

La evaluación de la ciencia y la investigación

Evaluation of science and research

LUIS SANZ MENÉNDEZ

Instituto de Políticas y Bienes Públicos (IPP)
Consejo Superior de Investigaciones Científicas (CSIC)
Luis.Sanz@csic.es

Desde la dirección de la *Revista Española de Sociología* se me ha pedido que prepare una contribución sobre la evaluación de la ciencia y la investigación. Dado que los editores me conocen, e incluso pueden haber leído algunos de mis trabajos (Sanz-Menéndez, 2004; Sanz-Menéndez y De Moya, 2010), es razonable que sus expectativas sobre este artículo estén asociadas a la defensa de una posición favorable a la evaluación de la investigación.

El objeto de este artículo es, por un lado, analizar la centralidad de las diversas formas de evaluación en el proceso de producción, comunicación y difusión del conocimiento y, al mismo tiempo, llamar la atención sobre algunos de los problemas que este método imperfecto puede generar en el funcionamiento de la ciencia y, por tanto, plantear algunos de los desafíos actuales. Si se me permite una metáfora (o quizá una hipérbole), la evaluación es a la ciencia como la democracia a nuestra sociedad: sabemos que es un sistema de gobierno imperfecto, que tiene y que genera problemas en su aplicación, que es mejorable en su práctica en muchos países y organizaciones, pero lo cierto es que, y quizá es una asunción normativa, no podemos hacer ciencia sin la evaluación de la misma, porque sin ella no sería tal.

La estructura de este trabajo es la siguiente: en primer lugar se realiza un análisis del surgimiento de los procesos de evaluación en la actividad científica moderna, esencialmente asociados a la revisión por pares (*peer review*) como sistema de control de calidad y su adopción posterior como sistema de asignación de recursos. Se contextualiza la práctica de la evaluación en el contexto de los sistemas de I+D y se hace referencia a los diversos niveles en los que se practica. También se plantea la existencia de atributos nacionales en los sistemas de evaluación de la investigación asociados a la singularidad de las instituciones de I+D, así como de presiones recientes derivadas de los indicadores bibliométricos y de los *rankings* como sistemas de evaluación. En la sección cuarta, se repasan algunas instituciones españolas de evaluación de la investigación. Finalmente, se presentan algunos de los desafíos a los que se enfrenta la práctica de la evaluación de la investigación.

SURGIMIENTO Y DESARROLLO DE LA EVALUACIÓN (LA REVISIÓN POR PARES)

El origen de la evaluación de la investigación se asocia a los procedimientos y prácticas establecidos por la Royal Society en Reino Unido. El momento “fundacional” de esta práctica institucionalizada, en forma de revisión por pares expertos (*peer review*), 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); esto es, surgió como un sistema de control de calidad.

En un proceso evolutivo, y ya en el siglo XX, la evaluación quedó asociada al método científico y pasó a extender su campo de acción: primero a utilizarse como procedimiento de asignación de fondos para la investigación y posteriormente para evaluar otros objetos, como los programas, los resultados de las organizaciones o los investigadores en el contexto de sus carreras.

Sin embargo, la expansión de la práctica, los objetos y los métodos de la evaluación no ha estado exenta de tensiones y polémicas. A finales de los setenta, en Estados Unidos existía una cierta preocupación por los procedimientos de evaluación utilizados en la asignación de fondos de investigación por la National Science Foundation (NSF) o los National Institutes of Health (NIH), como demuestran las revisiones realizadas de esas prácticas (Cole *et al.*, 1978); incluso algunos afirmaban (Chubin y Hackett, 1991) que la revisión por pares 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.

El interés por la evaluación de la investigación se ha extendido y continúa atrayendo la atención de los gobiernos y de las agencias de financiación. La preocupación por los métodos de evaluación de la investigación propició el desarrollo de trabajos (OECD, 1987) que contribuyeron a ampliar la perspectiva y a introducir elementos que conectaban la forma tradicional 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.

El interés de la comunidad científica sobre los procesos de evaluación en las revistas académicas (Campanario, 2002) o la utilización de las métricas y de la bibliometría (Nederhof, 2006) no han hecho sino crecer, alimentados por la dinámica competitiva de la ciencia. Adicionalmente, otros factores externos presionan y, en el contexto de los efectos de la crisis, se han añadido dimensiones al debate sobre los resultados y el valor del gasto en investigación (“*value for money*”) y la necesidad de responder a la pregunta de qué obtienen los ciudadanos de los recursos públicos que se asignan a las actividades de I+D. Otra tendencia de interés, con efectos en nuestro concepto de evaluación, se refiere al impacto de la sociedad de la información y el acceso a ingentes cantidades de información que permiten la construcción de indicadores cuantitativos y sistemas de *rankings* o *ratings*, tanto de instituciones (Sanz-Menéndez y De Moya, 2010) como de personas.

LA EVALUACIÓN EN SU CONTEXTO

Así pues, 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.

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 han 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, la promoción académica, e incluso para el otorgamiento de premios, como el Nobel (Zuckerman, 1977).

Además del control de calidad, desde los inicios, la evaluación del mérito por parte de los pares 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 más 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. Así pues, 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.

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, consejos de investigación, etc.), se produjo la generalización de la evaluación de las propuestas de financiación. La revisión por pares, en su forma pura o modificada, se convirtió en una práctica ordinaria para la atribución de fondos para investigación.

La revisión por pares es el nombre que se da al juicio del mérito científico por otros científicos que trabajan en o cerca del campo en cuestión. La premisa es que el juicio sobre ciertos aspectos de la ciencia, por ejemplo, su calidad, es una decisión experta que solamente pueden tomar aquellos que tienen suficiente conocimiento sobre el desarrollo cognitivo del campo, su agenda de investigación y sus practicantes dentro del mismo (OECD, 1987).

El método se basa en la determinación del mérito de las propuestas y es la propia comunidad científica la que da sus opiniones y asesora 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). Con el paso de los años, se han desarrollado nuevas formas de intervención estatal en la política de I+D: la investigación por objetivos, la investigación estratégica,

la investigación priorizada, etc. Se han consolidado 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, para la mejor utilización de los recursos.

Además, el desarrollo de programas de I+D 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; se trata de la evaluación de resultados o de impacto. En definitiva, la multiplicidad de objetivos para las actividades de producción científica y, por tanto, la diversidad de criterios de evaluación, crea un contexto más complejo para el desarrollo de la misma.

Aunque la evaluación de la investigación gira en torno a dos ámbitos fundamentales: a) el control de calidad del proceso de producción de nuevo conocimiento y de sus resultados y efectos y b) la asignación de recursos, tanto monetarios, como de recompensas y reputacionales. Sin embargo, hoy en día, la evaluación de la investigación implica una concepción más amplia e integradora que supere la tradicional identificación de la misma con la revisión por pares, dado que esta es solo una de las posibles técnicas a ser utilizadas, siendo otras los trabajos bibliométricos, las encuestas, los modelos econométricos, etc., cuya utilización se ha desarrollado con fuerza en los últimos años.

LAS NUEVAS TENDENCIAS Y LA PRESIÓN DE INDICADORES CUANTITATIVOS Y RANKINGS

Uno de los procesos que más está afectando a la evaluación de la investigación es la consolidación de los indicadores cuantitativos (bibliométricos y de otro tipo) y de los *rankings* institucionales y personales, como métodos de medida de la reputación y de la calidad.

La aplicación de indicadores cuantitativos y especialmente la de los *rankings* (por ejemplo: ARWU Academic Ranking of World Universities, THE-Time Higher Education, QS-World University Ranking, Leiden Ranking, SIR-Scimago Institutions Rankings, Goggle Scholar citations, Research Gate, etc.) es una industria creciente. *Rankings* de universidades, de programas doctorales o de investigadores, casi siempre asociados con la idea de calidad, se han hecho populares e influyen en diversos públicos y audiencias: a los estudiantes que van a elegir dónde estudiar, a los gobiernos que financian, etc.

Los *rankings* son muy populares entre los gobiernos y los medios de comunicación, pero con frecuencia los responsables universitarios y los académicos los examinan como medidas muy pobres de los resultados, de la calidad o de la reputación. A pesar de las objeciones y limitaciones señaladas (Weingart, 2005), los *rankings*, una vez que se difunden, devienen, naturalizados y transforman el entorno de las organizaciones (y de los académicos), porque afectan a la reputación de sus instituciones (e investigadores).

Los *rankings* no sustituyen otras formas de evaluación, como la revisión por pares, pero es indudable que se han convertido en señales de calidad en un entorno de la investigación que es mundial y donde, universidades, centros de investigación e investigadores quieren estar en buenas posiciones en estos *rankings* o en las redes sociales. Hasta su proliferación, las instituciones seguían diversos procedimientos para crear su propia reputación y posicionarse en el sistema de estatus nacional e internacional. La reputación institucional y académica era más dependiente de las opiniones de los profesionales y colegas de mayor estatus; el reconocimiento y el sistema de estatus estaba basado en una agregación no sistemática de la reputación y el mérito.

Por definición, el estatus es un bien posicional que necesariamente es comparativo, relativo y recíproco; son las comparaciones las que construyen el sistema de estatus y también otorgan valor simbólico a las organizaciones (e investigadores). En las universidades y en la investigación, las comparaciones de calidad son un elemento central de la medición, porque en estos contextos la información sobre la reputación, la productividad o los resultados son difíciles de observar, medir e interpretar (Sauder y Espeland, 2009).

El poder de los números y de los *rankings* reside en que, a pesar de las llamadas a la prudencia en su uso e interpretación, hacen el estatus explícito y tienen consecuencias: crean una jerarquía formal que determina la posición de los que han sido medidos. Haciendo públicos sus juicios sobre el estatus, los *rankings* llevan a las instituciones e investigadores a ser sensibles a sus posiciones y a actuar en consecuencia. Además, imponiendo una métrica compartida, aunque sea parcial, los *rankings* contribuyen a crear o a unificar el campo organizativo, ya se refiera a universidades o centros de investigación, y producen presiones isomórficas. Finalmente, los *rankings* tienen el efecto de crear “buenas” y “malas” reputaciones.

Así pues, la capacidad de universidades, instituciones e investigadores de crear una reputación en base a valores o criterios distintos de esos que se usan en la construcción de *rankings* es crecientemente limitada, porque las evaluaciones de terceros tienen un efecto superior que las autoevaluaciones. Ya hay evidencias (Sauder y Lancaster, 2006) de que la introducción de los *rankings* institucionales ha alterado la estructura del sistema de estatus e incluso los valores (las medidas) que constituyen esos sistemas.

Los indicadores bibliométricos tienen algunos problemas y su uso para evaluar la calidad ha sido muy criticado, especialmente si se separan de la revisión por pares; sin embargo, cuando se los compara con las estrategias basadas en las consultas, la bibliometría tiene la ventaja de manejar un gran número de datos y hechos (publicaciones y citas). Se han señalado diversos problemas de estos indicadores, que se derivan de la cobertura de las revistas en las bases de datos, los problemas metodológicos de sesgo del país o disciplina, de los sesgos de las citaciones hacia Estados Unidos o la menor representación de otros idiomas distintos del inglés (van Raan 2005), o los límites específicos del uso de indicadores bibliométricos en las ciencias sociales (Clements *et al.*, 1995; Hicks, 1999; Nederhof, 2006).

Así pues, la evaluación puede verse hoy como un método (o mejor un proceso con diversos métodos) para proveer información de carácter evaluativo sobre algunos aspectos relacionados con la investigación y sus diversos objetos a partir de algunos criterios u opiniones (que van desde la excelencia investigadora a los efectos socioeconómicos).

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 y a sus actores, dotándoles de mayor coherencia y facilitándoles el aprendizaje continuo; todo ello además de favorecer la transparencia y la rendición de cuentas.

MODELOS NACIONALES DE EVALUACIÓN DE LA INVESTIGACIÓN Y SU INSTITUCIONALIZACIÓN

La investigación es una actividad internacional, sin embargo, las prácticas de evaluación de la investigación están integradas en el funcionamiento de los sistemas nacionales de I+D, por lo que existe una significativa variedad de arreglos institucionales y organizativos en los que estas actividades se desarrollan en los diversos países, así como de los objetos que se evalúan, de los criterios que se utilizan y de las consecuencias que tienen; es por ello que se habla de sistemas nacionales de evaluación de la investigación [véase al respecto el monográfico de *Research Evaluation*, 5 (1), April 1995].

Por ejemplo, en el Reino Unido, la evaluación de la investigación se ha convertido en un mecanismo esencial de distribución de los recursos para la financiación de I+D en las universidades, a través del denominado *Research Assessment Exercise (RAE)*, que en 2014 continuará con el *Research Excellence Framework (REF)*, que establece la asignación financiera que recibe cada universidad en el contexto de una evaluación de resultados anteriores, llevada a cabo por paneles. Modelos similares de financiación de las universidades se encuentran, por ejemplo, en Australia, donde cuentan con mayor peso los indicadores cuantitativos.

En Francia, la práctica más extendida es la que utiliza 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 resaltan el interés por evaluar los efectos socioeconómicos de la investigación o del funcionamiento del conjunto del sistema de innovación, como es el caso de Finlandia.

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 socioeconómicos. En general, 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 a los expertos y de designar los paneles de selección.

En España, en los ochenta y 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 administrativa singular que se responsabilizaba de la evaluación de las actividades de investigación (la Agencia Nacional de Evaluación y Prospectiva

—ANEP—) (Fernández de Caleyá, 2001) y que se debía encargarse de organizar de modo relativamente autónomo y 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. Sin duda, la creación de la ANEP resolvió algunos problemas de credibilidad y legitimación de la distribución de fondos públicos competitivos a los investigadores.

Posteriormente, se institucionalizaron otros mecanismos de evaluación de resultados, como la Comisión Nacional de Evaluación de la Actividad Investigadora (CNEAI) (Cruz-Castro y Sanz-Menéndez, 2007; Osuna, Cruz-Castro y Sanz-Menéndez, 2011), o de acreditación de los profesores de universidad y de evaluación de la calidad universitaria, con la instauración de la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) para el conjunto de España, así como de entidades similares de evaluación y de financiación de la investigación en las comunidades autónomas [por ejemplo: la Agència de 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 (ACAP), etc.].

Así pues, de una situación, a finales de los ochenta, caracterizada por el “cuasimonopolio” de la evaluación de las actividades de la investigación por parte de la ANEP, solamente acompañada de la CNEAI (encargada de los sexenios), se ha pasado a una multiplicación de organismos de gestión de la evaluación 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.

Aunque 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), más recientemente el debate y el uso de indicadores bibliométricos y rankings ha pasado a ser cotidiano (Chinchilla et al., 2013; Corera et al., 2013); lo cierto es que, tras unas décadas, la evaluación se ha institucionalizado, se ha normalizado en todas las áreas de conocimiento y disciplinas y, también, se ha burocratizado.

ALGUNOS DESAFÍOS PARA LA EVALUACIÓN DE LA INVESTIGACIÓN

Los sistemas de I+D y la práctica de la evaluación de la investigación a escala internacional se enfrentan a una serie de desafíos (Langfeldt, 2006) que se mueven en diversos niveles: la dinámica social que hace de la evaluación un mecanismo de control de la comunidad científica, una perspectiva asociada a la incertidumbre de juzgar la calidad de la investigación, especialmente cuando es ex ante, y una perspectiva organizativa relacionada con las diversas consecuencias de cómo se organiza la evaluación y que instrumentos se utilizan.

Por concretar, 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 control de la calidad de la investigación debe ser capaz de evitar que su aplicación frene el desarrollo de ideas innovadoras en la ciencia. En segundo lugar, la creciente interdisciplinariedad de la actividad investigadora. Una de las características del nuevo modo de producción del conocimiento (Gibbons *et al.*, 1994) ha llevado a que uno de los desafíos que el sistema tiene ante sí sea el control de calidad de la investigación generada fuera de los ámbitos disciplinares. En tercer lugar, la evaluación de la investigación como actividad se enfrenta a las dificultades de diferenciar 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 la sociedad legítimamente haya establecido.

En otro orden de cosas, 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*) y, sobre todo, desarrollar métodos y técnicas complementarias a la tradicional revisión por pares, que permitan enriquecer sin caer en la simplificación del proceso que parecen entrañar algunos *rankings*.

En lo práctico, la expansión de la evaluación genera la sobrecarga de los expertos. La mayoría de las tareas de evaluación de la investigación suelen ser ejecutadas por un limitado colectivo de pares, aunque de forma creciente se utilizan instrumentos como los bibliométricos. 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.

Representan también un desafío 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, así como la transparencia de los procesos de evaluación, 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.

Tampoco puede olvidarse que, a pesar de la existencia de evaluación, se descubren actos de fraude científico (Abbot, 2008) y plagio (Fanelli, 2009). Así pues, la evaluación no previene e inmuniza al sistema científico contra la corrupción. La existencia de reglas éticas y del escrutinio por parte de los pares es imprescindible.

Por último, la generalización de la bibliometría está siendo objeto de críticas, especialmente cuando la utilización de elementos como el factor de impacto (Baum, 2011) de las revistas en que se publica pasa a ser el elemento central de evaluación de las carreras individuales, generando los incentivos equivocados (Stephan, 2012). Esto ha llevado a declaraciones formales, como la San Francisco Declaration on Research Assessment – DORA (<http://am.ascb.org/dora/>) sobre los riesgos de valorar a los investigadores en sus carreras por el factor de impacto de las revistas en que publican.

En las prácticas de la acreditación de la ANECA, estamos lejos de estos excesos; es más, es probable que en algunas disciplinas como la sociología, la utilización de la revista en la que se ha publicado, como criterio de evaluación y reputación, está subutilizada frente a otros criterios de adscripción o de cumplimiento burocrático.

Es necesario evitar los desplazamientos pendulares, que se pase de un rechazo a la evaluación de la investigación a una identificación unívoca entre evaluación *ex ante*, por pares,

o a una identificación exclusiva de la evaluación con los indicadores bibliométricos. La evaluación por pares informada con indicadores bibliométricos está en el horizonte. Una debilidad es la limitada relevancia de la evaluación de resultados de las organizaciones e instituciones y la casi nula consecuencia de la misma, lo que dificulta las estrategias de acción colectiva de sus miembros.

Se trata también 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 también quizá de repensar su “viabilidad”, dado que en la mayoría de las agencias de financiación de la investigación es la agencia misma la que se encarga de organizar y gestionar sus necesidades de evaluación con la ayuda de expertos o científicos. Además, se ha consolidado en sistema descentralizado, fragmentado y diversificado de evaluación, pero no puede olvidarse que la evaluación es meramente instrumental en el sistema de producción de conocimiento que es la ciencia.

También es necesaria la sensibilización sobre la diferenciación de 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, dado que a veces cuesta más la evaluación que los fondos que se otorgan.

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, así como a fortalecer la limitada acumulación de conocimiento codificado sobre los procedimientos y criterios que se utilizan por los editores de revistas, los gestores de las políticas de financiación, o los evaluadores de personas o instituciones en el proceso de evaluación (FECYT, 2003).

Otro desafío al que se enfrenta la práctica de la evaluación de la investigación en España es su propio examen y evaluación, con el objeto de sacar conclusiones y aprender. En España las principales críticas que se hacen a la práctica de la evaluación se refieren 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 programas de I+D, de los coordinadores de la ANEP o de los responsables de las comisiones de la ANECA. La cooptación por la autoridad administrativa, 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 oportunidades 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. Otros temas son 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.

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

En definitiva, el principal desafío es cómo mejorar la calidad de la evaluación y su institucionalidad porque es la pieza clave de la credibilidad de la investigación.

REFERENCIAS BIBLIOGRÁFICAS

- Abbott, A. (2008), "The fraudster returns", *Nature*, 452: 672, 10 april 2008.
- Baum, J. A. C. (2011), "Free-riding on power laws: questioning the validity of the impact factor as a measure of research quality in organization studies", *Organization*, 18: 449-466.
- Campanario, J. M. (2002), "El sistema de revisión por expertos (peer review): Muchos problemas y pocas soluciones", *Revista Española de Documentación Científica*, 25 (3): 166-184.
- Clemens, E. S., Powell, W. W., McIlwaine, K. y Okamoto D. (1995), "Careers in Print: Books, journal, and Scholarly Reputations", *American Journal of Sociology*, 101 (2): 433-494.
- Cole, S., Rubin, L. y Cole, J. R. (1978), *Peer review in the National Science Foundation: Phase one of a study*, Washington D. C., The National Academy of Sciences.
- Corera, E., Chinchilla, Z., de Moya, F. y Sanz-Menéndez, L. (2013), "Producción científica y excelencia con liderazgo: Ranking general y por áreas de las instituciones universitarias españolas (2007-2011)", en *Informe CyD 2012. Fundación Conocimiento y Sociedad*, Barcelona, Fundación CYD: 236-242.
- Cruz-Castro, L. y Sanz-Menéndez, L. (2008), "Research Evaluation in Transition: Individual versus Organisational Assessment in Spain", en Whitley, R. y Gläser, J. (eds.) (2008), *The Sociology of the Sciences Yearbook [The Changing Governance of the Sciences. The Advent of the Research Evaluation Systems]*, 26: 205-223.
- Chinchilla, Z., Corera, E., de Moya, E. y Sanz-Menéndez, L. (2013), "La producción científica española en el contexto internacional y la posición de sus instituciones de investigación en el ranking mundial" en *Informe CyD 2012. Fundación Conocimiento y Sociedad*, Barcelona, Fundación CYD: 243-252.
- Chubin, D. E. y Hackett, E. (1991), *Peerless Science. Peer Review and U.S. Science Policy*, Albany (N. Y.), State University of New York Press.
- Fanelli, D. (2009), "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data", *PLoS ONE* 4 (5), e5738. <http://dx.doi.org/10.1371/journal.pone.0005738> Acceso 30 diciembre 2013.
- FECYT (2003), *Modelos de protocolos para la evaluación de actividades de I+D e innovación*, Madrid, FECYT.
- Fernández de Caleyá, R. (2001), "Los comienzos de la evaluación científica en España", *QUARK, Ciencia, Medicina, Comunicación y Cultura*, 22-23: 26-28.

- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. y Trow, M. (1994), *The new production of knowledge*, London, Sage.
- Hicks, D. (1999), "The difficulty of achieving full coverage of international social science literature and the bibliometric consequences", *Scientometrics*, 44 (2): 193-215.
- 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].
- Langfeldt, L. (2006), "The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments", *Research Evaluation*, 15 (1): 31-41.
- Nederhof, A. J. (2006), "Bibliometric monitoring of research performances in Social Sciences and the Humanities: A review", *Scientometrics*, 66 (1): 81-100.
- OECD (1987), *Evaluation of Research. A selection of current practices* (A report by Michael Gibbons and Luke Georghiou), Paris, OECD.
- Osuna, C., Cruz-Castro, L. y Sanz-Menéndez, L. (2011), "Overturning some assumptions about the effects of evaluation systems on publication performance", *Scientometrics*, 86 (3): 575-592.
- Rip, A. (1994), "The Republic of Science in the 1990s", *Higher Education*, 28 (1): 3-23. [v. e. "La república de la ciencia en los años noventa", en L. Sanz y M. J. Santesmases (eds.), 1996, número especial "Ciencia y Estado", *Zona Abierta*, 75-76: 57-89].
- Sanz-Menéndez, L. (1997), *Estado, ciencia y tecnología en España (1939-1997)*, Madrid, Alianza Editorial: 427.
- Sanz-Menéndez, L. (2004), "Evaluación de la investigación y sistema de ciencia", *Boletín de la Sociedad Española de Bioquímica y Biología Molecular (SEBBM)*, 140: 6-10.
- Sanz-Menéndez, L. y de Moya, F. (2010), "A new industry: University rankings in the social sciences" en (2010), *World Social Science Report*, París, UNESCO: 242-244.
- Sanz-Menéndez, L. (1995), "Research actors and the state: research evaluation and evaluation of science and technology policies in Spain", *Research Evaluation*, 5 (1): 79-88.
- Sauder, M. y Espeland, W. N. (2009), "The Discipline of Rankings: Tight Coupling and Organizational Change", *American Sociological Review*, 74 (1), 63-82.
- Sauder, M. y Lancaster, R. (2006), "Do Rankings Matter? The Effects of U.S. News & World Reports Rankings on the Admission Process of Law Schools", *Law & Society Review*, 40 (1): 105-134.
- Stephan, P. (2012), "Perverse incentives", *Nature*, 484: 29-31, 5 april.
- Van Raan, A. F. J. (2005), "Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometrics methods", *Scientometrics*, 62 (1), 133-143.
- Weingart, P. (2005), "Impact of bibliometrics upon the science systems: inadvertent consequences?", *Scientometrics*, 62 (1), 117-131.
- Zuckerman (1977), *Scientific Elite. Nobel laureates in the United States*, New Brunswick (N. J.), Transactions, 1996 (new edition).
- Zuckerman, H. y Merton R. K. (1971), "Patterns of Evaluation in Science. Institutionalisation, Structure and Functions of Referee System", *Minerva*, 9 (1): 66-100. [v. e. en R. K. Merton, 1977, *La Sociología de la Ciencia. Investigaciones teóricas y empíricas*, Madrid, Alianza; 2: 579-621].

Luis Sanz Menéndez es doctor en Sociología (UCM), profesor de investigación del CSIC y, desde 2004, director del Instituto de Políticas y Bienes Públicos (IPP) del CSIC. Desde octubre de 2007 es

presidente del Comité de Política Científica y Tecnológica (CSTP) de la OCDE. Con anterioridad ha sido profesor en la Facultad de Ciencias Políticas y Sociología de la UCM, así como investigador y profesor visitante en diversas universidades como la de California en Berkeley, la de Georgia en Athens, la de Colorado en Denver, la de Rutgers en Newark o el Instituto Tecnológico de Georgia en Atlanta. Ha publicado más de un centenar de trabajos de investigación sobre sistemas de I+D, políticas de ciencia, tecnología e innovación, carreras académicas y la evaluación de la investigación en revistas internacionales, amén de *Estado, ciencia y tecnología en España (1939-1997)*, Madrid, Alianza Universidad, 1997. Más información en <http://www.ipp.csic.es/personal/luis.sanz>