24965/gapp.11011>

## La metodología de evaluación en AEVAL

### *The evaluation methodology in the The National Agency of Public Policies and Quality of Services Evaluation (AEVAL)*

Rico Callado, Javier

Politólogo y evaluador de políticas públicas (España – Spain)

ORCID: <https://orcid.org/0000-0001-9756-2808>

[javier.rico.callado@gmail.com](mailto:javier.rico.callado@gmail.com)

#### NOTA BIOGRÁFICA

Licenciado en Ciencias Políticas por la Universidad de Granada y especialista en políticas públicas y sociales por la Pompeu Fabra. Especialista en análisis avanzado de datos. Evaluador de políticas públicas en la extinta AEVAL y en el Instituto para la Evaluación de Políticas Públicas. Autor de distintas evaluaciones, guías de evaluación y artículos de políticas públicas, sociales y evaluación.

---

#### RESUMEN

El artículo analiza la metodología diseñada y aplicada por la Agencia Estatal de Evaluación de Políticas Públicas y la Calidad de los Servicios en la evaluación de programas y políticas públicas a lo largo de sus años de existencia (2007-2017). Una metodología ecléctica, principalmente pluralista, aunque adoptando parcialmente otras perspectivas y paradigmas. Fue el fruto de un aprendizaje y práctica de evaluación continuo. Se reflexiona sobre distintos elementos institucionales y organizativos que condicionaron tanto su metodología como su práctica evaluativa. Se analiza el diseño de las evaluaciones de AEVAL, con especial referencia a los métodos y técnicas utilizados, y su proceso de evaluación. Finalmente se destacan algunas de las fortalezas y debilidades de la metodología de la Agencia.

#### PALABRAS CLAVE

Evaluación; metodología de evaluación; AEVAL; proceso de evaluación; enfoque de evaluación; diseño de evaluación.

---

#### ABSTRACT

The article analyzes the methodology designed and applied by The National Agency for the Evaluation of Public Policies and Quality of Services in the evaluation of programs and public policies throughout their years of existence (2007-2017). An eclectic methodology, mainly pluralistic, although partially adopting other perspectives and paradigms. It was the result of continuous learning and evaluation practice. It reflects on different institutional and organizational elements that conditioned both its methodology and its evaluation practice. The design of AEVAL evaluations, with particular reference to the methods and techniques used, and their evaluation process, are also analysed. Finally, some of the strengths and weaknesses of the Agency's methodology are highlighted.

#### KEYWORDS

Evaluation; evaluation methodology; AEVAL; evaluation process; evaluation approach; evaluation design.

## SUMARIO

1. INTRODUCCIÓN. 2. METODOLOGÍA DEL ESTUDIO. 3. EL MODELO DE EVALUACIÓN DE AEVAL. 4. EL CONTEXTO INSTITUCIONAL DE AEVAL Y SU INCIDENCIA EN SU METODOLOGÍA Y EL DISEÑO DE LAS EVALUACIONES. 4.1. LIMITADA PRÁCTICA Y CULTURA DE EVALUACIÓN. 4.2. ESTRATEGIA Y PLANES DE EVALUACIONES. 4.3. CAPACIDADES EN EVALUACIÓN. 5. EL DISEÑO DE LAS EVALUACIONES AEVAL, SU OPERACIONALIZACIÓN Y LOS MÉTODOS DE ANÁLISIS. 6. EL PROCESO DE EVALUACIÓN EN AEVAL. 7. FORTALEZAS Y DEBILIDADES DEL MODELO AEVAL. REFERENCIAS BIBLIOGRÁFICAS.

## 1. INTRODUCCIÓN<sup>1</sup>

A lo largo de sus diez años de singladura la Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios (en adelante AEVAL) fue construyendo una metodología y modelo de evaluación ecléctico, propio, fruto del aprendizaje continuo, la práctica evaluadora desarrollada y factores de distinto orden.

Esta metodología, como no podía ser menos, tuvo un papel capital en una organización llamada a convertirse en el referente gubernamental en materia de evaluación de las políticas públicas y la difusión de la cultura de la evaluación en España, a través de su práctica evaluadora y la elaboración de guías metodológicas. Si el segundo requerimiento de las evaluaciones es que las evidencias encontradas respondan a las necesidades, visiones y prácticas de los *stakeholders*, la primera, y más importante de ellas, es que generen suficientes evidencias creíbles para ganar reputación científica (Chen, 2015, p. 26). Y el objetivo de la metodología del AEVAL era precisamente lograrlo, para dotar de solidez a su trabajo y desarrollar sus funciones de la forma más eficaz.

El modelo de evaluación de AEVAL no sólo debe comprenderse en términos de aproximaciones epistemológicas y metodológicas, sino fundamentalmente en el contexto de un proceso institucional de modernización administrativa, enmarcado desde el punto de vista político entre los principios de buen gobierno y como herramienta para fortalecer la rendición de cuentas y la eficacia de las políticas públicas. Una metodología orientada también a difundir un lenguaje común en evaluación, sus valores, reglas de juego y apertura a la realidad y a la lógica de la acción pública en sus diferentes niveles (AEVAL, 2010, p. 44). En definitiva, una metodología condicionada por el contexto institucional, la naturaleza y características de dicha Agencia y por todo un conjunto de dimensiones o «políticas de evaluación» que fueron articulándose progresivamente. En particular la definición, requisitos y objetivos de la evaluación; la construcción de capacidades en evaluación; y la gestión y proceso de las evaluaciones.

Recientemente el Gobierno ha aprobado el Proyecto de Ley de Institucionalización de la Evaluación de las Políticas Públicas en la Administración General del Estado, que contempla nuevamente la creación de una Agencia Estatal de Evaluación de Políticas Públicas. Este artículo pretende realizar una modesta contribución al debate sobre un sistema de evaluación público y la creación de una nueva Agencia de Evaluación, analizando la metodología de evaluación adoptada por AEVAL a lo largo de su existencia (2007-2017), así como sus fortalezas y debilidades. La estructura del artículo es la siguiente: en primer lugar, se enmarcará dicha metodología en el seno de los grandes paradigmas de la evaluación, y se constatará, a grandes rasgos, cómo fue llevada a la práctica por AEVAL en sus evaluaciones. En segundo lugar, se reflexionará en torno a algunos elementos institucionales y organizativos que condicionaron su metodología y práctica evaluativa. En tercer lugar, se analizará el diseño de las evaluaciones y cómo se operacionalizaban, junto con algunas referencias a los métodos y herramientas utilizados. En cuarto lugar, se analizará brevemente el proceso de evaluación. Finalmente se apuntarán algunas de las fortalezas y debilidades de la metodología de la Agencia. Se trata, en todo caso, de realizar una aproximación más bien general y reflexiva.

No se pretende reflexionar sobre el éxito o fracaso de AEVAL en cuanto a Agencia de evaluación, hecho que dependió en gran medida de su encaje institucional. Tal y como destacó Chelimsky (2009, p. 52), el éxito de la evaluación en el seno del gobierno depende en gran medida del marco político en la que se desarrolla dicha evaluación, hasta el punto que algunos fracasos de las estructuras y organismos de evaluación pueden atribuirse a cuestiones vinculadas a las presiones del contexto gubernamental y la adaptación a dicho entorno, y no sólo a los méritos de la metodología y el proceso de evaluación en sí mismos. Estos aspectos son analizados por Juan Antonio Garde en su artículo.

<sup>1</sup> Queremos agradecer a los evaluadores anónimos y a los editores de la revista sus sugerencias y comentarios a la primera versión de este artículo.

## 2. METODOLOGÍA DEL ESTUDIO

El método de investigación utilizado es el estudio cualitativo de caso. Dentro de la pluralidad de enfoques de este tipo de método (Creswell, 2013) se sigue el enfoque constructivista propugnado por Stake (2005), caracterizado por un tratamiento holístico del fenómeno, la interacción del investigador con el objeto de estudio y la aplicación de métodos de análisis inductivos y flexibles.

La metodología desarrollada por AEVAL se analizará tomando como referencia los documentos en los que se encuentra reflejada: Guía práctica para el diseño y la realización de evaluaciones de políticas públicas (2015), Fundamentos de evaluación de políticas públicas (2010) y La función evaluadora: principios orientadores y directrices de actuación de la evaluación de políticas y programas (AEVAL, 2009).

El análisis del grado en que las evaluaciones efectivamente realizadas a lo largo del período de existencia de AEVAL (2007-2017) reflejan los elementos metodológicos contenidos en estos documentos se realizará tomando como referencia doce evaluaciones de AEVAL, un 30 % de todas las evaluaciones seleccionadas, mediante un muestreo estratificado por conglomerados. Nueve corresponden al primer período (2007-2011), época en la que más evaluaciones realizó AEVAL. Y dos al segundo período (2012-2017). En todo caso, no se empleará una red conceptual o marco analítico concreto, por cuanto, si bien algunos autores abogan por esta sistemática (Yin, 2009), propia de enfoques postpositivistas, tiende a limitar y restringir en exceso el análisis. Por el contrario, se opta por seguir la recomendación de Stake y emplear un marco conceptual flexible desde la perspectiva constructivista.

Asimismo, se han realizado distintas entrevistas semiestructuradas en profundidad con distintos responsables y personal con funciones relevantes en el ámbito de la evaluación desarrollada por AEVAL. En dichas entrevistas se ha indagado principalmente la construcción de su metodología y su aplicación práctica, los factores institucionales que las condicionaron, así como los puntos fuertes y debilidades del modelo AEVAL. Todo ello acompañado de la experiencia del propio autor.

## 3. EL MODELO DE EVALUACIÓN DE AEVAL

La evaluación, junto con los análisis de políticas públicas, ha evolucionado ampliamente desde sus orígenes en la segunda mitad del siglo xx. Una de sus particularidades ha sido la rápida sucesión de teorías, modelos o enfoques de evaluación y el constante enfrentamiento entre paradigmas –la denominada «guerra de paradigmas» (Caracelli, 2000, p. 101)–. Una disciplina, por ende, en la que existe un escaso consenso. Es, principalmente, una desavenencia epistemológica y ontológica y consecuentemente también con respecto a qué método o aproximación debe utilizarse en evaluación. Una disciplina en la que, además, nos encontramos ante una situación que podría catalogarse de profundamente insatisfactoria, ambigua, carente de claridad conceptual y precisión (Belcher y Palenberg, 2018, p. 1).

¿Cuál fue la respuesta de AEVAL a estas discrepancias metodológicas? El modelo AEVAL fue ecléctico y una mixtura de los distintos paradigmas en evaluación. Adoptó principalmente la evaluación pluralista que presenta tres notas características. Por una parte, la evaluación es un proceso colectivo que promueve la intervención de los distintos actores y legitima los intereses existentes, y en el que los juicios de valor se formulan procurando llegar a un acuerdo (AEVAL, 2010; Weiss, 1983). En segundo lugar, su carácter político y el estar intrínsecamente unida a la toma de decisiones públicas. Y finalmente por adoptar la premisa de que no existe una única forma de abordar el análisis de las intervenciones públicas, sino una pluralidad de perspectivas. En este sentido se rechaza la jerarquía de métodos y evidencias que sitúa los estudios experimentales o las aproximaciones positivistas y postpositivistas en primer lugar.

El denominado modelo o enfoque AEVAL se bosquejó inicialmente en su documento La función evaluadora: principios orientadores y directrices de actuación de la evaluación de políticas y programas (AEVAL, 2009) y en Fundamentos de evaluación de políticas públicas, editado en 2010. Se plasmó definitivamente en la Guía práctica para el diseño y la realización de evaluaciones de políticas públicas del año 2015, que contiene la forma en que la Agencia afrontó los distintos debates y paradigmas en evaluación. Fue, al mismo tiempo, el fruto de su práctica evaluadora, puesto que uno de los elementos que caracterizó a la Agencia fue la necesidad de realizar evaluaciones desde el primer momento.

En todo caso, su metodología se fue construyendo progresivamente. A lo largo de los años 2007-2008 predominaron las aproximaciones heterogéneas y los informes respondían a un formato breve y ejecutivo, más próximo a las recomendaciones y análisis de expertos. No será hasta el año 2009-2010 cuando se

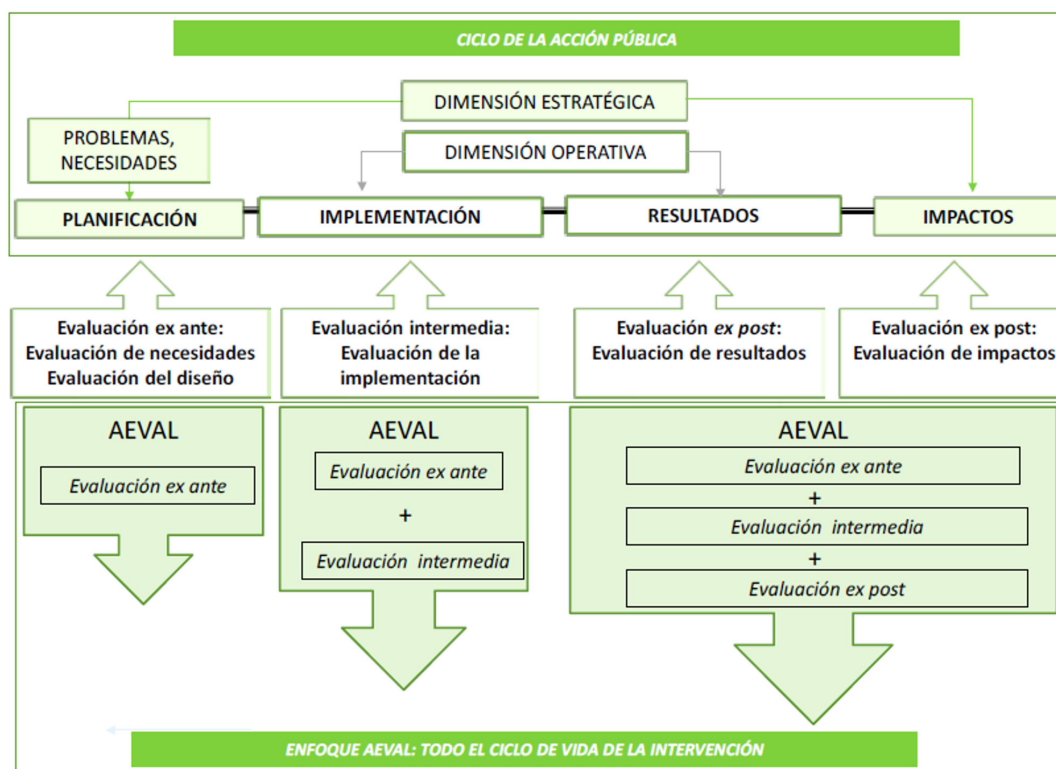
comienza a perfilar su modelo de evaluación y las evaluaciones siguen una cierta pauta y estructura común, con un mayor peso y presencia explícita en los informes de las herramientas y los métodos de análisis.

Uno de los rasgos distintivos del modelo AEVAL es su abordaje de la evaluación desde una perspectiva de políticas públicas. La evaluación es, ante todo, un instrumento de análisis de políticas, de carácter multidisciplinar. El predominio de la complejidad, la interconexión de los problemas públicos y la importancia de los aspectos políticos de la acción pública motivó que AEVAL intentase integrar en su metodología el análisis político-estratégico junto con el análisis de aspectos más operativos del despliegue de las intervenciones, como son los objetivos, recursos y resultados. En definitiva, se consideraba que no era posible alcanzar un juicio global sobre una intervención sin analizar en paralelo los componentes políticos y decisionales de la esfera política (AEVAL, 2015). La complejidad de las políticas públicas obligaba a que la evaluación se asentase, por tanto, en una pluralidad de enfoques y métodos de evaluación.

En esta perspectiva de políticas públicas, el análisis de los problemas públicos –que no dejan de ser un constructo social y político (Instituto para la Evaluación de Políticas Públicas, 2020, p. 20)–, su diagnóstico y el diseño de las intervenciones tenía particular trascendencia para AEVAL. En unas políticas y programas en los que las decisiones públicas no siempre son adoptadas sobre la base a criterios puramente racionales, las intervenciones no necesariamente están dotadas de una teoría causal o del cambio y no es infrecuente la inexistencia de un adecuado diseño previo o la ausencia de un diagnóstico del problema que se pretende abordar. De este modo se pretendió incorporar a la evaluación los postulados que apuntan a que el análisis del diseño es capital para comprender los resultados obtenidos por las intervenciones públicas. Se asume que los problemas de numerosas intervenciones son, en definitiva, los malos diseños.

Así, para AEVAL, la evaluación no puede focalizarse exclusivamente en los resultados o impactos, sino que es necesario que abarque todo el ciclo de la política pública o el programa. Y en los supuestos en los que el encargo de evaluación se centra en los potenciales resultados es necesario realizar al mismo tiempo una evaluación intermedia o de implementación y una evaluación *ex ante*. Circunscribir la evaluación exclusivamente a una fase del ciclo provoca que se pierdan causas explicativas de los fenómenos y obvia el carácter interrelacional de los distintos componentes de las políticas. Esto ocasiona que la evaluación no pueda subestimar la necesidad de considerar los sistemas globales y la política pública o programa como un todo.

GRÁFICO 1. EVALUACIÓN INTEGRAL. CICLO DE LA ACCIÓN PÚBLICA



Fuente: AEVAL (2015): Guía práctica para el diseño y la realización de evaluaciones de políticas públicas.

Se trata, en definitiva, de un enfoque de evaluación integral que pretende evitar el análisis fragmentado de unas políticas o programas que, para su adecuada comprensión, deben ser examinados de forma integral y holística. La propia definición de evaluación de AEVAL enfatiza la necesidad de comprensión global de las intervenciones y pretende alcanzar un juicio valorativo sobre todo el ciclo de la política (diseño, puesta en práctica y efectos, en cuanto a resultados e impactos) (AEVAL, 2015), definición que guarda similitudes con la formulada por Rossi et al. (2003). Esta aproximación resulta plenamente acertada y meritoria, pero al mismo tiempo representa un esfuerzo hercúleo e implica un evidente riesgo: sobredimensionar la evaluación, que puede resultar inabarcable, en términos de suficiente profundidad y calidad analítica, cuando el objeto de evaluación resulta complejo, se está frente a una política con perfiles amplios o no se disponen de los recursos necesarios para abordarla.

En esta perspectiva el papel de las partes interesadas o *stakeholders* se vuelve capital, no ya sólo en la formulación de los problemas y en el diseño, sino también en la implementación. AEVAL destacó la importancia que tienen determinados actores en la conformación y el desempeño de las políticas. Y asimismo reconoció la multiplicidad de perspectivas, intereses y valores, que es una de las características esenciales de la evaluación pluralista. Ello posibilitaba, además, dotar de utilidad a sus evaluaciones, que en la definición de su alcance incorporaba los planteamientos de la evaluación centrada en la utilización (*utilization-focused evaluation*). Esta visión se aproxima a lo que Chen denominó perspectiva de evaluación integral (Chen, 2015, p. 29). No obstante, esta visión se reducía en algunas de las evaluaciones de AEVAL a una mera identificación de los actores intervinientes, y a una descripción del papel representado en la intervención, con una escasa profundidad.

AEVAL no adoptó la evaluación tradicional o clásica basada en el paradigma positivista o post-positivista, para el que es posible desentrañar la realidad y los vínculos causales de las políticas con respecto a los efectos producidos de forma cierta, objetiva, neutra y cuantificable a través de la aplicación del método científico a los fenómenos sociales<sup>2</sup>; el predominio o énfasis de la medición a través métodos cuantitativos, en particular de los métodos experimentales o cuasi-experimentales. Y que aboga por las evaluaciones centradas en los resultados o efectos –preferiblemente netos–, y los objetivos perseguidos.

No se trató, en todo caso, de un rechazo frontal. Así, este paradigma está también presente en sus evaluaciones, en particular aquellas más centradas en los resultados. En algunas de estas evaluaciones se enfatiza la utilización del método científico positivista para aprehender fenómenos sociales y se recurre a la utilización de herramientas y métodos cuantitativos que refuercen el carácter científico-técnico de la evaluación: métodos y herramientas cuasi-experimentales, grupo de comparación intacto geográficamente local, modelizaciones econométricas y regresiones de todo tipo para identificar relaciones causales entre la intervención pública y los efectos observados. Se trata, por tanto, de resolver el problema de atribución desde un punto de vista cuantitativo.

Del mismo modo determinadas evaluaciones de AEVAL, en el análisis de los productos y los resultados, tomaban como referencia los objetivos perseguidos, siguiéndose parcialmente el método de evaluación basado en el logro de estas metas (*goal-based evaluation*), en las que la medición y cuantificación del logro de los objetivos introducía elementos de certeza y comprensibilidad. En este aspecto se observa en sus evaluaciones un predominio de la perspectiva de la intervención, en la que prima la descripción de los cambios producidos y se considera que la intervención es el origen del sistema causal. La secuencia lógica de análisis toma como punto de partida la intervención y asciende a través de la cadena de resultados: inputs, actividades, productos, resultados e impactos. En este engranaje cobra particular relevancia el contexto en el que se insertan las intervenciones públicas. Por el contrario, no está muy presente en AEVAL la perspectiva de sistema, en la que el punto de vista pasa de la intervención a los cambios de interés<sup>3</sup>.

El rechazo del paradigma positivista, en el caso de AEVAL, se asienta en determinados supuestos. Uno de ellos, ya esbozado, es el considerar las políticas públicas como procesos de acción complejos, con respecto a los que la evaluación debe trascender el análisis del grado de cumplimiento de los objetivos definidos para ser útil. Lo importante no es sólo la eficacia o la eficiencia (cuestión diferente es cómo se mida), sino el cómo y el por qué se logran los resultados, no ya sólo buscados, sino los realmente producidos, y cómo operan las intervenciones en un contexto determinado. Es preciso, por tanto, un pensamiento lateral y una

<sup>2</sup> Un paradigma fundamentado, además en la premisa de «racionalidad y suficiencia de la acción gubernamental, obviando el contexto, lo complejo, lo local», en palabras de AEVAL (2015, p. 26).

<sup>3</sup> En esta perspectiva se toma como referencia dichos cambios, se inicia un análisis hacia atrás de las diferentes causas que contribuyen al cambio, que pueden estar relacionadas con la intervención o no.



práctica emergente más que un pensamiento lineal «racional». Para AEVAL –como se verá más adelante–, el único criterio de valor no puede ser la eficacia (el cumplimiento de los objetivos y el logro de resultados), sino que, dada la diversidad de destinatarios y utilidad de las evaluaciones, se deben incorporar otros criterios y perspectivas.

Otro de los supuestos de este paradigma es la relevancia de los contextos y los actores, esenciales para abrir la «caja negra» y para determinar los mecanismos que producen el cambio, incorporando las intervenciones distintas relaciones políticas y sociales que involucran a numerosos *stakeholders*, hecho que sin duda afecta a los modelos de análisis. Al fin y a la postre las políticas y los programas se integran en procesos sociales y organizacionales de múltiples capas, cuya influencia debe ser tomada en cuenta, y donde los sistemas deben analizarse de forma abierta (Calidoni-Lundberg, 2006). Los contextos en los que se desenvuelven las políticas públicas tienen una notable influencia en la intervención. Esta visión, propia de la evaluación realista, está presente en AEVAL, aunque no es desde luego predominante ni alcanza todas sus potencialidades analíticas. En particular no se observa que sus evaluaciones indaguen en las cuestiones internas que operan en el seno de las intervenciones públicas, las transformaciones e interacciones producidas, ni las responsabilidades, poderes y potencialidad de los programas y los sujetos cuya conducta se pretende explicar (Pawson y Tilley, 1997).

Finalmente, la consideración de que las aproximaciones puramente positivistas –en particular los métodos de magnitudes (experimentales o cuasi experimentales)–, adolecen de limitaciones. Una de ellas es de orden metodológico y está relacionada con la dificultad de aprehender la realidad de forma objetiva a través de la observación y obtener mediciones precisas de los resultados de interés, obviando además otras aproximaciones válidas para dilucidar la atribución de los resultados y soslayar los análisis de contribución. En suma, buscar la solución a todos los desafíos conceptuales a través de la estandarización de la medición. Una búsqueda de la precisión que es, a la vez, dominante, seductora y equivocada, al menos desde la perspectiva constructivista (Patton, 2020, p. 2). Así, en el anexo II de la Guía de evaluación, al analizar las evaluaciones de impacto, si bien AEVAL destaca el carácter riguroso de estos métodos y su valor analítico para esclarecer los impactos causales obtenidos, apunta a que el «estudio de los impactos debe complementarse con otras aproximaciones de tipo cualitativo o modelos que permitan conocer otros aspectos determinantes en los impactos logrados» (AEVAL, 2015, p. 156), indagando sobre el contexto o el cómo, la distribución de los impactos o la dificultad de utilizar estos métodos en la evaluación de políticas amplias y sucesivas en el tiempo. La actitud de AEVAL es de prevención hacia posturas reduccionistas en las que la evaluación queda limitada a algunos elementos centrales, principalmente la intervención y los resultados: una vez implementado el programa, sólo queda analizar estos últimos. Así, el papel de la evaluación se estrecha, y su labor queda circunscrita a evaluar rigurosamente si el cambio produce los resultados predeterminados (Chen, 2015, p. 26).

Frente al predominio de las técnicas cuantitativas, propio de este paradigma, para AEVAL las distintas perspectivas teórico-científicas se consideran en un plano de igualdad, sin la preponderancia o hegemonía de ninguna de ellas. Este aspecto, denominado por AEVAL triangulación pluralista, remite a la triangulación teórica –distinta a la triangulación de métodos que se verá más adelante–. Bajo esta asunción los análisis cuantitativos y cualitativos deben combinarse en la evaluación, en un planteamiento «utilitarista» y pragmático que admite todo tipo de fuentes de información, técnicas y herramientas de recopilación y análisis de información, siempre y cuando sean pertinentes en el marco del diseño de la evaluación y se disponga de la información necesaria y los recursos para implementarlas.

Por otro lado, también se aprecia en AEVAL cierta presencia del método crítico-racional, en el que los valores y los fines sociales deben estar también presentes en la evaluación.

Una de las características que condicionó la metodología de AEVAL fue el enfrentarse a la evaluación no sólo de programas sino también de políticas complejas, en ocasiones con perfiles poco definidos, de notable amplitud o incluso a evaluaciones de objetos compuestos por todo un conjunto de acciones dispersas que obligaban a reconstruir una lógica de intervención escasamente perfeñada o inexistente<sup>4</sup>. Ello obligó a superar la tradicional tendencia proyecto-céntrica que ha caracterizado a la evaluación y a adoptar un enfoque más amplio.

---

<sup>4</sup> Un ejemplo de ello lo constituyó la Evaluación de la gestión y funcionamiento de las Confederaciones Hidrográficas (E17//2009), en la que el objetivo consistía en obtener una visión integrada del desempeño de las funciones y competencias de dichas confederaciones desde la perspectiva de su adecuación a los principios de la Directiva Marco del Agua. Esto es, poner en relación unos objetivos recientes con una estructura anteriormente existente y que carecía de competencias exclusivas para lograr esos objetivos (AEVAL, 2015).

Ante la multiplicidad de objetos de evaluación o los requerimientos dispares, AEVAL optó por una pluralidad de aproximaciones. En definitiva, dado que los métodos derivan al fin y a la postre del tipo de problemas que analizan, la opción adoptada fue pragmática, aceptando que ningún enfoque o teoría estaba en condiciones de dar respuestas por completo satisfactorias. Y que la evaluación debe ser adaptativa en el sentido de adoptar métodos acordes con el grado de desarrollo del objeto de evaluación, del tipo de evaluación que se realiza y de lo que de ella se demanda. En definitiva, no existió, para AEVAL, un único camino metodológico, por lo que era imprescindible desarrollar y combinar varias perspectivas y aceptar, sin ninguna ambigüedad, el pluralismo metodológico.

#### 4. EL CONTEXTO INSTITUCIONAL DE AEVAL Y SU INCIDENCIA EN SU METODOLOGÍA Y EL DISEÑO DE LAS EVALUACIONES

No cabe duda de que los factores institucionales condicionan numerosas decisiones en materia de evaluación. En el caso de AEVAL estos factores tuvieron, sin duda, un peso importante en la configuración de su metodología, en una Agencia que progresivamente fue alcanzando un mayor grado de madurez organizativa y experticia para acometer las funciones de evaluación y la misión que tenía encomendada. A continuación, se repasan brevemente estos factores y su influencia en la metodología AEVAL.

##### 4.1. Limitada práctica y cultura de evaluación

En España en general y en sus administraciones públicas en particular se carecía en los inicios de este siglo de un sistema de evaluación consolidado o cohesionado, más allá de una constelación de organizaciones –principalmente públicas–, con mandatos y prácticas en evaluación no interrelacionadas, con una cultura de evaluación que emergía lentamente, pero con diversidad de enfoques conceptuales, prácticas y mandatos (Feinstein y Zapico-Goni, 2010).

La progresiva implantación de la evaluación a nivel institucional fue reactiva, derivada de los requerimientos impuestos por la UE vinculados a la recepción de fondos comunitarios, así como también la pertenencia a otras organizaciones internacionales como la OCDE. El retraso en el despliegue de la evaluación pudo obedecer a factores de cultura político-administrativa y la lenta implantación de la Nueva Gestión Pública (Casado y Del Pino, 2021). A la prioridad en consolidar un sistema democrático que requería crear múltiples instituciones, estructuras, administraciones y servicios públicos, un tiempo más para «hacer y crear que, para reflexionar y analizar, y, por tanto, evaluar» (Bustelo, 2020, p. 305). Avances modestos y dispersos en materia de institucionalización que, además, coincidieron a partir de 2008 con una crisis económica, y donde la evaluación se desenvolvía un entorno público administrativo en el que quizás era considerada más como un coste que como una inversión (Ruiz, 2012).

La naturaleza de AEVAL y su carácter de organismo público insertado en el seno de una estructura gubernamental fuertemente burocratizada y jerarquizada condicionó sus relaciones con los distintos actores y destinatarios de las evaluaciones, en particular los gubernamentales, no ya sólo a nivel directivo o estratégico, sino también con los ministerios y unidades de gestión que actuaban como referentes en las distintas evaluaciones.

En este contexto, AEVAL adoptó algunas estrategias metodológicas destinadas a limitar las posibles resistencias durante el proceso de evaluación o la escasa receptividad ante las evidencias y las recomendaciones realizadas. Una de ellas fue pretender incorporar al enfoque y al alcance de la evaluación aquellos requerimientos y necesidades específicas manifestados por los destinatarios a nivel directivo y de gestión de los ministerios. En ocasiones como hipótesis de evaluación, o como apartados o bloques específicos de los informes<sup>5</sup>. AEVAL, al mismo tiempo que realizaba una evaluación, promovía la cultura evaluativa, potenciando la utilidad de la evaluación desde el punto de vista gerencial en los distintos niveles político-administrativos.

---

<sup>5</sup> Como ejemplos pueden citarse la Evaluación sobre la participación de la Administración General del Estado en el Sistema para la Autonomía y Atención a la dependencia (E13/2008), en la que una parte sustantiva de la evaluación lo constituyó el sistema informático del SISAAD, que, aun teniendo importancia a nivel de gestión y de los resultados, no era uno de los aspectos más críticos en el desarrollo del Sistema. Y la Evaluación de las medidas de racionalización y mejora de la gestión de la Incapacidad Temporal (E22/2009), que incorporó análisis particularizados relativos a los procesos de baja (en concreto si las bajas se producían los viernes o los lunes).

Asimismo, sus evaluaciones incorporaban abundantes componentes jurídico-administrativos, normativos y procedimentales muy alineados con las administraciones públicas. Este hecho no siempre fue acertado, al centrarse las evaluaciones en ocasiones tal vez en demasía en aspectos de gestión frente a otros elementos cruciales, dada la necesidad de acotar el alcance de la evaluación en el caso de objetos de evaluación complejos, transversales y políticas de gran amplitud.

## 4.2. Estrategia y planes de evaluaciones

Los planes de evaluación de AEVAL se ajustaron y adaptaron a los mandatos de evaluación anuales aprobados por el Consejo de Ministros –que hasta 2012 estuvieron vinculados al Programa Nacional de Reformas–, y que conformaban la parte fundamental de sus planes anuales.

En primer lugar, dichos mandatos contenían objetos de evaluación, objetivos evaluativos, tipos de evaluación y enfoques muy dispares. Este fue, sin duda, uno de los condicionantes institucionales de mayor calado para AEVAL. Encargos en los que existían, además, motivaciones, intereses y resistencias de distinto orden. Si bien AEVAL acordaba con los departamentos ministeriales las políticas o programas que se incorporaban al mandato del Consejo de Ministros, la capacidad de evaluar determinadas políticas o programas dependía en última instancia del nivel político.

Estos encargos motivaron la construcción de una metodología capaz de adaptarse a una enorme pluralidad de realidades evaluativas, que comprendiese la naturaleza política de las evaluaciones (Weiss, 1998) y los obstáculos y oportunidades que se desprendían de la posición institucional de AEVAL en cuanto a organismo de evaluación de la Administración General del Estado. Pero también representaron un factor que podía lastrar el diseño de sus evaluaciones y restringir su alcance: qué se iba a analizar y cómo.

Las evaluaciones de AEVAL se encomendaban por múltiples propósitos. Junto a la clásica evaluación sumativa para conocer el proceso y los resultados de los programas o políticas, otros encargos estaban encaminados a un análisis de intervenciones en funcionamiento para su mejora continua (*formative evaluation* o *constructive evaluation*)<sup>6</sup>. En otras ocasiones se encargaban para poner en valor un programa o mostrar sus méritos y resultados valiosos, como forma de fortalecer la posición de los decisores o gestores de determinadas áreas o incrementar su presupuesto. Otras evaluaciones podían nacer de la desconfianza de un decisor con respecto a la intervención que realmente se estaba desarrollando, su gestión, implementación y resultados, por lo que se acudía a una visión externa que contrastase la información transmitida a través de la estructura jerárquica.

Existen dos fases en las que las presiones políticas pueden afectar gravemente a la evaluación, tal y como certeramente destacó Chelimsky (2009, p. 54). Una de ellas es su diseño, momento en el que se puede intentar interferir un diseño sólido o incluso imponer uno inapropiado. El otro es la etapa final: los hallazgos de la evaluación, su lenguaje o los intentos para evitar su publicación. Aspectos que implican el riesgo de que un solo conjunto de *stakeholders* defina las necesidades de evaluación, introduzca sesgos y que ésta no sea útil para otros destinatarios de las evaluaciones en una sociedad plural. Ambos supusieron un reto para AEVAL, que en particular debió hacer frente en casos puntuales a resistencias al contenido final de las evaluaciones.

Si bien el alcance de la evaluación y el diseño en el caso de AEVAL se construía a través de un proceso de negociación con actores institucionales, principalmente niveles directivos de la AGE, era necesario evitar esta serie de condicionantes del diseño y el alcance de la evaluación. En particular que se intentase vedar la evaluación y análisis de determinados aspectos que pudiesen mostrar las limitaciones de la intervención; que se restringiese innecesariamente la evaluación cuando debía tener una perspectiva más amplia; o el intento de hacer prevalecer un método de análisis, como la demanda de evaluaciones «de impacto» cuando el grado de maduración del programa no lo permitía, puesto que la intervención se había aprobado recientemente y no había desplegado efectos. Este último condicionamiento no fue algo exclusivo de AEVAL, sino que suele producirse en el contexto general de las evaluaciones (Pawson, 2002).

Estos hechos obligaban a AEVAL a tener una cierta prevención y a esclarecer si, por ejemplo, ya se había adoptado una decisión relativa al programa, y por tanto el alcance o ciertas preguntas de evaluación constituirían un mero escaparate o se intentaba justificar ese cambio de la política o el programa.

<sup>6</sup> En AEVAL predominan las evaluaciones que entrarían en lo que H. T. Chen (2015) catalogaba como evaluaciones constructivas (*constructive evaluations*) bien de proceso o de resultado. Estas evaluaciones tienen como finalidad la mejora continua del programa. Proveen información sobre la debilidad o fortaleza relativa de la estructura de los programas y de los procesos de implementación (*constructive process evaluations*), o de cómo contribuyen a los resultados (*constructive outcome evaluations*).



Por otro lado, los planes de evaluación de AEVAL durante su segundo período (2012-2017) incorporaron distintos estudios orientados más bien a analizar cuestiones relativas al funcionamiento y organización de la propia administración o dar respuesta a necesidades muy específicas, actuando más como un organismo instrumental de análisis: es el caso del diagnóstico de situación y rediseño del SIA o de una metodología para caracterizar las funciones desarrolladas por los servicios centrales de la AGE.

La dependencia de los planes de evaluación de AEVAL de los mandatos del Consejo de Ministros y, en general, del impulso político-administrativo dado a la Agencia motivó que, si bien durante un primer período (2007-2011) se acometiesen hasta 33 evaluaciones con objetos muy diversos, su labor evaluadora en un segundo período (2012-2017) fuese menor. Además, los objetos de evaluación no se caracterizaron quizás por ser los más relevantes desde la perspectiva del gasto, su relevancia social o en relación a las principales problemáticas existentes en el seno de la sociedad española. En el año 2014 no hubo mandato del Consejo de Ministros.

**TABLA 1. PLANES DE EVALUACIÓN DE AEVAL EN SU SEGUNDA ETAPA (2012-2017)**

<b>Año 2013</b>	
E39/2013.	Manual para la racionalización y eliminación de duplicidades.
E38/2013.	Diagnóstico de situación y rediseño del Sistema de Información Administrativa.
E37/2013.	Diseño de una metodología para caracterizar las funciones desarrolladas por los servicios centrales de la AGE.
E36/2013.	Evaluación del Plan de Medidas para la Mejora de los Servicios de Sanidad Exterior.
E35/2013.	Evaluación de la formación dirigida a los empleados públicos de la Administración General del Estado.
E34/2013.	Plan de Ahorro y Eficiencia Energética 2008-2012.
<b>Año 2015</b>	
E40/2015.	Evaluación del Plan de Calidad de los Recursos Humanos de los Servicios Generales de la Administración de la Comunidad Autónoma de las Illes Balears.
<b>Año 2016</b>	
E43/2016.	Evaluación de la implementación del Programa Nacional de Desarrollo Rural 2014-2020.
E42/2016.	Metodología de evaluación y seguimiento de la transparencia de la actividad pública.
E41/2016.	Seguimiento de la formación dirigida a los empleados públicos: Administración General del Estado, Organizaciones Sindicales, Entidades Locales y Comunidades Autónomas. Años 2013-2014.

*Fuente:* Elaboración propia a partir de datos de AEVAL.

### 4.3. Capacidades en evaluación

El carácter de institución pública burocratizada de AEVAL incidió en distintas políticas que mediatizaban su labor evaluadora, como su política de recursos humanos, su presupuesto o los mecanismos de contratación.

En materia de personal, la relación de puestos de trabajo del Departamento de Evaluación de AEVAL estuvo compuesta casi exclusivamente por empleados públicos, salvo dos plazas. La plantilla efectiva del Departamento de Evaluación a lo largo de su existencia osciló entre 17 y 12 evaluadores, una cifra que puede considerarse modesta para responder a los objetivos asignados a la Agencia.

Esta circunstancia presentaba ventajas e inconvenientes. Entre las primeras cabe destacar el que los evaluadores fuesen profundamente conocedores de la realidad administrativa, procesos y ámbitos sectoriales en los que se desarrollaban las evaluaciones. Utilizaban, además, un lenguaje común con los gestores y niveles directivos de la AGE. El inconveniente era que la relación de puestos de trabajo restringía la incorporación de determinados perfiles. No existían prácticamente visiones ajenas a la administración, ni personal con una práctica y conocimientos dilatados en materia de evaluación. De hecho, a pesar de la pluralidad en cuanto a formación académica y trayectorias profesionales, en un principio existió cierto déficit en perfiles

vinculados al análisis de políticas públicas y en el uso de técnicas de análisis de datos. En todo caso AEVAL fue construyendo progresivamente capacidades en evaluación, no sólo desde la práctica, sino articulando mecanismos formativos especializados y nuevas incorporaciones que permitiesen incrementar los conocimientos en materia tanto de evaluación como en la utilización de técnicas de análisis. Estas carencias iniciales eran suplidas con el recurso a expertos externos en las distintas materias.

En cuanto a sus recursos económicos, el gasto en términos de obligaciones reconocidas fue reduciéndose desde los 5,2 millones de euros en 2008, a los 2,8 millones en 2017. No obstante, estas cifras absolutas pueden resultar engañosas, por cuanto entre 2009 y 2013 una partida importante del gasto fue el arrendamiento de su sede física.

Una de las características del proceso de evaluación de AEVAL era el carácter interno de las evaluaciones, realizadas por el propio personal de la Agencia con alguna excepción puntual. Para AEVAL todos los aspectos metodológicos de las evaluaciones debían realizarse en el seno del Departamento de Evaluación, puesto que era quien conocía el objetivo de la evaluación, sus preguntas, qué evidencias eran relevantes para responderlas y qué derroteros debían seguir las evaluaciones. Por ello se contrataban aspectos concretos, principalmente el uso de algunas técnicas o tratamiento puntual de datos, dentro de un entorno controlado. También se subcontrataban algunos procesos de las encuestas, que eran ampliamente utilizadas por AEVAL en sus evaluaciones, no sólo para captar percepciones, sino también para capturar datos. Era el caso principalmente del trabajo de campo y las primeras tabulaciones de resultados. Pero tanto el diseño de la investigación, como la confección de los cuestionarios, el muestreo y la explotación de los resultados de las encuestas se realizaba bien por AEVAL o de forma conjunta entre el personal de AEVAL y las empresas contratadas.

## 5. EL DISEÑO DE LAS EVALUACIONES AEVAL, SU OPERACIONALIZACIÓN Y LOS MÉTODOS DE ANÁLISIS

El diseño de las evaluaciones –entendido como el sistema mediante el que se va creando la estructura lógica de la evaluación, basado en una serie de interrogantes e hipótesis (Spiel et al., 2015)–, constituye una de las piezas centrales de todo el proceso de evaluación. El diseño de AEVAL estuvo caracterizado por varios aspectos cruciales.

El primero de ellos fue la importancia de definir con precisión y con un criterio realista el alcance de la evaluación y caracterizar el objeto: conocer y comprender la naturaleza de la intervención, su propósito, objetivos y ámbito en el que se desarrollaba. Al establecer el alcance, AEVAL prestaba atención a los objetivos de la evaluación, su justificación y potencial utilidad, incorporando las expectativas identificadas en el encargo de evaluación. El análisis del contexto jugaba un papel muy destacable. En definitiva, se trataba de establecer el campo, contornos y prioridades en las que debía centrarse la evaluación, para garantizar su calidad y que se focalizase en áreas de interés prioritarias, evitando además un gasto innecesario de los recursos evaluativos en áreas de interés marginales, la dispersión y el análisis de aspectos puramente tangenciales.

El segundo aspecto clave fue la forma en que se operacionalizó la evaluación, a través de preguntas y criterios de evaluación, plasmados en la matriz de evaluación. En este sentido AEVAL empleó algunos de los recursos tradicionales de esta disciplina, en los que el diseño se articula en torno a preguntas de evaluación consideradas como el elemento central que guiaba todo el proceso. La estructura de la matriz de preguntas de evaluación de la Agencia habitualmente se ajustó a una amplia pregunta coincidente con el criterio de evaluación que se pretendía indagar. Esta pregunta, a su vez, se descomponía en subpreguntas de evaluación con distinto grado de precisión en función tanto del alcance y el tipo de evaluación que se realizaba como de otros criterios. Las preguntas de evaluación abordaban entre otros aspectos, las hipótesis de la intervención o teoría del programa.

Las preguntas no eran, sin embargo, inalterables. A lo largo del análisis de la intervención surgían cuestiones no consideradas inicialmente que debían ser incluidas. Por otro lado, la constatación de la imposibilidad fáctica de disponer de datos provocaba su reformulación o eliminación. También se modificaban o suprimían subpreguntas cuando los propios hallazgos de la evaluación mostraban su irrelevancia. O al contrario, se incorporaban preguntas claramente pertinentes<sup>7</sup>.

<sup>7</sup> Así, en la Evaluación de medidas de racionalización de la incapacidad temporal (IT) (E22/2009), las evidencias apuntaban a una clara interconexión entre la incapacidad por contingencias comunes y por contingencias profesionales y a un posible trasvase o

Hablar de diseño de evaluación en AEVAL significa hablar de evaluación por criterios de evaluación, puesto que no sólo se configuraban como elemento metodológico básico, sino que la estructura de la parte central de sus informes –en su segunda etapa–, dedicada al análisis, interpretación y síntesis, se adecuaba a los criterios.

La evaluación por criterios –también denominada método de criterios preordenados–, adoptada por AEVAL tenía determinados rasgos que intentaban superar algunas de las críticas formuladas con respecto a esta estructura de evaluación. Junto a los criterios tradicionales en la evaluación (que para la Agencia eran la pertinencia, la coherencia interna, la coherencia externa y/o complementariedad, la implementación, la eficacia, eficiencia y sostenibilidad) se incorporaron otros vinculados a valores sociales tales como la equidad, la participación y la transparencia. O incluso se definían algunos criterios *ad hoc*, dadas las peculiaridades de la evaluación, como por ejemplo la apropiabilidad, en el caso de cambios administrativos o de cultura que debían producirse en las estructuras administrativas como consecuencia de la aprobación de leyes o normas que tenían precisamente ese objetivo<sup>8</sup>.

Para AEVAL los criterios de evaluación eran áreas de análisis de gran amplitud, que en sí mismas no tenían que limitarse a un parámetro concreto, y donde el estándar debía caracterizarse por la multidimensionalidad y complejidad de las políticas públicas. Por ejemplo, el criterio de eficacia no se reducía al logro de los objetivos previstos (que podían además ser distintos a los declarados oficialmente) sino que abarcaba el grado en que el problema, demanda o necesidad que motivó la intervención era resuelto, o el análisis de los posibles efectos no buscados.

Algunos criterios tenían notable importancia para AEVAL, al adoptar una perspectiva de evaluación integral de todo el ciclo de la intervención pública. Entre ellos figuraban en particular, como se desprende de su Guía de Evaluación, los relacionados con el diseño y la identificación del problema. Tal es el caso del criterio de «pertinencia» entendido como el grado en que el conjunto de medidas que componen una intervención están orientadas a resolver la necesidad o problemática existente en el contexto en el que ésta se produce. O el grado en que ésta es congruente con las necesidades existentes. También fue importante en sus evaluaciones el criterio de coherencia interna, por cuanto comprendía gran parte de las cuestiones vinculadas al diseño de las intervenciones. Y asimismo el criterio de coherencia externa o complementariedad. La tabla 2 contiene los criterios de evaluación utilizados por AEVAL.

**TABLA 2. CRITERIOS DE EVALUACIÓN UTILIZADOS POR AEVAL**

Tradicionales	Otros criterios
Pertinencia	Equidad
Relevancia	Participación
Coherencia interna	Transparencia
Coherencia externa	Criterios <i>ad hoc</i>
Complementariedad	
Implementación	
Cobertura	
Eficacia	
Eficiencia	
Sostenibilidad	

*Fuente:* Elaboración propia a partir de AEVAL (2015).

deficiente selección de los procesos de IT, con claros efectos en la protección de los trabajadores y con alta incidencia en el sistema sanitario. Esto motivó la incorporación de algunas preguntas que indagasen estas cuestiones.

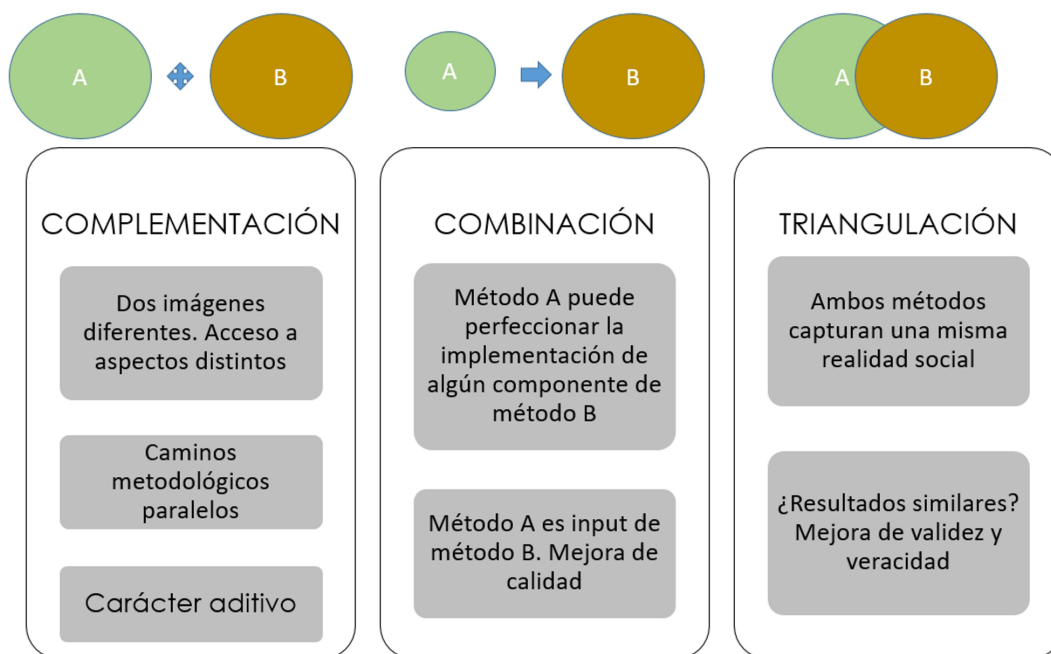
<sup>8</sup> Ejemplos de criterios *ad hoc* pueden observarse en la Evaluación de los programas de formación de la cooperación española al desarrollo (E23/2009) y la Evaluación de la gestión y funcionamiento de las demarcaciones de costas para la protección del Dominio Público Marítimo Terrestre AEVAL (E28/2010).

La tercera cuestión importante en el diseño de evaluación era evitar el predominio de un método, técnica o herramienta de investigación. La evaluación sigue un proceso lógico-racional, en el que debe predominar qué se quiere conocer y cómo se descompone el objeto de conocimiento en sus distintas dimensiones. Los métodos o herramientas concretas deben ocupar un lugar secundario, puesto que representan las «flechas en el carcaj del evaluador» (Cook et al., 2010). AEVAL abogó por evitar la evaluación condicionada por el método y los sesgos que conlleva.

En este plano AEVAL defendió cuatro estrategias, a saber: los análisis de atribución, contribución, análisis de cambio y el meta-análisis. El último de ellos en realidad no estuvo muy presente en la generalidad de sus evaluaciones, más allá de evaluaciones concretas que mostraron las evidencias que otros estudios y evaluaciones sobre la materia objeto de análisis<sup>9</sup>. Y ello a pesar del gran valor que tienen tanto las metaevaluaciones como los meta-análisis, desde la perspectiva de las políticas basadas en la evidencia, puesto que referencian resultados concretos de las intervenciones diseñadas, bien desde el punto de vista numérico, bien desde el punto de vista narrativo.

Para AEVAL resultaba fundamental recurrir a la triangulación u operacionalización múltiple, que remite a la conveniencia de emplear más de un método concreto al objeto de estudio, de tal forma que se garantice la validez de las evidencias encontradas. Desde el punto de vista de las herramientas de análisis ello implica tanto la triangulación de datos (utilizar distintas fuentes de datos para corroborar la información) como fundamentalmente la triangulación entre métodos: utilizar distintas herramientas de análisis con el objetivo de superar o neutralizar las limitaciones inherentes a cada uno de ellos.

GRÁFICO 2. COMPLEMENTACIÓN, COMBINACIÓN Y TRIANGULACIÓN



Fuente: Elaboración propia a partir de Bericat (1998).

Esta multiplicidad de técnicas o aproximaciones se observa en sus evaluaciones. Los métodos cuantitativos estuvieron presentes en muchas de ellas. En particular cabría destacar, en cuanto a evaluaciones que prestaron particular atención a la atribución, la Evaluación de la política de bonificaciones y reducciones de cuotas de la Seguridad Social destinada a las personas en situación de discapacidad (E19/2009), en la que se utilizó el *Propensity Score Matching* junto con otras técnicas; este mismo método se empleó también en la Evaluación de la formación dirigida a los empleados públicos de la Administración General del Estado (E35/2013), que utilizó además distintas regresiones logísticas y lineales para realizar un amplio análisis no

<sup>9</sup> Un ejemplo puede encontrarse en la Evaluación sobre la política de bonificación y reducción de cuotas de la Seguridad Social (E14/2008).

sólo de los efectos de la formación, sino de cómo se incardina con otros factores en relación al desempeño de los empleados públicos.

Por su parte la Evaluación de las medidas de racionalización y mejora de la gestión de la Incapacidad Temporal (E22/2009) recurrió a todo tipo de regresiones y técnicas de clasificación no sólo para analizar los resultados de alguna de las medidas, sino para caracterizar adecuadamente el problema, la prevalencia de la incapacidad temporal por contingencias comunes y contrastar determinadas hipótesis de evaluación. La utilización de técnicas de comparación con grupos controlados o similares, que permiten indagar también sobre los potenciales efectos netos se aprecia en la Evaluación de las acciones financiadas con cargo a los presupuestos generales del Estado en las áreas de influencia socioeconómica de la red de Parques Nacionales (E15/2008) o la Evaluación del Programa de Ayudas para Actuaciones de Reindustrialización (REINDUS) (E27/2010). La evaluación del Plan Estratégico de Seguridad Vial 2005-2008 (E16/2009) –que utilizó la teoría del programa (Theory Based Evaluation) para analizar la lógica de la intervención y las hipótesis causales–, contenía una modelización ARIMA que estudió tanto el impacto de algunas variables en las series temporales de víctimas de tráfico, como el impacto de las dos reformas legislativas más trascendentes.

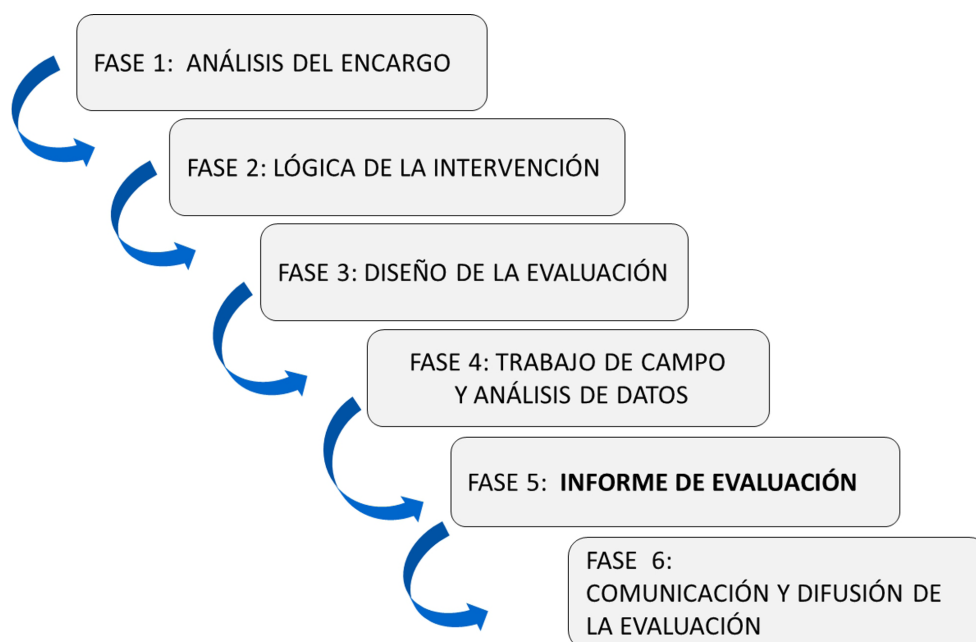
Un caso peculiar lo ofrece la evaluación de las trabas administrativas para la creación de empresas (E31/2011), en la que se utilizó la técnica de cliente misterioso, que permitió tener una imagen precisa y real del coste de la creación de empresas en términos monetarios, de tiempo y procedimientos con respecto a unos tipos concretos de empresas<sup>10</sup>.

El recurso a técnicas cualitativas de todo tipo estuvo muy presente en la práctica totalidad de evaluaciones. Para AEVAL la perspectiva cualitativa operaba en un plano de igualdad junto a la cuantitativa. Y ello por cuanto a veces el aspecto crucial en la evaluación no es la validez interna, prioritaria para las técnicas experimentales o cuasi experimentales, sino la validez externa, puesto que aborda aspectos sobre la generalización de los efectos encontrados y aporta a los decisores información muy útil sobre factores relevantes que puede apoyar sus futuras decisiones.

## 6. EL PROCESO DE EVALUACIÓN EN AEVAL

El proceso de evaluación en AEVAL se realizaba siguiendo la secuencia que se refleja en el gráfico 3.

**GRÁFICO 3. PROCESO DE EVALUACIÓN AEVAL. FASES**



*Fuente:* Elaboración propia a partir de AEVAL (2015).

<sup>10</sup> También se empleó esta técnica en la evaluación de la calidad del servicio de los museos de titularidad estatal (E08/2007).



La primera fase de análisis del encargo permitía realizar una primera aproximación al objeto de evaluación. Contenía la caracterización del objeto de evaluación y la delimitación o alcance de ésta, para la que se utilizaban distintos criterios tales como la temporalidad, relevancia de las medidas (en el caso de que fuesen múltiples se priorizaban aquellas que tenían un mayor impacto en la consecución de los objetivos) o la territorialidad, dada la naturaleza de Estado compuesto de España. Asimismo, el análisis del encargo realizaba un avance del enfoque de la evaluación, de la metodología y las herramientas a utilizar en la evaluación. Y finalmente la estrategia organizativa.

En esta fase se trazaba un primer mapa de actores que se iba completando a lo largo de la evaluación. Además, contenía un análisis de evaluabilidad, aunque no de modo formal y explícito.

Tras la fase de análisis del encargo se definía la composición del equipo de evaluación, en cuya selección se utilizaban dos criterios: la cualificación y la multidisciplinariedad. También se designaba el responsable de la evaluación.

La segunda fase consistía en el análisis y reconstrucción de la lógica de la intervención. Se situaba el foco en la génesis de la intervención, reconstruyéndola y analizando el contexto en el que surgía y se desarrollaba, se identificaba y definía el problema, sus causas y efectos, los instrumentos de intervención pública existentes para resolver el problema, junto con la teoría e hipótesis de la intervención y finalmente el despliegue de la intervención.

La tercera fase comprendía el diseño de la evaluación propiamente dicho: se establecían las preguntas, hipótesis, criterios y matriz de evaluación, se seleccionaban las herramientas y técnicas necesarias, se planificaba la evaluación y se establecían los posibles recursos externos y contrataciones necesarias para la culminación de la evaluación.

La cuarta fase correspondía a la recopilación y análisis de la información y en ella se pretendía cubrir las necesidades de información de la matriz de preguntas de evaluación.

En una quinta fase se daba forma definitiva al informe de evaluación, puesto que realmente se iba construyendo a lo largo de todo el proceso de evaluación, práctica que fue también fruto de la experiencia adquirida. Finalmente se procedía a la difusión y comunicación de la evaluación.

En el proceso de evaluación de AEVAL predominaba el pragmatismo, la necesidad de modular correctamente tanto el alcance de las evaluaciones como la planificación ordenada de las actividades que debían realizarse, aspecto sin duda sustantivo en unas evaluaciones siempre condicionadas por los recursos de tiempo, de personal y económicos, y con fechas de entrega establecidas. Con el trascurso del tiempo se fue implementando en AEVAL un proceso por el que, una vez finalizada cada una de las etapas, se elaboraba un informe de fase que permitía a la dirección del Departamento de Evaluación y de la propia Agencia conocer con detalle el estado en el que se encontraba la evaluación. Estos informes parciales se integraban en el informe final de evaluación, una vez revisados.

## 7. FORTALEZAS Y DEBILIDADES DEL MODELO AEVAL

El modelo AEVAL presenta, sin duda, fortalezas y debilidades. Uno de los mayores logros de AEVAL fue el haber sido capaz de diseñar en un plazo de tiempo relativamente corto una metodología de evaluación consistente, compleja, plural, adaptable a los distintos objetos, integral y, sin duda, pragmática.

Uno de los aspectos del modelo AEVAL no suficientemente ponderado era su carácter de agencia de evaluación generalista, cuyas evaluaciones no se circunscribían a un sector de intervención pública o políticas concretas, sino que abarcó la práctica totalidad de sectores. Con objetos de evaluación muy dispares. En la que el tipo de intervención a evaluar podía consistir en una política pública, cualquier instrumento de despliegue (un plan, un programa, una ley o cualquier otra medida). Se evaluaban intervenciones situadas en cualquier fase del ciclo: diseño, implementación, finalizadas o continuas. Desde múltiples perspectivas o centrándose en aspectos muy delimitados, acometiendo evaluaciones *ex ante* (las menos) de implementación o proceso (las más) y de resultados (limitadas). En definitiva, AEVAL no podía y no debía apostar por un tipo de evaluación específico, como, por ejemplo, la evaluación de resultados. Esto permitió que la Agencia acumulara una experiencia evaluativa de indudable riqueza.

Fruto de este modelo y en el plazo de tiempo que existió se realizaron un importante número de evaluaciones, treinta y nueve, que podrían considerarse como «verdaderas» utilizando la terminología de Stufflebeam y Sinkfield (1995), esto es, procesos evaluativos que realmente se corresponden con el concepto de

evaluación de políticas públicas<sup>11</sup>, lo que no es un escaso bagaje, hecho absolutamente novedoso en España. Más aún si cabe si se toma en consideración la amplitud de su perspectiva o de los objetos de evaluación y los limitados recursos disponibles. Su metodología, por ende, fue utilizada ampliamente en distintos ámbitos no sólo administrativos y de decisión pública, sino también en ámbitos académicos.

Empero, el modelo AEVAL presentaba algunas debilidades, tal y como se desprende del análisis realizado sobre una muestra representativa de sus evaluaciones y las opiniones expresadas por distintos responsables e integrantes de la Agencia. Una de ellas era, sin duda, que la perspectiva plural y amplia representada por su modelo no siempre se llevó a la práctica tal y como estaba formulado. Sus evaluaciones fueron muy heterogéneas en su objeto y perspectivas, y en ocasiones se obvió la profundidad analítica postulada en sus documentos metodológicos.

El análisis de las evaluaciones seleccionadas en este estudio muestra que las evaluaciones estaban en ocasiones excesivamente centradas en los aspectos de gestión procedimental, las normas legales que articulaban las intervenciones y los aspectos administrativos y burocráticos. Esta visión tan centrada en las estructuras en las que se insertaba AEVAL, y a las que pertenecía gran parte de su personal, pudo obviar o restringió otras visiones y planteamientos más amplios que hubiesen aportado una mayor riqueza a algunas de sus evaluaciones. Si bien es cierto que los gestores y decisores eran uno de sus principales destinatarios, no eran los únicos, y por ello las evaluaciones de la Agencia debían responder a esa pluralidad de destinatarios.

Por otro lado, y derivado de lo anterior, se observa que la aproximación realizada en algunas evaluaciones estuvo excesivamente centrada en la perspectiva operacional de los programas (su puesta en práctica) e incluso existió un predominio de lo descriptivo. Y, además, con unos análisis de implementación focalizados en un enfoque *top-down*, esto es, de arriba a abajo, centrado en las decisiones de las autoridades, en la regularidad o conformidad legal del proceso y no un enfoque o visión *bottom up*, que defiende que debe analizarse también el grado de participación de los actores, el nivel de conflicto o las actividades en red.

En algunas de las evaluaciones analizadas la perspectiva de los efectos y resultados pudo quizás no estar adecuadamente resuelta, o al menos con toda la profundidad que sería deseable. El análisis de los resultados, con no ser el único aspecto relevante de la metodología AEVAL, no encontró un adecuado acomodo en aquellas evaluaciones excesivamente centradas en los aspectos apuntados más arriba. Tampoco se abordó en otros casos destacados elementos del diseño de las intervenciones que podían ser aspectos críticos de los programas o políticas sometidos a evaluación.

Finalmente se aprecia que determinadas evaluaciones operaban en un ámbito de investigación evaluativa estrecho y con un planteamiento excesivamente lineal, siguiendo el modelo estándar. Y por otro lado en algunas de ellas se observa un déficit del análisis de las interacciones del programa con el entorno y el contexto en que se insertaban.

Estas debilidades no son, en todo caso, generalizables a la totalidad de sus evaluaciones, y en parte son debidas a distintos factores, como el proceso de formación de la metodología a lo largo de varios años; el hecho de que fuese una Agencia de nueva creación; o a las características, perfiles y formación del personal que las llevaba a cabo. Cuando hablamos de evaluación de políticas públicas tendemos a centrarnos en las grandes cuestiones: el método, el enfoque, el proceso, los grandes propósitos de la evaluación. Y obviamos algunas cuestiones básicas. Una de ellas es la necesidad de fortalecer las capacidades en evaluación, hecho de particular trascendencia en organismos o unidades de evaluación. Capacidades que, como otros aspectos, AEVAL fue desarrollando progresivamente.

## REFERENCIAS BIBLIOGRÁFICAS

AEVAL (2009). *La función evaluadora: principios orientadores y directrices de actuación de la evaluación de políticas y programas*. Ministerio de la Presidencia. <https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Evaluacion.pdf#page=1>

AEVAL (2010). *Fundamentos de evaluación de políticas públicas*. Ministerio de Política Territorial y Administración Pública. <https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia0.pdf#page=1>

<sup>11</sup> Son aquellos procesos evaluativos que producen evidencia científica a partir de la investigación sistemática, tiene por objeto intervenciones públicas identificables, incorporan un juicio de valor sobre la base de criterios de valor y proporcionan orientaciones prácticas en relación con la intervención evaluada.

- AEVAL (2015). *Guía práctica para el diseño y la realización de evaluaciones de políticas públicas*. Ministerio de Hacienda y Administraciones Públicas. [http://www.aeval.es/es/difusion\\_y\\_comunicacion/publicaciones/Guias/Guias\\_Evaluacion\\_Politicas\\_Publicas\\_y\\_Calidad\\_SSPP/Guia\\_evaluaciones\\_AEVAL.html](http://www.aeval.es/es/difusion_y_comunicacion/publicaciones/Guias/Guias_Evaluacion_Politicas_Publicas_y_Calidad_SSPP/Guia_evaluaciones_AEVAL.html)
- Belcher, B. y Palenberg, M. (2018). Outcomes and Impacts of Development Interventions: Toward Conceptual Clarity. *American Journal of Evaluation*, 39(4), 478-495. <https://doi.org/10.1177/1098214018765698>
- Bericat, E. (1998). *La integración de los métodos cuantitativo y cualitativo en la investigación social: significado y medida*. Ariel.
- Bustelo, M. (2020). Spain. En R. Stockmann, W. Meyer y L. Taube (eds.), *The Institutionalisation of Evaluation in Europe* (pp. 303-327). Palgrave Macmillan. [https://doi.org/10.1007/978-3-030-32284-7\\_12](https://doi.org/10.1007/978-3-030-32284-7_12)
- Calidoni-Lundberg, F. (2006). *Evaluation: definitions, methods and models. An ITPS framework* [Working Paper, R 2006: 002]. Swedish Institute for Growth Policy Studies.
- Caracelli, V. J. (2000). Evaluation use at the threshold of the twenty-first century. *New Directions for Evaluation*, 88, 99-111. <https://doi.org/10.1002/ev.1194>
- Casado, J. M. y Del Pino, E. (2021). Evolución, situación actual y retos de la evaluación de políticas públicas en las Administraciones españolas (2000-2021). *Cuadernos Económicos de ICE*, 102, 13-38. <https://doi.org/10.32796/cice.2021.102.7308>
- Chelmsky, E. (2009). Integrating Evaluation Units into the Political Environment of Government: The Role of Evaluation Policy. En W. M. Trochim, M. M. Mark y L. J. Cooks (eds.), *Evaluation policy and evaluation practice* (pp. 51-66). New Directions for Evaluation.
- Chen, T. H. (2015). *Practical Program Evaluation Theory-Driven Evaluation and the Integrated Evaluation Perspective* (2.ª ed.). SAGE Publications Ltd.
- Cook, T. D., Scriven, M., Coryn, C. L. S. y Evergreen, S. D. H. (2010). Contemporary Thinking About Causation in Evaluation: A Dialogue With Tom Cook and Michael Scriven. *American Journal of Evaluation*, 31(1) 105-117. <https://doi.org/10.1177/1098214009354918>
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3.ª ed.). SAGE Publications Ltd.
- Feinstein, O. y Zapico-Goni, E. (2010). *Evaluation of Government Performance and Public Policies in Spain* [ECD Working Paper Series, 22]. World Bank. <https://openknowledge.worldbank.org/handle/10986/27913>
- Instituto para la Evaluación de Políticas Públicas (2020). *Guía de evaluación de resultados de políticas públicas*. Ministerio de Política Territorial y Función Pública. [https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia\\_de\\_Evaluacion\\_de\\_Resultados.pdf#page=1](https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia_de_Evaluacion_de_Resultados.pdf#page=1)
- Patton, M. Q. (2020). Evaluation Use Theory, Practice, and Future Research: Reflections on the Alkin and King AJE Series. *American Journal of Evaluation*, 41(4), 581-602. <https://doi.org/10.1177/1098214020919498>
- Pawson, R. (2002). Evidence-based Policy: In Search of a Method. *Evaluation*, 8(2), 157-181. <https://doi.org/10.1177/1358902002008002512>
- Pawson, R. y Tilley, N. (1997). *Realistic Evaluation* (1.ª ed.). SAGE Publications Ltd.
- Rossi, P. H., Lipsey, M. W. y Freeman, H. E. (2003). *Evaluation: A Systematic Approach* (7.ª ed.). SAGE Publications, Inc.
- Ruiz Martínez, A. (2012). Panorámica actual de la evaluación de las políticas públicas. *Presupuesto y Gasto Público*, 68, 13-23. [https://www.ief.es/docs/destacados/publicaciones/revistas/pgp/68\\_01.pdf](https://www.ief.es/docs/destacados/publicaciones/revistas/pgp/68_01.pdf)
- Stake, R. E. (2005). *Investigación con estudio de casos*. Morata.
- Stufflebeam, D. L., y Shinkfield, A. J. (1995). *Evaluación sistemática: Guía teórica y práctica*. Ministerio de Educación y Ciencia. Paidós.
- Spiel, C., Schober, B., y Bergsmann, E. (2015). Program Evaluation. En J. D. Wright (ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2.ª ed., pp. 117-122). Elsevier Ltd.
- Weiss, C. H. (1983). The stakeholder approach to evaluation: Origins and promise. *New Directions for Program Evaluation*, 17, 3-14. <https://doi.org/10.1002/ev.1322>
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2.ª ed.). Prentice Hall.
- Yin, R. K. (2009). *Case study research: Design and methods* (4.ª ed.). SAGE Publications Ltd.