

# ¿Cómo puede contribuir el proceso de creación de códigos normativos al desarrollo de una cultura de evaluación? Elementos para el debate en el caso español

---

María Bustelo Ruesta\*

**Palabras clave:** *Estándares y orientaciones de evaluación, códigos normativos, institucionalización de la evaluación, metaevaluación, ética y evaluación.*

## 1. Introducción<sup>1</sup>

Las normas, estándares, guías orientativas y/o códigos en evaluación, que comienzan a surgir en la década de los años ochenta y noventa en algunos países, cumplen una doble función. Por un lado, orientan a los profesionales sobre la práctica profesional y son, además, una guía de orientación ética. Por otro, son un indicador de la institucionalización de la función de evaluación en un determinado país y, en parte, de la consolidación de la evaluación como ámbito académico y profesional. Aunque, como se verá más adelante, también se hará referencia a la primera función de orientación ética, es en la segunda función de consolidación e institucionalización de la evaluación donde se va a concentrar este artículo.

La consolidación profesional de una disciplina se puede medir a través de una serie de factores tales como el desarrollo de programas académicos superiores dedicados a la formación de profesionales (que además conllevan la elaboración de manuales, libros de texto, lecturas y casos prácticos), la edición de publicaciones periódicas especializadas y la existencia de asociaciones profesionales, además, lógicamente, de la presencia de un mercado

---

\* Profesora de Ciencia Política y de la Administración, Universidad Complutense de Madrid.

<sup>1</sup> Una primera versión de este artículo, en forma de ponencia y con el título «El potencial impacto del establecimiento de normas, estándares y códigos en la creación de una cultura de evaluación», se presentó en el IX Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública (Madrid, noviembre 2004).

de trabajo y la existencia de profesionales que son contratados como tales. En esta misma línea, la elaboración de códigos de conducta o listados de normas, estándares o principios éticos puede ser también no sólo un indicador importante, sino también un elemento generador y provocador de una mayor madurez y consolidación institucional y profesional de una disciplina. Profesiones fuertemente consolidadas desde hace tiempo (como son la medicina, la abogacía o la arquitectura) o más recientemente (como la psicología) han desarrollado este tipo de códigos. Además de fomentar un sentido de pertenencia e identificación entre los profesionales, acotan y regulan la práctica profesional entre sus respectivos campos. En el caso de la evaluación de programas y políticas públicas, los países con una función de evaluación más desarrollada y consolidada, en las que se han dado un conjunto de los factores anteriormente mencionados (programas académicos específicos, publicaciones especializadas, mercado de trabajo y, sobre todo, asociaciones profesionales), son los que han desarrollado este tipo de códigos principalmente en la última década.

Estos códigos en el campo de la evaluación están normalmente vinculados a las asociaciones profesionales de evaluadores, o bien a organizaciones supranacionales que desarrollan la evaluación en diferentes ámbitos, como pueden ser la Comisión Europea o la OCDE<sup>2</sup>. Los primeros códigos en elaborarse en el campo de la evaluación de programas y políticas fueron en el ámbito anglosajón. El primero y más importante de todos, las normas del *Joint Committee*, en su primera edición en 1981 y su segunda en 1994 (*Joint Committee* 1981 y 1994). Luego le siguieron los códigos de la Asociación Americana de Evaluación (1994); la Asociación Canadiense de Evaluación (1996); y la Asociación Australiana de Evaluación (1997)<sup>3</sup>. También se aprobaron en 2001 unas orientaciones de evaluación por parte de la Asociación Africana de Evaluación, que son una adaptación a las normas del *Joint Committee* y en las modifican aproximadamente un 40% de las normas previamente existentes<sup>4</sup>.

En el ámbito europeo, se han aprobado códigos u orientaciones de las Sociedades de Evaluación francesa (2003), alemana (2001), suiza (2000) y británica (Beywl y Speer 2004; Widmer 2004)<sup>5</sup>. Mientras los modelos suizo y alemán siguen la estructura de las normas del *Joint Committee*, el francés está formulado en formato de grandes principios éticos (pluralismo, independencia, competencia, respeto, transparencia y responsabilidad), y el inglés hace una interesante separación de normas según vayan dirigidas a evaluadores, comanditarios, participantes, o implicados en una auto-evaluación. Las normas de entidades supranacionales, como son las de la OCDE fueron revisadas en 1998, al igual que las de la Comisión Europea que fueron revisadas en 2000. En el caso de la OCDE, las orientaciones tienen el propósito de mejorar el uso de las evaluaciones que se realizan en el ámbito de la cooperación internacional y la revisión de 2000 de la Comisión Europea se hace específicamente para mejorar la comunicación externa de los resultados de las evaluaciones realizadas en el marco de la Comisión. Asimismo, existen también códigos que están vinculados a determinados sectores. Como ejemplos, se pueden mencionar las orientaciones para evaluar programas de prevención de drogodependencias, elaboradas por el Observatorio Europeo de Drogodependencias, y los principios éticos para evaluar programas de cooperación al desarrollo, elaborados por la NORAD (Agencia Noruega de Cooperación)<sup>6</sup>.

<sup>2</sup> Véase Comisión Europea (2000).

<sup>3</sup> Normas (*Standards*) elaboradas por el *Joint Committee on Standards for Educational Evaluation* (1981 1.ª ed., y 1994 2.ª ed.); Principios Orientadores para Evaluadores (*Guiding Principles for Evaluators*) de la Asociación Americana de Evaluación (adoptados y editados por la asociación en 1994, revisados y ratificados en julio de 2004); Orientaciones para una actuación ética (*Guidelines for ethical conduct*) de la Sociedad Canadiense de Evaluación (1996); Orientaciones para una actuación ética en evaluaciones (*Guidelines on Ethical Conduct of Evaluation*) de la Sociedad Australasiática de Evaluación (1997), que completan con un código ético de la misma sociedad aprobado en 2000.

<sup>4</sup> Véase (<http://www.afrea.org>). También Love y Russon (2004) y Rouge (2004).

<sup>5</sup> Véanse las páginas web de la Sociedad Francesa de Evaluación (<http://www.sfe.asso.fr>), la Sociedad Suiza de Evaluación (<http://www.seval.ch>), la Sociedad Alemana de Evaluación (<http://www.degeval.de>) y la Sociedad Británica de Evaluación (<http://www.evaluation.org.uk>).

<sup>6</sup> Véase EMCDDA (1998) y NORAD (1997).

## 2. Algunos elementos básicos a tener en cuenta respecto a los códigos

Teniendo en cuenta en qué consisten los códigos, una primera cuestión a resaltar es que pueden estar compuestos por diferentes tipos de normas: los principios (*principles*) se suelen referir a cómo deben comportarse los evaluadores, los estándares (*standards*) explicitan qué atributos debe tener y cumplir una buena evaluación, y las orientaciones (*guidelines*) formulan recetas para el cumplimiento de estándares y principios, es decir se refieren al cómo hacer para cumplirlos. Por lo tanto, las normas como las del *Joint Committee* serían estándares y las de las asociaciones norteamericana y francesa serían principios. Las orientaciones se darían prácticamente, de forma más o menos explícita, en todos los casos. Otra de las cuestiones importantes es la de si se trata de normas mínimas (que hacen referencia a qué se debe cumplir cómo mínimo para una evaluación) o normas máximas (cómo debería ser una buena evaluación). Todos los códigos mencionados consisten en normas máximas.

Respecto al grado de especificidad y al contenido de los códigos o normas, se puede decir que resulta casi inevitable que exista una cierta tensión entre la necesidad de que los códigos sean relevantes para el contexto específico en el que deben ser utilizados, y la de tener unas normas válidas para todo el mundo y en diferentes tipos de situaciones. Quizá esta tensión puede ser resuelta, por un lado, elaborando códigos nacionales que tengan en cuenta la variedad de situaciones que pueden darse en sistemas político-organizativos específicos. Por otro, se deben tener en cuenta las prácticas reales de evaluación en el contexto concreto; de ahí la importancia de realizar análisis meta-evaluativos previos a dicha elaboración.

Otra de las cuestiones importantes a tener en cuenta es a quién o quiénes van dirigidos los códigos. Algunos, como el propio del *Joint Committee* y, por lo tanto, también los que han tomado dicho código como referencia, están muy centrados en la figura del evaluador/a, y ésta es una de sus mayores debilidades. El de la OCDE, como es lógico, está más dirigido a los países miembros y a los equipos que, de cada país, generan los informes de evaluación. Desde luego, parece interesante pensar en la comunidad de evaluación de forma más amplia, como es el caso de la Sociedad Británica de Evaluación, en el que la manera de organizar sus orientaciones ha sido según a quién iban dirigidas: evaluadores, comanditarios, participantes e implicados en auto-evaluaciones. A efectos de crear y generar cultura de evaluación, tan importante es promover una buena conducta por parte de quienes hacen las evaluaciones, como por parte de quienes las encargan y participan en las mismas desde diferentes perspectivas y posiciones.

Otra cuestión de suma importancia, pero muchas veces más delicada, es la de la aplicación de los códigos. Es evidente que cualquier código debe tener una notable dosis de autoridad profesional y de credibilidad en la comunidad de evaluación, dado que no suelen tener ningún poder sancionador. De ahí la importancia de que la elaboración de los códigos sea liderada por parte de una instancia cualificada y con una dosis grande de legitimidad y credibilidad en la comunidad evaluadora. También es importante que el proceso de elaboración y revisión incluya la participación de las diferentes personas implicadas en la práctica evaluativa y que se base en las realidades del contexto concreto en el que se van a aplicar las normas. Tiene sentido, pues, que la elaboración de los códigos se lidere por parte de las asociaciones o sociedades nacionales, siempre y cuando éstas representen al conjunto de las personas implicadas profesional y académicamente en la realidad de la práctica evaluativa de un país, en los diferentes sectores.

### 3. La función de orientación ética de los códigos

Una primera función que cumplen los códigos deontológicos<sup>7</sup> es que permiten la introducción y desarrollo de los temas éticos, que resultan especialmente relevantes en la práctica de la evaluación. No se trata sólo de elaborar teóricamente sobre estos asuntos, sino también de orientar en la práctica en los conflictos y dilemas éticos que se van a producir, inevitablemente y dado el contexto político en el que está inmersa la evaluación, en cualquier situación evaluativa. Esto se ve de forma explícita y clara en los códigos elaborados. Por ejemplo, el *Joint Commmittee*, además de plantearse, con la elaboración de sus Normas, orientar el diseño, empleo y crítica de las evaluaciones y de animar a la utilización de una variedad de métodos en evaluación, pretende ayudar a los evaluadores a identificar y afrontar la realidad política. Textualmente este comité afirma que «las agendas políticas y el dinero son fuentes de poder que pueden corromper la evaluación en cualquier circunstancia. La adhesión a las Normas ayudará a minimizar las posibilidades de una mala utilización de ese poder por parte de los evaluadores o sus clientes» (*Joint Committee* 1994: 4).

Sin embargo, cuando los códigos tienen un alto contenido en temas éticos o, en otras palabras, dan especial prioridad a esta primera función, como es el caso de los Principios Orientadores de la Asociación Americana de Evaluación, se suelen formular con un alto grado de generalidad. Esta circunstancia, que es en cierto modo inevitable e incluso deseable para evitar la rigidez, es una de las mayores debilidades de los códigos. Una de las consecuencias de dicha generalidad es que puede parecer que estos códigos sean una forma de regular lo obvio. Ciertamente, los Principios Orientadores de la AEA podrían ser aplicables a cualquier práctica profesional y están formulados de una forma muy general<sup>8</sup>. Esta generalidad de los principios de la AEA también tiene que ver con el afán de ser comprehensivos, es decir que todo el mundo pueda verse incluido y reflejado. Así, podríamos decir que, en nuestros días lo «políticamente correcto» consiste en ser flexible con las diferentes aproximaciones y modelos, tener en cuenta la voz de los grupos minoritarios y aceptar las diferentes perspectivas, incluso aunque éstas sean diametralmente opuestas. De hecho, aunque la evaluación es uno de los campos en los que mejor han empezado a convivir las tradiciones cualitativa y cuantitativa, ya que muchos evaluadores utilizan una mezcla de métodos y técnicas, sigue habiendo evaluadores cuantitativistas y cualitativistas que hablan lenguajes distintos. Este intento de integración de todas las posibles tendencias o, mejor dicho, el miedo a ser políticamente incorrectos, ha hecho que queden unos principios algo desabridos, dejando las orientaciones demasiado abiertas. De hecho, alguna de las críticas de los Principios Orientadores de la AEA van en esta dirección. Peter H. Rossi coincide con Chelimsky en que los Principios Orientadores de la AEA no se preocupan de forma especial por la integridad de los resultados ni en hacer al evaluador lo suficientemente valiente a la hora de presentar sus resultados (Chelimsky 1995; Rossi 1995). Así, considera que los Principios de la AEA son débiles y que «los códigos potentes y detallados existen en los campos profesionales en los que se ha alcanzado consenso sobre la sustancia y el método, mientras que los códigos débiles afloran en las asociaciones en las que los miembros están divididos en dichas materias» (Rossi 1995: 55). Finalmente, esta generalidad también está relacionada con la idea de no «sentar cátedra» cuando se trata de valores, especialmente cuando se viene de una tradición en las ciencias sociales en la que se ha pretendido, durante muchos años, que los valores tuvieran muy poco que ver con la práctica científica (House 1995). En este sentido, uno de los Principios Orientadores de la AEA es especialmente criticado por general: el referido a las «responsabilidades hacia el bienestar público y general».

<sup>7</sup> Utilizamos aquí el término «código deontológico» en sentido amplio, incluyendo todo tipo de normas, códigos o listados de principios, estándares u orientaciones en evaluación.

<sup>8</sup> Para profundizar en las diferencias entre los códigos del *Joint Committee*, y los de las asociaciones americana, canadiense y australasiática, véase Bustelo (1998).

En cualquier caso, esta generalidad a la que nos referimos está menos presente en las normas del *Joint Committee* y en las orientaciones éticas de la Sociedad Australasiática de Evaluación y más presente en las orientaciones de las asociaciones americana y, en parte también, en la canadiense. Así, se pueden distinguir dos tipos de códigos: unos, como son los de las asociaciones americana y canadiense, que tienen más que ver con principios orientadores generales para una actuación ética en evaluación y otros, como son las normas del *Joint Committee* y las orientaciones de la Sociedad Australasiática, en las que se especifica más cómo debe hacerse una evaluación<sup>9</sup>. Curiosamente, aunque el título de los códigos canadiense y australasiático es similar («orientaciones para una actuación ética»), sus características son bien distintas, siendo el primero muy general y el segundo mucho más específico.

En este sentido, creemos que se puede establecer una diferencia entre un tipo de códigos que se ciñen más a la idea de ofrecer principios u orientaciones éticas generales, y otros, que denotan una visión más amplia del quehacer evaluativo, y en los que subyace la idea de que hacer una evaluación éticamente consiste en hacerla bien, y que ofrecen por lo tanto, orientaciones más específicas y concretas para hacer bien una evaluación. De hecho, en los últimos manuales de evaluación, como es el de Carol H. Weiss, se deja entrever esta idea más amplia de la relación entre ética y evaluación. En el capítulo que dedica a cuestiones éticas («Evaluar con integridad»), esta autora señala seis ideas fundamentales sobre cómo se debe realizar una evaluación ética en una sociedad democrática: 1) conocer el programa o política que se va a evaluar y su contexto; 2) mantener la alta calidad técnica y la relevancia del estudio de evaluación; 3) utilizar el equilibrio y el juicio, siendo conscientes de que los evaluadores trabajan en un contexto en el que el programa es prioritario y las condiciones no siempre son buenas para el estudio sistemático; 4) considerar el uso desde el principio; 5) comportarse éticamente a lo largo del estudio; y 6) «sobrevivir y cosechar recompensas» (con esta máxima, Weiss pretende demostrar que, a pesar de todas las demandas que se requieren del evaluador, la evaluación merece la pena ya que es un trabajo que genera comprensión y mayor calidad de vida, y pretende mejorar la manera en la que se abordan los problemas públicos) (Weiss 1998: 320 y siguientes.). Como se puede observar, las cuestiones éticas a las que se refiere Weiss van más allá del «comportamiento ético» al que se refiere en el punto 5. Es decir, evaluar éticamente implica el propio concepto y concepción de la evaluación.

Muy ilustrativo resulta el caso de la Sociedad Australasiática de Evaluación, en cuyo caso, la aprobación de sus «Orientaciones para una Actuación Ética en Evaluación» en 1997 es completada por un «Código Ético» que se aprueba en diciembre de 2000, el cual es más similar en formato y contenido a los otros códigos más generales. Mientras las orientaciones «están dirigidas a todos aquellos que encargan, preparan, dirigen y utilizan las evaluaciones, además de a los que investigan, enseñan y publican sobre evaluación», el código es «una declaración de los valores y principios que los miembros [de la Sociedad Australasiática] sostienen en su trabajo de evaluación y en su membresía en la Sociedad»<sup>10</sup>. Los propósitos de las orientaciones son dos. Por un lado, son orientaciones para un comportamiento ético para la toma de decisiones en evaluación, promoviendo una práctica evaluativa ética. «A través de la estimulación de la conciencia y la discusión sobre asuntos éticos, las orientaciones pretenden promover la mejora continua en la teoría, la práctica y el uso de la evaluación». Por otro lado, las orientaciones «pretenden ayudar a la gente a reconocer y resolver cuestiones éticas particulares que aparecen en el curso de una evaluación». A través del «Código Ético», la Sociedad pretende promover la práctica evaluativa ética, apoyar los derechos de sus miembros y mantenerse como la principal organización en evaluación en su ámbito territorial. El «Código» se refiere a responsabilidades con el campo de la evaluación y el público en general, además de a responsabilidades con la Propia Sociedad Australasiática y los miembros. Así, todos los miembros de la Sociedad están obligados a adherirse y comportarse según dicho código.

<sup>9</sup> Las normas del *Joint Committee* se dividen en normas de utilidad, factibilidad o viabilidad, corrección o legitimidad y precisión. En el caso de las «Orientaciones para una actuación Ética en Evaluación» de la Sociedad Australasiática de Evaluación, los principios y orientaciones están divididos según el momento del proceso de evaluación al que se refieren: encargo y preparación de una evaluación, realización de la misma, y comunicación de los resultados.

<sup>10</sup> Véase <<http://www.aes.asn.au/ethics.cfm>>

En esta concepción más amplia sobre lo que significa hacer éticamente una evaluación, los códigos, como es el caso de las normas del *Joint Committee* o las orientaciones de la Sociedad Australasiática de Evaluación, se convierten en una guía orientativa sobre cómo hacer evaluaciones. En definitiva, crean una masa crítica y generan una base teórico-práctica sobre cómo pueden y deben hacerse las evaluaciones. Representan, pues, una base filosófica sobre lo que es y significa la evaluación, además de orientar sobre cómo deben hacerse. Por lo tanto, los códigos, además de una cuestión de regulación ética y de defensa profesional, se convierten, así, en una cuestión de identidad, de creación de identidad profesional y académica, algo extremadamente importante en disciplinas y prácticas de nuevo cuño, como es el caso de la evaluación. Esta última circunstancia está relacionada con la segunda función de los códigos, que se trata a continuación.

## 4. La función de institucionalización y consolidación de la evaluación

Efectivamente, los códigos deontológicos son también importantes para la creación de un sentido de pertenencia e identificación a una determinada comunidad profesional. Esta función de la deontología profesional, la cual nos parece clave, es además reconocida por parte de las organizaciones que han generado los códigos en el campo de la evaluación. Por ejemplo, la Asociación Americana de Evaluación considera abierta y explícitamente sus Principios Orientadores como un medio para la *socialización* de sus miembros. Más allá, no sólo se pretende que los Principios Orientadores sean una buena excusa para la discusión entre profesionales, sino que también sean un catalizador para la *identificación* de los profesionales como tales; es decir, que se genere grupo y sentido de pertenencia a una profesión que ha ido adquiriendo definición, autonomía y potencia creciente en los últimos treinta años. Esta circunstancia, lógicamente, va a contribuir a una mayor consolidación de la evaluación como campo profesional y académico de forma visible e identificable.

Además de esta función deontológica que pueden cumplir los códigos, tanto en su versión más estrecha de códigos éticos, como en la más amplia mencionada anteriormente, queremos insistir aquí en otras posibles funciones que potencialmente cumplen los códigos y que también contribuyen a la consolidación e institucionalización de la evaluación. Sobre todo si se consideran los códigos en el sentido más amplio de guías u orientaciones sobre cómo deben realizarse las evaluaciones, este conjunto potencial de normas o estándares pueden cumplir las funciones que se señalan a continuación

### (i) Contribución a la creación de una cultura de evaluación

Una de las razones por las que suele haber *ausencia* de una cultura de evaluación es porque no existen prácticamente *referencias de prácticas concretas*, identificadas como evaluación y, si las hay, éstas están aisladas y se conocen por parte de poca gente. Además de los estilos de gestión pública que se han ido forjando durante años, y del complejo contexto político en el que se da la evaluación de programas y políticas públicas, muchas veces el problema de ausencia de cultura es tan básico como que no se sabe realmente lo que es evaluación, ni cómo puede hacerse. Las personas que se encargan de la gestión pública, que formulan y ejecutan las políticas, o están en torno a las mismas, no han sido formadas específicamente en evaluación, y difícilmente pueden socializarse en una realidad que es prácticamente inexistente. Si además no existen prácticas similares a la evaluación en las instituciones o, como suele ocurrir, si existen, éstas no son reconocidas como tales, resulta difícil introducir una cultura de evaluación sin algunos mecanismos de ayuda.

En este sentido, parece especialmente adecuado generar mecanismos que ayuden a crear una base de lo que es y significa la evaluación, su importancia y posibilidades para las instituciones y administraciones públicas y para un adecuado desarrollo democrático de las mismas en su relación con la ciudadanía. Se trata, sobre todo, de desarrollar mecanismos que ayuden a definir lo que es la evaluación y a distinguirla de otras prácticas que no lo son, que generen alguna referencia sobre cómo deben y pueden hacerse las cosas en este campo. Los códigos, especialmente los que se desarrollan de forma más específica respecto a cómo se hacen las evaluaciones, pueden ser, en nuestra opinión, un excelente mecanismo de ayuda para ir introduciendo una cultura de evaluación en las administraciones públicas.

Relacionado con esto, la elaboración de códigos puede ser también una manera de fomentar el *fortalecimiento institucional para las unidades o instituciones dedicadas a la evaluación de programas y políticas*, las cuales pueden ejercer su autoridad y competencia en el campo de la evaluación a través de este mecanismo. Tal es el caso de los códigos elaborados por la OCDE y la Comisión Europea, y también, en otro nivel, el del *Joint Committee* de Estados Unidos. En el resto de los países han sido las Sociedades o Asociaciones de Evaluación las que han promovido la elaboración de códigos, entre otras cosas, porque tampoco existen muchas instituciones dedicadas exclusivamente a la evaluación de manera transversal. Como se verá más adelante, parece interesante reflexionar el papel que pueden cumplir instituciones de este tipo en la elaboración de códigos, y tal puede ser el caso de la Agencia Estatal de Evaluación, adscrita al Ministerio de Administraciones Públicas español.

Una de las cuestiones claves en la creación de una cultura de evaluación es la necesaria *integración de principios generales de evaluación* —conjunto de principios teóricos y metodológicos más o menos compartidos por la comunidad científica y profesional especializada en evaluación— *con las variables contextuales específicas* de cada contexto nacional. Los códigos deontológicos, siempre que se hagan de forma participada, permiten integrar dichos principios generales de evaluación con las variables específicas contextuales que se dan en cada contexto evaluativo. A pesar de la multidimensionalidad y variedad de prácticas evaluativas dentro de cada país o región, los sistemas nacionales de evaluación suelen tener algunas características contextuales específicas, las cuales pueden ser reflejadas o contempladas en estos códigos.

Por último, y relacionado con lo anteriormente expuesto, el *proceso de creación de códigos*, su elaboración y posterior revisión puede ser una excelente *oportunidad para el debate y la exposición de experiencias concretas*. Efectivamente, el propio proceso de elaboración y posterior revisión de códigos deontológicos puede acabar siendo más rico que el propio resultado. En primer lugar, puede —y debe— dar lugar a la participación de personas que pertenecen, de una manera u otra, a las comunidades de evaluación, bien sea haciendo, encargando, utilizando o investigando evaluaciones. En este sentido, la elaboración de un código deontológico puede dar lugar a intercambios y debates fructíferos, es decir, ser una buena oportunidad para el debate entre profesionales y académicos dedicados a la evaluación, y para la exposición y comunicación de experiencias concretas. Además, todos los códigos existentes no han nacido con una vocación permanente, sino que se someten a revisiones periódicas. Aunque estas revisiones casi nunca suelen conllevar cambios radicales, vuelven a ser una excelente oportunidad para que la comunidad evaluativa discuta y ponga en cuestión las normas, estándares y principios con las que quiere convivir, es decir, para que haga y extienda filosofía —y, por lo tanto, cultura— de evaluación.

## **(ii) Creación de lenguaje común y clarificación de criterios, cuestiones y elementos clave de la evaluación**

Toda práctica de nuevo cuño necesita de un lenguaje común, que se va creando a través de las prácticas y los profesionales, incluso aunque en muchos casos se adquieran términos ya acuñados en otro idioma (como es el

caso, en evaluación, de la importación de muchos términos ingleses al español). Hay que ponerse de acuerdo en los términos y en su significado. La elaboración de códigos deontológicos puede ayudar enormemente a la creación de un lenguaje común. En una reunión sobre normas y estándares de la Sociedad Europea de Evaluación en noviembre de 2003, Hans Lundgren de la OCDE, afirmaba que el glosario que se hizo para acompañar las «Orientaciones para una buena práctica en evaluación» (*Best Practices Guidelines for Evaluation, Public Management Service*, OCDE 1998) había resultado, en su opinión, un producto más importante que los propios principios.

También en un código deontológico en sentido amplio se expresan los *principios y elementos clave de la evaluación* (como son, en el caso de las normas del *Joint Committee*, la utilidad, viabilidad, legitimidad y precisión de las evaluaciones), todas las cuestiones a tener en cuenta a la hora de realizarla (como por ejemplo, credibilidad de los evaluadores, procedimientos prácticos, documentación del programa, acuerdos formales, claridad del informe y comunicación de resultados) y se ofrecen orientaciones que facilitan la elaboración y definición sobre qué y cómo evaluar de los programas y políticas públicas (como por ejemplo, identificación de *stakeholders* y de valores, conflicto de intereses, fuentes de información sistemática y análisis, así como la propia meta-evaluación). La elaboración y posible adopción de algún tipo de código significa poner en común y asumir de forma más o menos compartida dichos elementos clave en evaluación.

### (iii) Una forma de socialización e identificación profesional específica en el campo de la evaluación

La evaluación de programas y políticas públicas tiene una serie de características específicas (Ballart 1992; Bañón 2002; Monnier 1995; Vedung, 1997), que le confieren una identidad propia. Entre esas características específicas se encuentran su carácter político, el enjuiciamiento sistemático y su carácter práctico y aplicado, que la diferencian de otras prácticas de análisis, gestión e investigación. Así conviene distinguir la evaluación de programas y políticas públicas de la investigación social, el seguimiento o monitoreo, la gestión de la calidad, los observatorios, etc., y otorgar a la evaluación una identidad propia como campo científico y profesional diferenciado (Bustelo 2002).

La elaboración de códigos puede contribuir enormemente a la creación de esta identidad propia, máxime si éstos se formulan de manera transversal, como es el caso de prácticamente todos los códigos mencionados en este estudio<sup>11</sup>. Creemos que los códigos que sean aplicables a cualquier campo de actividad tienen más oportunidades de crear una identidad propia para la evaluación de programas y políticas públicas como ámbito profesional y académico diferenciado.

En relación con esto, parece importante también la contribución que los códigos pueden hacer para la identificación de especialistas evaluadores, es decir, profesionales que aunque puedan también desarrollar su práctica en alguna área o campo de programas o políticas (por ejemplo, políticas de formación profesional, género, o de servicios sociales), tengan como primera y principal especialización la evaluación de programas y políticas públicas. Por lo tanto, la elaboración de códigos de evaluación por parte de gente que tenga experiencia en diferentes campos de política y actividad no sólo es enriquecedor, sino imprescindible para otorgar una identidad propia. Esta transversalidad que caracteriza a algunos códigos es una excelente ayuda para la *socialización e identificación profesional*.

<sup>11</sup> Solamente las normas de la Sociedad Alemana de Evaluación se inician por una investigación en el campo de la educación y formación vocacional, aunque posteriormente se amplían a todos los campos. De igual forma, el primer código elaborado, el del *Joint Committee*, también se elaboró inicialmente en el campo de los programas y políticas educativas, aunque de forma rápida pasó a ser una referencia en la evaluación de todos los campos y sectores.

#### **(iv) Una cierta garantía en la relación de comanditarios-evaluadores-participantes en la evaluación**

Los códigos representan un documento de referencia, es decir, algo a lo que se puede recurrir en la relación de clientes y evaluadores para clarificar cuestiones o facilitar los puntos que deben ser tratados, además de ser algo que otorga cierta autoridad tanto a evaluadores, como a comanditarios y otros participantes, en situaciones complejas y difíciles, ya que especifica los elementos clave a tener en cuenta al hacer una evaluación, lo que es esperable y qué tipo de normas deberían cumplirse.

Esta circunstancia creemos que es especialmente importante cuando no existe una fuerte cultura de evaluación y hay una clara falta de referencias prácticas en evaluación, ya que los códigos pueden verdaderamente orientar en la práctica a los *stakeholders* o agentes críticos de una evaluación.

#### **(v) Apoyo a la formación/aprendizaje y a la investigación en evaluación**

Por último, en esta sección se discute la utilidad de los códigos deontológico más allá de su propia función, es decir, cómo pueden contribuir a labores de formación e investigación en el campo profesional y académico de la evaluación. En este sentido, los códigos son una excelente fuente para la *formación de evaluadores*. En primer lugar, y especialmente los que especifican más sobre cómo debe hacerse una evaluación, como es el caso del *Joint Committee* y similares, permiten hacer un repaso de la teoría y práctica de la evaluación, además de discutir cada una de las cuestiones clave en evaluación. Además, permiten dilucidar e introducir las cuestiones éticas en evaluación, asunto de vital importancia en la formación de evaluadores/as. Así, uno de los mejores métodos para la formación ética de futuros evaluadores, es el trabajo sobre casos prácticos que exponen dilemas éticos concretos<sup>12</sup> tomando como herramientas los códigos deontológicos existentes.

Por otro lado, la elaboración de códigos deontológicos en evaluación está íntimamente ligada a la *meta-evaluación*. En primer lugar, la propia realización de códigos es un ejercicio de tipo meta-evaluativo, ya que se trata de pensar en, analizar y valorar cómo se hacen y se deben hacer las evaluaciones. En segundo lugar, porque cualquier código que se precie y que realmente quiera integrar las variables específicas del contexto en el que va a ser aplicado, debe estar basado en la práctica evaluativa real, sus fortalezas y debilidades, en dicho contexto. Por lo que la meta-evaluación como técnica de análisis de evaluaciones se convierte en una herramienta imprescindible para dicho análisis. Y en tercer y último lugar, porque los códigos son una excelente guía (de hecho, cada una de las normas se puede convertir en un criterio de meta-evaluación) para la realización de meta-evaluaciones en sus dos sentidos: para evaluar y/o asegurar la calidad de las evaluaciones y para estudiar la función de evaluación en torno a determinados sectores o instituciones. Esta última función de la meta-evaluación, el estudio de la función de evaluación, tiene, además, importantes implicaciones académicas, además de las profesionales.

## **5. El papel que podría cumplir la elaboración de un código de evaluación en España**

En los últimos años en España la evaluación se ha comenzado a desarrollar, y aunque existen algunos indicadores de cambio y rápido desarrollo, nos enfrentamos a una realidad con una función de evaluación todavía

---

<sup>12</sup> Un excelente recurso es la sección sobre «Retos éticos» en la revista *The American Journal of Evaluation*.

poco consolidada. La Sociedad Española de Evaluación se crea en octubre de 2001, se crean algunos cursos de posgrado específicos en evaluación a partir de 2002<sup>13</sup>, y en mayo de 2004 se crea la Comisión para el estudio y creación de la Agencia Estatal de Evaluación de la Calidad de los Servicios y de las Políticas Públicas<sup>14</sup>, que culminará con la creación de dicha Agencia. Sin embargo, a pesar de estos evidentes signos de consolidación incipiente, no existe todavía una práctica evaluativa sistemática que no sea aislada y aleatoria, ni tampoco existen prácticamente publicaciones especializadas sobre la materia, ni una verdadera cultura de evaluación en nuestras instituciones públicas. Por ahora, tampoco existe ningún código en España, ni se ha comenzado ningún proceso de elaboración de normas, aunque se haya hablado de ello e incluso se haya creado una comisión de trabajo al respecto en la Sociedad Española de Evaluación. En este último apartado del artículo, se discute el potencial papel que podría cumplir un posible código o guía orientativa en este momento en España.

En este momento en España, generar un proceso de elaboración de algún tipo de normas podría contribuir esencialmente a crear cultura e institucionalizar la evaluación. En este caso, además del producto que finalmente se obtendría, sería especialmente importante y fructífero el proceso de creación de dicho código. Aunque la Sociedad Española de Evaluación ha ido creciendo —con más de un centenar de socios— en estos últimos años, todavía hay gente relacionada con la evaluación en distintos ámbitos que todavía no conoce o no se ha acercado a la misma. Si el proceso de elaboración y discusión de las normas se hace, al menos en una de sus primeras fases, de una forma participada, éste puede ser un catalizador de las diferentes prácticas de evaluación que se están llevando en cabo en los diferentes sectores. Sería a su vez una manera de conocer de forma más sistemática y amplia lo que se está haciendo en España en evaluación y por parte de quién.

Además de la Sociedad Española de Evaluación que, al igual que las diferentes asociaciones profesionales en otros países, debe tener un papel clave en este proceso, aquí es interesante pensar en un posible apoyo en el proceso de la Agencia Estatal de Evaluación. Creemos que esto podría resultar en un beneficio mutuo, fortaleciendo por un lado, la institucionalización de la función de evaluación y, por otro, el contacto institucional con las comunidades en práctica de evaluación. Por supuesto, siempre salvaguardando la independencia de ambas entidades. En un momento posterior podría estudiarse la cuestión de su aplicación, proceso al que se le daría prioridad en las revisiones posteriores.

En un potencial código español también habría que resolver, como en el resto de los códigos generados en los últimos años y a los que hemos hecho referencia, la tensión entre reflejar la realidad actual de la evaluación en España (lo que hay y es) con lo que debería ser y haber, extremos hoy por hoy todavía muy lejanos. Ya se ha comentado anteriormente que, en nuestra opinión, los códigos deben reflejar y aspirar al máximo (normas máximas) y deben reflejar, por tanto, el debería ser, pero, para que resulten útiles y relevantes para el contexto local, deben estar basados, por otro lado, en la realidad de la práctica de la evaluación de un país (lo que es). Esta tensión creemos que puede resolverse, en parte, haciendo hincapié en orientaciones especialmente dedicadas a los problemas más acuciantes o las áreas más sensibles de la práctica profesional. Por eso es extremadamente importante basar el proceso de elaboración de los códigos en un proceso participativo en el que la gente que está trabajando en evaluación en los diferentes ámbitos, exponga los principales problemas con los que se encuentra en su práctica evaluativa.

Además de este proceso participativo, también deben utilizarse como base las meta-evaluaciones realizadas y por realizar en nuestro país, es decir, los estudios e investigaciones sobre procesos de evaluación en los diferentes ámbitos y sectores. En este sentido, el papel de la Agencia Estatal fomentando, encargando y realizando este tipo de estudios puede ser también crucial.

<sup>13</sup> Véase Máster en Evaluación de Programas y Políticas Públicas UCM (<http://www.ucm.es/info/csegae/>); Máster de Evaluación. Planificación, Intervención y Evaluación de Políticas Públicas y Programas Sociales, UAL (<http://www.ual.es/personal/bfernan/MasterdeEvaluacion.htm>)

<sup>14</sup> Orden del Ministerio de Administraciones Públicas 1463/2004, de 19 mayo (BOE de 25 de mayo de 2004).

Teniendo en cuenta la necesidad de basar un posible código español en el análisis de los problemas más acuciantes o las áreas más sensibles de la práctica evaluativa española, se presentan en esta parte del artículo algunas ideas a considerar en una potencial elaboración de normas españolas de evaluación. Para ello, me baso en una meta-evaluación realizada en el ámbito de las políticas de género en España (Bustelo 2004), destacando lo que, en mi opinión, son conclusiones que podrían ser extrapolables a otros ámbitos o sectores de política en la actualidad en España. Para ello, se organizan los principales problemas encontrados en cinco grandes bloques que se exponen a continuación.

### • **Cuestiones relacionadas con los propósitos y la concepción de la evaluación**

Un primer factor que influye muy a menudo negativamente en las evaluaciones es la falta de claridad a la hora de plantearse la evaluación: algo tan básico como el *para qué y por qué* se hace la misma no aparece de forma nítida. Un indicador evidente de esta falta de claridad es el planteamiento general de algunos de los estudios de evaluación que fueron meta-evaluados, en los que no se supo distinguir entre hacer investigación sobre el problema a abordar —en este caso la situación de desigualdad de la mujer y las desigualdades de género— y hacer evaluación de las actuaciones para abordarlo —los planes y las políticas de igualdad de género—, cuestión esta última directamente relacionada con la función de responsabilidad y rendición de cuentas de la evaluación. Por lo tanto, parece crucial que cualquier código o conjunto de normas u orientaciones insista en la clarificación del fin y la función de las evaluaciones.

Una segunda cuestión es que ninguna de las experiencias analizadas se planteó como una *evaluación más global de la política* de igualdad de género, sino como una evaluación concreta de un determinado plan. Efectivamente, en general en España, existe una tendencia a hacer evaluación de medidas, actuaciones o proyectos concretos, careciendo a veces de una visión más global que tenga en cuenta planes y políticas más generales.

En tercer lugar, otra de las razones que se identifican como clave de la falta de claridad en el fin y las funciones de las evaluaciones estudiadas, es la *ausencia de reconocimiento del carácter político* que todo proceso de evaluación tiene. Por lo general, tiende a considerarse la evaluación como un asunto eminentemente técnico, y esto es algo que inevitablemente hace que se desvirtúen los procesos de evaluación, haciéndolos en muchos casos residuales e inútiles para los fines que supuestamente se habían propuesto en un principio.

### • **Cuestiones relacionadas con la naturaleza valorativa de la evaluación**

Uno de los pasos cruciales en evaluación consiste en determinar cuáles son los *criterios de evaluación*, es decir qué es lo que se quiere evaluar, una vez se clarifica por qué y para qué queremos evaluarlo. En primer lugar, es importante ya que es lo que determina y guía el estudio, orientando la recopilación y el análisis de la información; en definitiva, se trata de la identificación y definición de las variables que se van a valorar. Pero además, en evaluación, se determinan también los criterios de valor, es decir, a través de la definición de los criterios estamos definiendo la base lógica con la que posteriormente vamos a emitir los juicios sobre el programa o política evaluada. En definitiva, los criterios de evaluación son los que van a definir la perspectiva de valoración a utilizar.

El análisis de las experiencias de evaluación metaevaluadas en el trabajo mencionado permitió detectar que, en general, existía una cierta confusión respecto a la definición de los criterios de evaluación. En la mayoría de los casos ni siquiera los criterios aparecían explícitamente reflejados en los informes de evaluación. En general, pen-

sar sobre los criterios de evaluación es un paso clave y que, sin embargo, se suele obviar o se suele dar por hecho, pareciendo que ya están preestablecidos unos criterios para la evaluación que son generales, preexistentes y válidos para cualquier situación, pero que además nadie explicita de una forma clara. Por lo tanto se deduce que no hubo una reflexión seria previa a los procesos de evaluación sobre realmente qué se quiere evaluar. Es más, en muchas ocasiones da la impresión de que el QUÉ evaluar (criterios de evaluación) se ha definido a través del cómo evaluarlo (metodología y, sobre todo, las técnicas utilizadas), dando prioridad a esto último. Así, parece que lo más importante son las técnicas o instrumentos para recoger la información y esto, en muchos de los casos, opaca los pasos previos (definición de los criterios de evaluación) y también algunos de los posteriores (valoración, emisión de juicios y recomendaciones).

De esta manera, también detectamos en las once experiencias metaevaluadas que existía un claro déficit de lo que denominamos «*traducción o elaboración práctica*». Además de recopilar información, analizarla e interpretarla, en evaluación también se debe valorar dicha información (*enjuiciamiento sistemático*) y, posteriormente, realizar sugerencias para la acción (*establecimiento de recomendaciones*). La denominación de «*traducción o elaboración práctica*» viene dada por el hecho de que son estos dos últimos pasos los que confieren un necesario carácter aplicado a la evaluación. Así, al ser la evaluación una forma especialmente aplicada de indagación, ya que se hace con la intención de mejorar y rendir cuentas sobre el programa o plan a evaluar, debe ser práctica y estar dirigida a la acción. En definitiva, las evaluaciones se deben hacer con la clara y directa intención de ser utilizadas, bien sea para tomar decisiones y modificar aspectos concretos del programa a evaluar, bien sea para contribuir al aprendizaje de las organizaciones que tienen como una de sus misiones la función de evaluar sus actuaciones públicas. Por lo tanto, parece evidente que uno de los «*productos*» de los procesos de evaluación debe ser un listado de recomendaciones que « *traduzca*» en términos prácticos lo que se ha concluido a través de dicho proceso o estudio. Este déficit que detectamos de una forma muy evidente en las experiencias de evaluación de las políticas de igualdad de género, en nuestra opinión es algo que tiende a ocurrir, en mayor o menor grado, en otros ámbitos y sectores. En definitiva, esta circunstancia que hace que en muchos casos se considere sobre todo la parte de recopilación y análisis de la información, despreciando otros pasos del análisis, resulta del hecho de no reconocer de forma clara la naturaleza valorativa de la evaluación. Cualquier código que se elaborara debería insistir en dicha naturaleza valorativa.

### • **Cuestiones relacionadas con la gestión de la evaluación**

La mayoría de las evaluaciones estudiadas, que para muchos organismos de igualdad representaban sus primeras experiencias, adolecieron en cierto modo de una visión estrecha y parcial de la evaluación, es decir, fueron evaluaciones con una tendencia a ser parciales y finalistas. Si tenemos en cuenta como se utilizaron los diferentes *tipos de evaluación*, vemos que en muchos casos se gestionaron inadecuadamente las posibilidades de cada tipo de evaluación, especialmente si tenemos en cuenta la tipología de evaluaciones según el contenido o fase del programa a evaluar. Así, muy pocas experiencias realizaron evaluación del *diseño o conceptualización* del programa. Sin embargo, resaltamos aquí la importancia de este tipo de evaluación, especialmente si se quiere reforzar una perspectiva de evaluación de políticas más que de planes y actuaciones concretas. En lo que se refiere a la evaluación de *proceso*, se puede concluir que ésta se hizo mucho menos de lo que se pretendía. En realidad las evaluaciones llamadas de proceso o implementación han sido, en muchos casos, evaluaciones de producto o *outputs*, limitándose a describir las actuaciones o actividades que se habían realizado —o no— como parte de los planes. Sin embargo, prácticamente no se evaluaron los mecanismos de funcionamiento a través de los cuales los planes o programas se ponían en marcha. Por último, la evaluación de *resultados* es la que más se desarrolló en su vertiente de evaluación de productos o *outputs*. Es decir, hubo una tendencia a evaluar las actuaciones que resultaban de los planes aunque, en su mayoría, este

tipo de evaluación se reducía a valorar el «grado de cumplimiento» y que consiste simplemente en comprobar si se realizan o no cada una de las acciones o medidas propuestas. En muchos casos, ni siquiera se hizo una valoración global del plan, sino de cada una de las actuaciones. Finalmente, no se realizó prácticamente evaluación de resultados en su vertiente de efectos o *outcomes*, sobre todo si hablamos de los efectos netos de los planes.

Además de aprovechar mejor las oportunidades que nos brindan los diferentes tipos de evaluación, se trataría de utilizarlos conjunta y sinérgicamente. Por lo general, cuando se ha hecho evaluación de la implementación o proceso de ejecución del plan y evaluación de resultados, se ha tendido a hacer dos estudios separados con procesos diferenciados. Y esto hace que la evaluación pierda mucha de su posible potencia.

Hay dos cuestiones relacionadas con el tiempo y que resultan cruciales en cualquier proceso de evaluación. En primer lugar, *el momento en el que se plantea la evaluación*, ya que ésta resulta una herramienta mucho más potente si se plantea desde el inicio, en la misma fase de diseño y planificación de los planes. Y esto es algo de lo que, con alguna excepción, adolecieron todas las experiencias analizadas, ya que todas las evaluaciones se plantearon hacia finales del período de vigencia del plan que se quería evaluar. Esta circunstancia acarrea múltiples dificultades. La segunda cuestión es la del *calendario de la evaluación*, si éste ha sido ajustado, o si ha habido algún retraso sobre el calendario previsto, que es lo que ha ocurrido en prácticamente todas las evaluaciones. Aunque en algunas no han sido tan significativos en otras dichos retrasos pusieron en duda la utilidad global del estudio. La solución a este problema de los tiempos sabemos que no es sencilla, ya que existen claros condicionamientos en el propio funcionamiento de las organizaciones públicas, como, por ejemplo, el necesario ajuste presupuestario al año fiscal, los casi siempre lentos procedimientos para la contratación externa o los propios ritmos organizativos. Sin embargo, creemos que prestarle más atención a esta variable es esencial para el éxito de las evaluaciones.

Una de las grandes dificultades que se mencionaron reiteradamente por las responsables de los organismos de igualdad encargados de las evaluaciones analizadas, era la *dificultad de recopilar la información sistemática* necesaria para realizar la evaluación. Al ser transversales los planes, una buena parte de la información que había que recopilar para la evaluación no se encontraba en los propios organismos de igualdad, sino en otras unidades administrativas. Para esta ardua tarea, que no es exclusiva de las políticas de igualdad, además de una buena dosis de autoridad y de persuasión, los organismos públicos que lideren las evaluaciones necesitan, por un lado, buenos *sistemas de información*, y, por otro, *estructuras de coordinación* adecuadas que los sustenten, alimenten y mantengan. Resulta claro pues que, más que un problema técnico relacionado con los instrumentos y técnicas para recopilar y analizar la información, existe un problema de decisión política y de un posterior diseño organizativo. El marco institucional para la evaluación es en sí un escenario político importante que no normalmente no se considera.

### • Cuestiones relacionadas con la inversión en evaluación

Para que las evaluaciones que se realicen sean útiles y relevantes se debe considerar la función de evaluación de manera central en las instituciones que lideren y protagonicen los procesos de evaluación. Sin embargo, en el caso de la meta-evaluación realizada, aunque una de las *principales funciones de los organismos de igualdad*, además de la información y sensibilización y la de investigación en temas de igualdad, fuera el impulso, coordinación, seguimiento y evaluación de los correspondientes planes, con alguna excepción, la función de evaluación es prácticamente residual en la mayoría de los organismos de igualdad que realizan las evaluaciones analizadas.

También es importante la cuestión de los *recursos invertidos* en los procesos de evaluación. En este sentido, el presupuesto empleado en los estudios de evaluación analizados fue muy heterogéneo según las experiencias pero, en cualquier caso, muy poco en términos relativos. Incluso en el caso en el que los gastos directos fueron los más altos, como es en el caso de Andalucía, el gasto en evaluación representaba sólo el 0,07% del presupuesto total del Plan evaluado y el 0,97% del presupuesto del Instituto Andaluz de la Mujer en los tres años de vigencia del Plan. En la elaboración de unas normas, habría que tener en cuenta y dar alguna referencia sobre cómo debe invertirse en evaluación.

### • Cuestiones relacionadas con la comunicación de los resultados

En general, los *informes finales* de evaluación no fueron muy completos y además su *difusión* no ha sido por lo general amplia. En la mayoría de los casos, los informes se han quedado para «uso interno». De hecho, el carácter público de los informes y el *acceso* a los mismos no está claro en muchos casos. Lo curioso es que no parece existir un criterio común al respecto. Como anécdota ilustradora de esta falta de criterios a la que nos referimos, mientras en una comunidad autónoma tuvimos que avalar nuestra petición de información con la garantía de la Universidad Complutense, en otra acabaron enviando la copia del informe a la dirección particular de la investigadora, después de muchos meses de espera y probablemente por temor a que la «vía oficial» no fuera la adecuada.

Respecto a la accesibilidad de los informes de evaluación, sería imprescindible que un posible código o conjunto de normas discutiera y pusiera en común los criterios al respecto, incluyendo el acceso y el carácter público de los informes de evaluación realizados en el seno de o para las instituciones públicas españolas.

## 6. Conclusiones

Como se ha puntualizado en la primera parte de este artículo, los códigos pueden contribuir sustancialmente al desarrollo de una cultura de evaluación y, en definitiva, a la mejora de la calidad y a la profesionalización de la evaluación en las administraciones públicas. Más allá, el mismo hecho de plantearse su elaboración puede promover procesos participativos de debate y discusión entre profesionales y académicos. Además, si queremos tener normas o códigos útiles y que estén basados en la realidad, su proceso de creación debe necesariamente animar análisis y metaevaluaciones de prácticas de evaluación en diferentes áreas, sectores y territorios.

También en este artículo se ha analizado el concreto caso español, destacando la importancia de integrar el contexto español y las prácticas reales de evaluación que se llevan a cabo en España. Para tal propósito, resulta crucial, además de hacer un proceso de elaboración participado con la comunidad de evaluación, incorporar los aspectos críticos que los estudios de meta-evaluación destaquen. En este sentido, y tomando como ejemplo una metaevaluación realizada en el ámbito de las políticas de género, se destacan algunas cuestiones clave a integrar en un potencial código, relacionadas entre otras con la concepción de la evaluación, su naturaleza valorativa y el acceso a los informes y a los resultados de la misma.

Por último, este artículo ha puesto énfasis en la importancia de las asociaciones y sociedades profesionales de evaluación en el proceso de creación de los códigos, así como el apoyo institucional que pueden recabar. En el caso español, también se ha querido destacar el importante papel que puede cumplir la Agencia Estatal de Evaluación en el apoyo a la creación de códigos en España.

## REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Ballart, X. (1992), *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudios de caso*. Madrid: Ministerio para las Administraciones Públicas.
- Bañón, R. (2002), «La evaluación de la acción pública como responsabilidad democrática», en R. Bañón, comp., *La evaluación de la Acción y de las Políticas Públicas*. Madrid: Díaz de Santos.
- Bustelo, M. (1998), «Deontología de la evaluación: el modelo de los códigos éticos anglosajones», *Gestión y Análisis de Políticas Públicas* 11-12: 141-156.
- Bustelo, M. (2002). «¿Qué tiene de específico la metodología de evaluación?», en R. Bañón, comp., *La evaluación de la Acción y de las Políticas Públicas*. Madrid: Díaz de Santos.
- Bustelo, M. (2004), *La evaluación de las políticas de igualdad de género en España*. Madrid: La Catarata.
- Beywl, W. y Speer, S. (2004), «Data and Literature-Based Reflections on Western Europe Evaluation Standards and Practices», en C. Russon y G. Russon, eds., *International Perspectives on Evaluation Standards. New Directions for Evaluation* (104). San Francisco: Jossey Bass.
- Comisión Europea (2000), *Good Practice Guidelines for the Management of the Evaluation Function*. Working Group on Guidelines. Network of Evaluators of the European Commission. (<http://www.oecd.org/puma/mgmtres/pac/index.htm>).
- Chelimsky, E. (1995), «Comments on the Guiding Principles», en Shadish *et al.* eds., *Guiding Principles for Evaluators. New Directions for Evaluation* (66). San Francisco: Jossey-Bass.
- EMCDDA (1998), *Guidelines for the Evaluation of Drug Prevention: A Manual for Programme-Planners and Evaluators (Manuals)*. Lisboa: European Monitoring Centre for Drugs and Drug Addition.
- House, E.R. (1995). «Principled evaluation: A Critique of the AEA Guiding Principles», en Shadish *et al.* eds., *Guiding Principles for Evaluators. New Directions for Evaluation* (66). San Francisco: Jossey-Bass.
- Joint Committee on Standards for Educational Evaluation (1994), *The Program Evaluation Standards. How to Assess Evaluations of Educational Programs*. Thousand Oaks: Sage.
- Love, A. y Russon, C. (2004), «Evaluation Standards in an International Context», en C. Russon y G. Russon, eds., *International Perspectives on Evaluation Standards. New Directions for Evaluation* (104). San Francisco: Jossey Bass.
- Monnier, E. (1995), *Evaluación de la acción de los poderes públicos*. Madrid: Instituto de Estudios Fiscales.
- NORAD (1997), *Evaluación de proyectos de ayuda al desarrollo. Manual para evaluadores y gestores*. Madrid: IUDC-UCM y CEDEAL.
- Rossi, P.H. (1995), «Doing Good and Getting It Right» en Shadish *et al.* eds., *Guiding Principles for Evaluators. New Directions for Evaluation* (66). San Francisco: Jossey-Bass.
- Rouge, J-C. (2004), «The Origin and Development of the African Evaluation Guidelines», en C. Russon y G. Russon, eds., *International Perspectives on Evaluation Standards. New Directions for Evaluation* (104). San Francisco: Jossey Bass.
- Vedung, E. (1997). *Evaluación de Políticas Públicas y Programas*. Madrid: Ministerio de Trabajo y Asuntos Sociales. Instituto Nacional de Servicios Sociales.
- Weiss, C.H. (1998), *Evaluation. Methods for Studying Programs and Policies. Second Edition*. Upper Saddle River (N.J.): Prentice Hall.
- Widmer, T. (2004), «The Development and Status of Evaluation Standards in Western Europe», en C. Russon y G. Russon, eds., *International Perspectives on Evaluation Standards. New Directions for Evaluation* (104). San Francisco: Jossey Bass.