

**Unidad de Políticas Comparadas (CSIC)**

Documento de Trabajo 04-07

## **Evaluación de la investigación y sistema de ciencia**

**Luis Sanz Menéndez**

Consejo Superior de Investigaciones Científicas (CSIC)  
Unidad de Políticas Comparadas,

SPRITE

(Spanish Policy Research in Innovation & Technology, Training & Education)

Julio de 2004

# Evaluación de la investigación y sistema de ciencia

Luis Sanz Menéndez\*

Unidad de Políticas Comparadas del CSIC, Madrid

A finales de los setenta ya existía en los Estados Unidos una cierta preocupación por los procedimientos de evaluación utilizados en la asignación de fondos de investigación por las agencias públicas de fomento de la I+D [la *National Science Foundation* (NSF) o los *National Institutes of Health* (NIH)], como demuestran las revisiones realizadas de estas prácticas (por ejemplo Cole *et al* 1978). El interés por los métodos de evaluación de la investigación propició el desarrollo de trabajos desde la OCDE (OECD, 1987) que contribuyeron a ampliar la perspectiva y a introducir elementos que conectaban la tradicional forma de la evaluación de la investigación, por medio de la revisión por pares, con la evaluación de las políticas de investigación y de otros objetos, lo que implicaba la introducción de nuevos instrumentos y herramientas, a la vez que se hacían más complejos los criterios de evaluación.

A principios de los noventa, en Estados Unidos, algunos afirmaban (Chubin y Hackett, 1991) que la revisión por pares (*peer review*) era un tema olvidado en la política científica, a pesar de su centralidad para el sistema de investigación, y que se prefería mantener una fe ciega en cómo se hacían las cosas en este campo a ponerlas en cuestión, incluso si estas parecían ilógicas, se hacían de forma injusta, iban contra el avance del conocimiento o en perjuicio de las carreras investigadoras.

En esos mismos años en España, en el contexto del lanzamiento de la primera política científica merecedora de este nombre (Sanz Menéndez, 1997), la demanda de un sistema de evaluación que estuviese garantizado por el Estado (Sanz Menéndez, 1995) condujo a la creación de una estructura singular que se responsabilizase de la evaluación de las actividades de investigación (la Agencia Nacional de Evaluación y Prospectiva - ANEP).

Desde entonces, el interés por la evaluación de la investigación (en sentido amplio) se ha extendido. Por un lado, es un tema que continúa atrayendo la atención de los gobiernos y de las agencias de financiación<sup>1</sup>; por otro lado, el interés de la comunidad científica sobre los procesos de evaluación en las revistas académicas no ha hecho sino crecer<sup>2</sup>. Sin duda la dinámica competitiva de la ciencia ha ayudado a este creciente interés.

La evaluación de la investigación, y de algunos aspectos muy asociados a la misma, parece ser hoy un asunto de moda en España. Presenciamos una explosión de organismos y nuevos mecanismos de evaluación de la investigación o de la calidad de las universidades que ha llevado a la creación de la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) para el conjunto de España, de la Agencia de

---

\* Contacto: Lsanz@iesam.csic.es

<sup>1</sup> Por ejemplo, la Comisión Europea ha otorgado a la evaluación un papel muy singular en el desarrollo de su política de I+D tratando de articular diversas dimensiones, desde la evaluación *ex ante* de las propuestas, al seguimiento de los programas y proyectos, pasando por la evaluación de los impactos e incluso, más recientemente, por la medición de la excelencia en investigación.

<sup>2</sup> Véase Cole (1998) o en castellano Campanario (2002).

Gestió d'Ajuts Universitaris i de Recerca (AGAUR) y la Agencia per a la Qualitat del Sistema Universitari de Catalunya (AQU) en Cataluña, la Agencia de Calidad, Acreditación y Prospectiva de las Universidades de Madrid, la Agencia Canaria de Evaluación de la Calidad y Acreditación Universitaria (ACECAU), así como de entidades similares de evaluación y de financiación de la investigación en otras Comunidades Autónomas. También se ha observado el reforzamiento de los procedimientos autónomos de evaluación de la investigación en instituciones como el Fondo de Investigaciones Sanitarias (FIS).

## **El contexto de la evaluación de la investigación**

Para comprender la evaluación de la investigación, su dinámica de cambio y sus desafíos, debemos situarla en el contexto del sistema de I+D y su interacción con las políticas que tienden a promoverlo. Hace más de medio siglo que los sociólogos de la ciencia llamaron la atención sobre el papel central que la evaluación –leída como control de calidad- jugaba en el sistema de producción de conocimiento científico certificado que es la ciencia. El momento “fundacional” de esta práctica institucionalizada, en forma de revisión por pares expertos, se sitúa en 1665, cuando la Royal Society instauró un sistema por el cual la presentación de trabajos para su publicación en *Philosophical Transactions* debía realizarse con el informe favorable de un miembro de la Royal Society (Zuckerman y Merton, 1971).

Así pues, desde los primeros momentos del desarrollo de la práctica investigadora diferenciada y de la consolidación de la institución científica, la obtención de la “certificación” y la declaración del conocimiento como válido ha sido parte consustancial de la ciencia. La evaluación de la investigación se ha desarrollado inicialmente en el marco del sistema de comunicación científica y la aceptación de trabajos en revistas científicas estaba sometida al juicio de los pares que trataban de determinar sus contribuciones. La evaluación por pares de los artículos científicos o de los méritos curriculares ha sido el mecanismo central que la comunidad investigadora ha utilizado para la aceptación de las publicaciones científicas, el reconocimiento con los puestos, e incluso para el otorgamiento de premios (como los Nobel).

Además del control de calidad, desde los inicios, la evaluación del mérito por parte de los pares científicos ha cumplido también la función esencial de incidir en la dirección del desarrollo del conocimiento de las diversas disciplinas. Sin embargo, el desarrollo, la expansión y la consolidación de las instituciones de la ciencia ha llevado a que surjan tensiones con las demandas mas amplias que emergen de la sociedad y del sistema político, y que hacen de la evaluación un instrumento para determinar la asignación de recursos, un mecanismo para cambiar estructuras organizativas y definir nuevos incentivos, o para evaluar los resultados en ciertas áreas científicas con relación a las necesidades nacionales.

Repasemos cómo, a resultas de algunos cambios ocurridos en la política de I+D, se han diversificado los objetos y los criterios utilizados en la evaluación de la investigación.

En primer lugar, señalar que, con el desarrollo y extensión de los sistemas de financiación pública de la investigación y la creación de instituciones y agencias especializadas en el fomento de la I+D (*NSF, Research Councils, etc.*), se produjo la

generalización de la evaluación de las propuestas de financiación. El *peer review*, en su forma pura o modificada, se convirtió en práctica ordinaria para la atribución de fondos para investigación. El método se basaba en la determinación del mérito de las propuestas y era la propia comunidad científica la que daba sus opiniones y asesoraba las decisiones sobre la asignación de fondos; de este modo se conseguía la legitimidad necesaria para el funcionamiento del sistema de I+D (Rip, 1994).

Hasta entonces los criterios utilizados para la evaluación de la investigación seguían asociados principalmente al mérito o a la calidad científica, bien fuese del artículo o de la propuesta de investigación. Sin embargo, con el paso de los años, se desarrollaron nuevas formas de intervención estatal en la política de I+D: la investigación por objetivos, la investigación estratégica, la investigación priorizada, etc. En definitiva, se consolidaban a través del sistema político, nuevas demandas de la sociedad o la economía a los investigadores con relación a la evaluación de la investigación. Por otra parte, algunas presiones adicionales en el sistema de investigación, derivadas de la escasez de recursos para sostener el crecimiento exponencial de la ciencia, han contribuido también a reforzar las tendencias en favor de una mayor selectividad y, por tanto, de un mayor papel de la evaluación en la asignación de la financiación o del espacio de las revistas en definitiva para la mejor utilización de los recursos. Además, el desarrollo de programas de I+D, que definen objetivos que desbordan la producción de conocimiento o la financiación de los proyectos más meritorios, introdujo un elemento adicional que se relacionaba con la toma de decisiones de los responsables y el seguimiento de los resultados de las actividades o el cumplimiento de los objetivos. La multiplicidad de objetivos para las actividades de producción científica y, por tanto, la diversidad de criterios de evaluación crea además un contexto más complejo para el desarrollo de ésta.

A pesar de que la producción de conocimiento es una actividad internacional, las prácticas de evaluación de la investigación están integradas en el funcionamiento de los sistemas de I+D, por lo que existe una significativa variedad de arreglos institucionales y organizativos en los que se desarrollan estas actividades en los diversos países, así como de los objetos que se evalúan, de los criterios que se utilizan y de las consecuencias que tienen<sup>3</sup>. Por ejemplo, la evaluación de la investigación se encuentra muy institucionalizada en el Reino Unido, donde se han desarrollado interesantes avances metodológicos en el campo de la evaluación de programas de I+D (ROAME) o de los impactos socioeconómicos de la I+D. Por otro lado, la evaluación de la investigación se ha convertido en un mecanismo esencial de distribución de los escasos recursos para la financiación de la I+D en las universidades, por medio del denominado *Research Assessment Exercise* (RAE). En Francia, la práctica más extendida es la que pretende utilizar la evaluación como una herramienta de dirección y gestión estratégica de los recursos del sistema de I+D. Por ejemplo, en el caso de la evaluación de los centros de investigación, con el objetivo de racionalizar o redefinir sus misiones o sus estrategias de desarrollo. Otros países, especialmente los de menor tamaño, resaltan el interés por evaluar los efectos socio-económicos de la investigación o del funcionamiento del conjunto del sistema de innovación, como es el caso de Finlandia.

Los Ministros de Ciencia de OCDE reunidos el pasado mes de enero de 2004 han llamado la atención sobre la importancia del control de calidad en el sistema de I+D, así

---

<sup>3</sup> Véase el número monográfico de *Research Evaluation*, vol. 5, nº 1, April 1995, sobre los sistemas nacionales de evaluación.

como del intercambio de experiencias y desarrollos metodológicos sobre la evaluación de la investigación y de los programas de I+D para el aprendizaje colectivo. Hoy en día, evaluación de la investigación implica una nueva concepción más amplia e integradora que supere la tradicional identificación de la misma con la revisión por pares; ésta es sólo una de las posibles técnicas a ser utilizadas, siendo otras las encuestas, los modelos econométricos, los trabajos bibliométricos, etc. La evaluación puede verse hoy como un método para proveer información de carácter evaluativo sobre algunos aspectos relacionados con la investigación (los objetos: individuos, proyectos, centros, programas, políticas, etc.) contra algunos criterios u opiniones (que van desde la excelencia investigadora a los efectos socio-económicos). La evaluación de la investigación, de sus resultados, de sus actividades, de sus instituciones o de sus actores ha ganado relevancia, porque ofrece la posibilidad de contribuir a guiar y a gestionar el sistema de I+D, dotándole de mayor coherencia y facilitándole el aprendizaje continuo y, en asociación con otros instrumentos como la prospectiva, favoreciendo una visión estratégica.

## **La institucionalización de la evaluación de la investigación en España**

El panorama de la evaluación de la investigación en España ha estado dominado, casi en exclusiva por la evaluación *ex ante* de los proyectos de investigación (a los efectos de obtener financiación) y de los individuos (a los efectos de los sexenios) (Sanz Menéndez, 1995). En otros países, la evaluación de los proyectos de I+D está interiorizada en el sistema y se debate más sobre la evaluación de otros objetos, tales como los centros de investigación, las universidades o los programas de I+D, tanto desde el punto de vista de su calidad o excelencia como de sus impactos socio-económicos.

En general, en otros países, las agencias de financiación o los Consejos de Investigación son los responsables de organizar el proceso de evaluación de las propuestas de investigación, de seleccionar los expertos y de designar los paneles de selección. Sin embargo, en España se creó una unidad administrativa denominada ANEP, actualmente con rango de Subdirección General que, separada de la maquinaria de la gestión de los programas de I+D, se hiciese cargo de organizar de modo relativamente autónomo, con la colaboración de la comunidad científica, el proceso de revisión de las propuestas de financiación de la investigación con el objetivo de determinar el mérito o la calidad científica de las mismas. La creación de la ANEP resolvió algunos problemas de credibilidad y legitimación de la atribución de fondos públicos competitivos a los investigadores.

De una situación, a finales de los ochenta, caracterizada por el “quasi-monopolio” de la evaluación de las actividades de la investigación por parte de la ANEP, solamente desafiado por la creación de la Comisión Nacional de Evaluación de la Actividad Investigadora (encargada de los sexenios), se ha pasado a una multiplicación de organismos pertenecientes a entidades estatales y autonómicas que organizan y desarrollan actividades de evaluación (*ex ante* o *ex post*) de algunos objetos asociados a la actividad de investigación (individuos, departamentos, proyectos de investigación, titulaciones, programas de doctorado, universidades, e incluso programas o planes de I+D), con el propósito de determinar el mérito o la excelencia científica más que los efectos o impactos de la investigación.

Esta multiplicación de organismos que desarrollan, de forma estable o discontinua, las actividades de evaluación relacionadas con el sistema de investigación se ha producido más como resultado de procesos políticos, donde los actores regionales organizan sus esquemas de evaluación necesarios para el desarrollo de sus propias políticas, que como resultado de una crisis de funcionamiento o de las críticas derivadas de la actividad de la ANEP. La centralidad de la ANEP parecía resultar de la relevancia de sus decisiones con relación a la financiación de la investigación, entonces concentrada en la Secretaría General del Plan Nacional de I+D y en la Secretaría de Estado de Universidades e Investigación (Promoción General del Conocimiento). La expansión radical de la financiación de los gobiernos regionales en los noventa y las dificultades de la ANEP, dados sus escasos recursos organizativos para prestar el servicio puntual a las CC.AA., han contribuido a la pérdida de centralidad de la ANEP en el sistema español de I+D, aunque no de la evaluación de la investigación, que resurge en forma de un sistema distribuido.

## **Desafíos para la evaluación de la investigación**

Los sistemas de I+D a escala internacional se enfrentan a una serie de desafíos, con independencia de sus contextos institucionales específicos, que son también significativos para el caso español:

En primer lugar, siguen abiertos los posibles desajustes cognitivos, derivados de la denominada “tensión esencial” (Kuhn, 1977), por los cuales el proceso de evaluación como el control de calidad de la investigación (o de los resultados socioeconómicos) debe ser capaz de evitar que su aplicación frene el desarrollo de las ideas más innovadoras en ciencia.

El segundo desafío al que se enfrenta la evaluación es la creciente interdisciplinariedad de la actividad investigadora. Se ha dicho que la investigación transdisciplinar es una de las características del nuevo modo de producción del conocimiento (Gibbons *et al* 1994) y se ha señalado que uno de los desafíos que el sistema tiene ante sí es el control de calidad de la investigación generada fuera de los ámbitos disciplinares.

En tercer lugar, la evaluación de la investigación se enfrenta a las dificultades de diferenciación de dos elementos: por un lado, la calidad de un proyecto o de un trabajo y, por otro, la relevancia del mismo con relación a las prioridades de investigación que se hayan establecido.

En cuarto lugar, el sistema de evaluación de la investigación debe encontrar un mayor equilibrio entre los objetos a ser evaluados (frente al predominio de la evaluación *ex ante* de las propuestas de investigación) y, sobre todo, desarrollar métodos y técnicas complementarias a la tradicional revisión por pares.

Otro de los problemas derivados de la extensión de la evaluación es la sobrecarga de los expertos. La mayoría de las tareas de evaluación de la investigación se suelen ejecutar por un limitado colectivo de los pares, aunque de forma creciente se incorporan expertos en evaluación de programas o en bibliometría. En estas situaciones, la tendencia a la relajación de las prácticas de evaluación y a basar el juicio en elementos secundarios,

como la reputación, son peligros para el sistema. Frente a esto, se plantea como remedio el mantenimiento del anonimato, especialmente cuando se evalúan trabajos o propuestas de proyectos.

Es también esencial la integración de las técnicas de evaluación de la investigación y su desarrollo en el contexto del asesoramiento a la toma de decisiones, en la dirección y gestión del sistema de investigación, que ganaría si se vinculase a la prospectiva o a la evaluación de opciones tecnológicas.

Por último, la transparencia de los procesos de evaluación es un elemento importante, no solamente por la relación entre evaluador, objeto de la evaluación y decisor, sino sobre todo por el efecto de aprendizaje que tiene todo el proceso. En este contexto el derecho a la reclamación y al recurso contra la evaluación es una propiedad esencial del sistema de garantías.

Tras la descripción de estos desafíos comunes, que pueden favorecer una dinámica de convergencia entre las prácticas de evaluación y sus diferentes papeles en los sistemas nacionales de evaluación pasamos, sin ánimo de ser exhaustivo, a revisar algunos problemas específicos de la evaluación de la investigación en España

## **Algunos desafíos para la evaluación de la investigación en España**

Los desafíos generales para la evaluación de la investigación se presentan en España con singularidades y especificidades que merece la pena señalar.

En primer lugar, es necesario eliminar la identificación unívoca entre evaluación *ex ante*, por pares, de los proyectos de investigación con la idea de evaluación de la investigación. El desarrollo sistemático de la evaluación de los resultados de la investigación utilizando técnicas bibliométricas, la puesta en marcha de evaluación de programas de I+D o la utilización de la evaluación como mecanismo para reforzar los procesos de redefinición de las misiones de los institutos o centros de investigación son modalidades que deben alcanzar el adecuado desarrollo en España; incluso la revisión del ámbito de la evaluación individual y su sustitución por la evaluación de colectividades (el grupo de investigación, el departamento, etc.) que refuercen las conductas colectivas o no el individualismo también son importantes.

Se trata de redefinir el papel de la evaluación, su extensión y su coordinación, una vez desaparecida la centralidad o el monopolio institucional de que disfrutó la ANEP hace años, y de repensar la “viabilidad” de la forma organizativa ANEP, dado que en la mayoría de las agencias de financiación de la investigación es la agencia misma la que se encarga de organizar sus necesidades de evaluación. Se ha consolidado en sistema descentralizado, fragmentado y diversificado de evaluación por lo que cabe preguntarse si se debe mantener la singularidad de la ANEP o integrar efectivamente su funcionamiento en la gestión de los programas de financiación de la investigación de la nueva Agencia de Financiación de la Investigación propuesta por el PSOE; no puede olvidarse que la evaluación es meramente instrumental en el sistema de producción de conocimiento que es la ciencia.

Otro desafío al que se enfrenta la práctica de la evaluación de la investigación en España es a su propia examen y evaluación, con el objeto de sacar conclusiones y aprender. Por ejemplo, la ANEP ha funcionado durante 15 años y jamás se ha hecho una evaluación, revisión y auditoria de sus procedimientos. En España las principales críticas que se hacen se siguen refiriendo a la presencia de prácticas poco éticas de los evaluadores en la revisión, de los editores o gestores en la selección de los evaluadores, etc.; sin embargo, no parecen preocupar los temas centrales para la producción de conocimiento que se asocian con los sesgos cognitivos.

La transparencia es un valor esencial de la evaluación de la investigación y la selección de las personas que la lleven a cabo un elemento decisivo. Existen asignaturas pendientes, tales como los procedimientos que se siguen para la nominación de los gestores de programa, de los coordinadores de la ANEP o de los responsables de las Comisiones de la ANECA. La cooptación por la autoridad, a propuesta del saliente, es la práctica habitual que permite el establecimiento de redes de viejos amigos, cuyos problemas han sido señalados en otros países. La normativa administrativa española exige la publicidad y la igualdad de oportunidad de todos los implicados cara a cualquier procedimiento de selección de personal. Un concurso de méritos, con criterios claros, al que se puedan presentar todos los investigadores interesados podría ser un camino. La selección de los responsables de la evaluación no puede estar al margen de la evaluación y del escrutinio público. Otro tema, que debe ser objeto de debate explícito es definir las pautas para escoger a los pares para las evaluaciones: ¿Los más expertos en el campo?, ¿dos pares iguales para que refuercen la opinión o muy diversos?, etc.

Otro problema, en fase de mejoría<sup>4</sup>, es la limitada acumulación de conocimiento codificado sobre los procedimientos y criterios que utilizan por los editores de revistas o los gestores de las políticas en la selección de los pares para el proceso de evaluación. También es necesaria la sensibilización sobre la necesidad de diferenciar los instrumentos a utilizarse en la evaluación por pares, dependiendo del montante de la financiación solicitada, porque la multiplicación de agencias y entidades produce una sobrecarga de evaluación entre los actores que debe ser tratada sabiamente. El esfuerzo de evaluación debe ser proporcional y proporcionado; la utilización de las técnicas de “preselección” puede ser un paso –naturalmente dejando abierta la vía al recurso<sup>5</sup>.

No cabe duda que una mayor profesionalización del sistema de gestión y de evaluación de la investigación ayudaría a aumentar los efectos de retroalimentación de la evaluación sobre el sistema y sobre las políticas.

Por último, la multiplicación de entidades de evaluación, relevantes para el sistema de investigación, exige un esfuerzo mayor de coordinación, de vertebración de un sistema distribuido. Los desafíos están en la coordinación de las bases de datos, en la homogenización de las prácticas, en el intercambio de experiencia de modo que se fomente el aprendizaje. La generalización de las actividades de evaluación de la investigación evidencia que es una actividad que está socialmente distribuida entre los

---

<sup>4</sup> Recientemente la FECYT ha editado un volumen sobre *Modelos de protocolos para la evaluación de actividades de I+D e innovación* (Madrid 2003).

<sup>5</sup> El esfuerzo evaluador no puede ser idéntico para un proyecto que solicita 20.000 euros que para otro que solicita 2 millones.

actores del sistema de investigación; no cabe ya la centralización, se trata de organizar este sistema de inteligencia distribuida lo mejor posible.

### Referencias citadas

- Campanario, J.M. (2002) "El sistema de revisión por expertos (peer review): Muchos problemas y pocas soluciones", *Revista Española de Documentación Científica* vol 25 (3): 166-184.
- Chubin, D.E. y E. Hackett (1991) *Peerless Science. Peer Review and U.S. Science Policy*. Albany (NY): State University of New York Press.
- Cole, S., L. Rubin y J.R. Cole (1978) *Peer review in the National Science Foundation: Phase one of a study*. Washington D.C.: The National Academy of Sciences.
- Cole, S. (1998) "How Does Peer Review Work and Can It be Improved?", *Minerva* vol 36 (XX): 179-189.
- Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott y M. Trow (1994) *The new production of knowledge*. Londres: Sage.
- Kuhn, T.S. (1977) *The essential tension : selected studies in scientific tradition and change*. Chicago-Londres: The University of Chicago Press [v.e. *La tensión esencial: estudios selectos sobre la tradición y el cambio en el ámbito de la ciencia*. México: FCE].
- OECD (1987) *Evaluation of Research. A selection of current practices* (A report by Michael Gibbons and Luke Georghiou). Paris: OECD.
- Rip, A. (1994) "The Republic of Science in the 1990s", *Higher Education* vol 28 (1), 3.-23. [v.e. "La república de la ciencia en los años noventa", en L. Sanz y M. J. Santesmases eds., número especial "Ciencia y Estado", *Zona Abierta* 75/76, 1996: 57-89].
- Sanz-Menéndez, L. (1995) " Research actors and the state: research evaluation and evaluation of science and technology policies in Spain", *Research Evaluation* vol. 5 (1): 79-88.
- Sanz Menéndez, L. (1997) *Estado, ciencia y tecnología en España (1939-1997)*. Madrid: Alianza Editorial, 427 pp.
- Zuckerman, H. y R.K. Merton (1971) "Patterns of Evaluation in Science. Institutionalisation, Structure and Functions of Referee System", *Minerva* vol 9 (1): 66-100. [v.e. en R.K. Merton *La Sociología de la Ciencia. Investigaciones teóricas y empíricas*. Madrid: Alianza, 1977, vol. 2: 579-621].