

Evaluación de la investigación: el declive de un modelo exitoso

Evaluating research: the decline of a successful model

VERA SACRISTÁN ADINOLFI
Universitat Politècnica de Catalunya
vera.sacristan@upc.edu

EL DESPEGUE DE LA INVESTIGACIÓN EN ESPAÑA

Junto con el Consejo Superior de Investigaciones Científicas (CSIC) y demás organismos públicos de investigación (OPIs), las universidades constituyen hoy el núcleo principal de la investigación científica y la generación de conocimiento en España. Su innegable despegue se produjo a partir de los años ochenta y a él contribuyeron diversos factores.

La promulgación en 1983 de la Ley de Reforma Universitaria (LRU) comportó, de entrada, un cambio organizativo imprescindible en las universidades: desaparecieron las cátedras y se crearon los departamentos, agrupaciones mucho más amplias en función de las áreas de conocimiento. Se pasó así de una organización directamente vinculada a las materias docentes a una organización estructurada en función de los conocimientos científicos del profesorado. Simultáneamente, la LRU consolidó el requisito del doctorado para todo el profesorado permanente de las facultades y escuelas técnicas superiores, lanzando así un mensaje claro sobre la importancia de la investigación. Este cambio se completó con un nuevo diseño de los concursos-oposición que, más allá de todos los defectos que resultó tener, preveía que las plazas se convocaran con un “perfil” investigador y que en los concursos la actividad investigadora de las personas candidatas tuviera un peso mayoritario. Casi simultáneamente, la promulgación de la Ley de la Ciencia en 1986 fue de la mano del crecimiento de las convocatorias para la financiación de proyectos de investigación y de las plantillas de personal investigador. Poco después, un real decreto implantó en 1989 un complemento de productividad basado en la evaluación de la actividad investigadora del profesorado y el personal investigador del CSIC (los comúnmente llamados sexenios).

Concursos, sexenios y proyectos de investigación comportaron la puesta en marcha de distintos sistemas de evaluación de la actividad investigadora. Por parte de las universidades en el caso de los concursos, por parte de la administración con competencias en investigación

en el caso de los sexenios y de los proyectos. Evaluación personal en el caso de los concursos y sexenios, evaluación colectiva de los grupos de investigación en el caso de los proyectos.

EVOLUCIÓN DE LA EVALUACIÓN DE LA INVESTIGACIÓN EN ESPAÑA

Desde entonces, y durante dos décadas, la evaluación de la investigación corrió a cargo de las comisiones creadas *ad hoc* para resolver los concursos de profesorado y personal investigador, efímeras en el tiempo por naturaleza, y de dos entidades de carácter estable: la Agencia Nacional de Evaluación y Prospectiva (ANEP) y la Comisión Nacional Evaluadora de la Actividad Investigadora (CNEAI). Aún hoy, la ANEP evalúa los proyectos de investigación que solicitan financiación pública, y la CNEAI hace lo propio con la actividad de investigación del profesorado universitario y de las escalas científicas del CSIC para la concesión de los complementos de productividad.

La ANEP evalúa las solicitudes de los proyectos a partir de los informes de los investigadores principales de estos, que incluyen justificaciones, planes de trabajo y relaciones de publicaciones. Por su parte, la CNEAI evalúa la actividad de investigación de carácter personal a partir de cinco aportaciones —en general, publicaciones en revistas científicas— seleccionadas por cada solicitante en función de su relevancia.

Pero en los años transcurridos desde su creación, han aparecido otros agentes que también evalúan la investigación en diversos contextos. La Ley Orgánica de Universidades de 2001 otorgó a la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) y a las agencias análogas de las comunidades autónomas la competencia de acreditar (requisito imprescindible para poder optar a una plaza) para profesorado universitario contratado y, en particular, de evaluar la actividad investigadora de las personas candidatas. Posteriormente, la Ley Orgánica de Modificación de la Ley Orgánica de Universidades de 2007 otorgó a la ANECA la competencia de acreditar también al profesorado funcionario. Comisiones específicas evalúan a su vez los currículos de las personas candidatas a las convocatorias de personal investigador creadas por la reforma de 2001 de Ley de la Ciencia y su desarrollo posterior (programas Ramón y Cajal, Juan de la Cierva...). Esta no es más que una relación incompleta de la situación, en la que ni siquiera se han mencionado las evaluaciones que se llevan a cabo en el ámbito europeo (proyectos de investigación, contratación de personal investigador...) o en el autonómico.

La ANECA y las agencias autonómicas llevan a cabo su evaluación sobre la base de los currículos presentados por las personas solicitantes mediante sistemas de puntuaciones, esto es, con un importante componente métrico. Con la intención, lograda o no, de que el proceso sea más fácil para los evaluadores y más justo para los evaluados, los currículos siguen un formato único o, más exactamente, único para cada agencia evaluadora y objetivo de evaluación. Estos formatos convierten los currículos en una profusa enumeración de publicaciones, comunicaciones a congresos, proyectos financiados y demás actividades relacionadas con la investigación, en un modelo bastante distinto de la selección de aportaciones de la CNEAI y de la exposición de motivos y objetivos de la ANEP.

VENTAJAS E INCONVENIENTES

La combinación de todos estos elementos —nueva organización, aumento de la financiación y de las plantillas, evaluación— ha fomentado la investigación en nuestro país y su normalización en comparación con el resto de países del mundo. A este respecto, el lector interesado puede consultar la base de datos sobre la actividad científica que ofrece IUNE¹ o los trabajos de análisis del IVIE².

Algunos estudios muestran que lo que facilitó el sustancial despegue de la investigación en España y su actual normalización no fue tanto la evaluación de la investigación como su financiación³. Sin embargo, todos coinciden en que las políticas de evaluación condicionan la orientación y las actividades del personal investigador. Y la evaluación de la investigación está tomando una deriva que debe ser analizada con cuidado.

De entrada salta a la vista la gran cantidad de agentes y actos evaluadores, y lo primero que cabe preguntarse es si tiene sentido semejante proliferación. Pero hay muchas más cuestiones de interés a plantear: particularmente, cómo se llevan a cabo estas evaluaciones, con qué consecuencias y si podría hacerse mejor.

Una parte sustancial de la evaluación, tanto de los grupos de investigación como de las personas individualmente, se lleva a cabo analizando currículos. En otras palabras, quienes llevan a cabo las evaluaciones leen la documentación que aportan las personas que se someten a evaluación sin que se produzca interacción alguna entre evaluadores y evaluados. Solo en los concursos de acceso a las plazas de profesorado y de personal investigador las personas candidatas deben exponer ante las comisiones que los resuelven —y debatir con estas— el alcance de sus trabajos y sus propuestas. Todo lo demás es evaluación por “papeles”. Esto se explica fácilmente: dada la proliferación de actos evaluadores, resulta materialmente imposible resolver la cuestión de otro modo. Ahora bien, además de explicarse por motivos prácticos, ¿se justifica también desde el punto de vista científico?

Con los años y, más acusadamente, con la aparición de la ANECA y las agencias autonómicas, se ha ido generando una clara tendencia hacia la cuantificación. Quienes investigan llevan años percibiendo, en mayor o menor medida, un mensaje según el cual lo que deben hacer es publicar cuantos más artículos mejor en revistas científicas, indexadas, por supuesto. Esta cuantificación del proceso evaluador ha permitido, a su vez, su burocratización progresiva. Indicadores numéricos, índices de citas..., poco a poco vamos caminando hacia una objetivación de la evaluación. Este sistema progresivamente cuantificado de evaluación está ampliamente aceptado en diversos ámbitos, particularmente en el de ciencias donde, en líneas generales, los resultados tienden a corresponder con la opinión subjetiva que la comunidad científica tiene de las personas y grupos evaluados. Asimismo, sistemas de indicadores como el de la ANECA y otros semejantes tienen ventajas que no deben menospreciarse.

¹ Accesible en <http://www.iune.es/>.

² Véanse, por ejemplo: http://www.ivie.es/es/actividades/noticias/2012/ws_uup01.php o también <http://www.ivie.es/en/actividades/noticias/2013/rankings-universidades-espanolas.php>.

³ Osuna C., Cruz-Castro L., Sanz-Menéndez, L. (2011), “Overturning some assumptions about the effects of evaluation systems on publication performance”, *Scientometrics*, 86: 575-592.

Por un lado, permiten lanzar mensajes claros que orientan el trabajo del profesorado y el personal investigador, esto es, permiten que todo el mundo sepa *a priori* lo que se espera de su actividad. Asimismo, con este tipo de evaluaciones, una persona con méritos investigadores suficientes puede tener la casi certeza de ser evaluada positivamente (aunque el recíproco no siempre sea cierto). Finalmente, y esto también es muy importante, una evaluación de carácter marcadamente métrico ofrece un cierto nivel de seguridad jurídica, cuestión para nada despreciable.

Sin embargo, el sistema actual ha acabado por provocar también una serie de efectos negativos que es manifiestamente necesario corregir.

En primer lugar, por su importancia, está la proliferación de publicaciones, el troceamiento de los resultados de la investigación y la escasez de trabajos de alcance amplio, realmente iluminadores. En pocas palabras, el sistema está tendiendo a la cantidad, más que a la calidad y la relevancia de la investigación. Este es un fenómeno contrario al interés general del avance científico.

Por otro lado, cabe preguntarse sobre el acierto de un modelo que pone tanto énfasis en las revistas indexadas y las citas. En particular, a pesar de los matices entre unos ámbitos y otros, comporta una gran dosis de uniformización entre estos, que ha sido muy discutida en algunos casos, particularmente en humanidades y ciencias sociales, ámbitos en los que algunos estudios revelan que las revistas indexadas recogen menos de una cuarta parte de las publicaciones totales⁴. Pero no es mi intención reproducir aquí el debate sobre las ventajas e inconvenientes de la bibliometría en la evaluación de la investigación, al que se han dedicado y se siguen dedicando revistas especializadas enteras⁵, números especiales de revistas científicas, especializadas o no en temas de política científica⁶, libros, congresos y un sinnúmero de debates, y que producen vaivenes, incluso contradictorios, en los sistemas de evaluación de la investigación en países tan afines en muchos sentidos como Gran Bretaña y Australia⁷. Es más, a mi entender, el debate que enfrenta a partidarios de la bibliometría con partidarios de la revisión por pares tiene ciertas dosis de confusión: al fin y al cabo, la bibliometría se basa en publicaciones en las que la selección de los trabajos se hace por pares, y la evaluación por pares se centra, hoy en día, en el análisis de información

⁴ L. Butler, I. McAllister, "Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science", *Political Studies Review*, 7: 3-17, 2009.

⁵ Especializadas en evaluación de la ciencia, como Research Evaluation (<http://rev.oxfordjournals.org/>), o en medición de esta, como Scientometrics (<http://link.springer.com/journal/11192>), entre muchas otras.

⁶ Se hizo famoso el debate promovido por "Nature: Peer review debate", *Nature*, junio 2006 (<http://www.nature.com/nature/peerreview/debate/>), pero ha seguido habiendo muchos otros en revistas científicas de diversos ámbitos. *Science and Public Policy*, 34 (8), 2007: <http://spp.oxfordjournals.org/content/34/8.toc> y *Political Studies Review*, 7(1), 2009: <http://onlinelibrary.wiley.com/doi/10.1111/psr.2009.7.issue-1/issuetoc>, son sólo dos ejemplos.

⁷ El año 2008, en Gran Bretaña la evolución del Research Assessment Exercise (RAE) al Research Excellence Framework (REF), un sistema de evaluación quinquenal de la investigación de grupos y departamentos, comportó pasar de un sistema basado en la evaluación por pares a un sistema principalmente métrico y cuantitativo (aunque con excepciones para las artes, humanidades, ciencias sociales, matemáticas y estadística, que mantendrán mecanismos de evaluación por pares). Simultáneamente, también en 2008, se produjo en Australia un cambio en el sentido opuesto: el Research Quality Framework (RQF), rebautizado como Excellence in Research for Australia (ERA), pasó de basarse solo en criterios métricos a basarse en evaluación por pares (a los que se suministra documentación informativa con resultados métricos diferenciados por disciplinas).

mayoritariamente bibliométrica, lo cual explica la correlación existente entre los resultados de una evaluación y otra⁸.

Lo que me interesa destacar es que, aunque con matices distintos según los ámbitos científicos, finalmente el mensaje es en esencia el mismo en todos los casos y para todos los ámbitos. Nos dediquemos a las matemáticas o a la sociología, nos dediquemos a la investigación básica o a las aplicaciones, la actividad científica se mide mayoritariamente por las publicaciones en revistas indexadas. ¿Por qué? O quizás, yendo a la raíz del problema: ¿para qué? ¿Es eso lo más importante? ¿Para quién? El sistema de evaluación de la investigación debería siempre responder a esta pregunta: ¿para qué y para quién queremos la investigación?

También en lugar destacado se sitúa la incapacidad del sistema para valorar lo atípico, esto es, para reconocer a la persona o al grupo que tienen valor para la ciencia desde una posición original y atípica en el sistema. No se fomenta realmente el conocimiento, sino los progresos marginales, que se publican con más facilidad que las ideas renovadoras. Del mismo modo, resulta más difícil publicar trabajos interdisciplinarios que trabajos cuya temática se mantenga dentro del flujo central de una especialidad. Asimismo, todos los sistemas de evaluación procuran fomentar y, por consiguiente, valorar la internacionalización de la investigación. Sin embargo, un cierto papanatismo español considera válida la ecuación internacional = extranjero, y a veces puede llegar a parecer que revistas internacionales son solo las que se editan fuera, y en otro idioma. Del mismo modo, y a pesar de todas las loas a la cooperación entre grupos de investigación, países y áreas de conocimiento, en las evaluaciones se sigue considerando que los artículos publicados por un autor en solitario tienen más valor que los publicados en colaboración de investigadores y grupos de investigación, respondiendo a una concepción excesivamente individual y competitiva de la investigación, como si su evaluación fuera un concurso para ver quién es más listo, en vez de una herramienta para hacer avanzar la ciencia.

Tampoco se pueden dejar de lado, aunque ya a otro nivel, las trampas que, con los años, el sistema ha llegado a desarrollar: autores que se citan mutuamente, revisores que, escudados en el anonimato, exigen ser citados en los artículos que revisan, revistas supuestamente científicas que exigen pagos a los autores por publicar sus artículos, poca robustez de los índices de impacto (muy particularmente, su cómputo cortoplacista y la poca correlación entre el índice de impacto de una revista y el número de citas de un artículo concreto publicado en esta).

Finalmente, aunque también de gran importancia, está el progresivo desplazamiento del sujeto evaluador, que está pasando de ser comunidad científica a estar formado por estamentos especializados y burocratizados que, de seguir así, podrían llegar a dirigir a la comunidad científica desde posiciones administrativo-políticas, con la consiguiente pérdida de control democrático sobre las decisiones políticas y de independencia científica en las decisiones técnicas.

El fenómeno de la evaluación de la actividad científica, con las virtudes y los defectos aquí descritos, aunque en menor o mayor medida, es generalizado y universal. A mi entender,

⁸ Véase, por ejemplo, J. Taylor, "The Assessment of Research Quality in UK Universities: Peer Review or Metrics?", *British Journal of Management*, 22: 202-217, 2011.

ha prestado un gran servicio, pero está llegando a su agotamiento si no ha llegado ya. Como en todo aquello que ha funcionado razonablemente bien, pero necesita una reforma, habrá que introducir cambios en el sistema, y hacerlo con cuidado para que mejoren los aspectos que no funcionan adecuadamente sin que se estropeen los que funcionan bien.

ALGUNAS PROPUESTAS

No es este el lugar para desarrollar de forma pormenorizada todos los detalles de una propuesta alternativa o, simplemente, mejorada, del sistema actual de evaluación de la ciencia. Entre otras cosas porque no está claro que exista un consenso suficientemente extendido sobre su necesidad y, menos aún, sobre la orientación que debería tomar. Por ello he optado por presentar este apartado como una lista de ítems de tal modo que cada uno corresponde a una idea, una línea de trabajo, una orientación de mejora o de cambio que, a mi entender, debería considerarse.

1. Reducir la proliferación de actos y agentes evaluadores y dar más alcance a unas pocas evaluaciones. No es sensato que el conjunto del sistema de ciencia y tecnología (evaluados y evaluadores) dedique tantas energías a presentar/evaluar, una tras otra, un sinnúmero de solicitudes específicas. Sería más eficiente una evaluación menos troceada y dispersa, a consecuencia de la cual se diera más capacidad de actuación, más recorrido, a las personas y a los grupos de investigación evaluados positivamente.
2. Concentrar la evaluación en pocos ítems de los currículos, elegidos y argumentados libremente por el personal investigador que es evaluado — individualmente o en grupo —, con el objetivo de enfatizar la calidad y la relevancia frente a la tendencia actual hacia la cantidad.
3. No concentrar tan solo en artículos de revista las aportaciones presentadas por el personal investigador para evaluación de personas y proyectos, sino abrirlas más a libros, capítulos de libros, presentaciones en congresos, documentos de trabajo y publicaciones en Internet, especialmente en determinados ámbitos en los que estas herramientas tienen relevancia.
4. Abandonar el (supuesto) ideal de comparabilidad total de las evaluaciones y aceptar las diversidades entre ámbitos, tanto las de carácter estructural e intrínseco, como las de tipo coyuntural, bajo el principio de que la evaluación debe corresponder a los objetivos de política científica, que no tienen porqué ser los mismos ni a lo largo del tiempo ni para todos los ámbitos.
5. Escuchar a las personas que son evaluadas y debatir con ellas, y no solo leer la documentación que presentan. Hacerlo aunque solo sea periódicamente, o en momentos especiales de sus trayectorias, no limitados a los concursos de acceso a plazas. A mi entender, esta es una de las pocas maneras, si no la única, de evitar la mayor parte de los no pocos falsos positivos y más de un falso negativo que hoy se producen al evaluar sobre el papel.
6. Responsabilizar de esta evaluación a paneles de científicos expertos en cada materia que analicen y debatan y, desde luego, no se limiten a sumar puntuaciones ni se

- concentren en formalismos. En ese debate, las herramientas bibliométricas serán sin duda útiles, pero como un elemento más sobre el cual construir las valoraciones.
7. Soy consciente de las dificultades que la apuesta anterior comporta, en cuanto a la objetividad de la evaluación (si es que tal cosa puede existir) y desde el punto de vista de las garantías jurídicas, especialmente relevantes en el caso de evaluaciones relacionadas con aspectos laborales (selección, estabilización, promoción...). En este sentido, considero deplorables algunos de los sistemas actuales de selección de paneles de evaluadores por sorteo o incluso por ofrecimiento voluntario. ¿Qué garantías de imparcialidad puede ofrecer un sistema de ofrecimiento voluntario? ¿Cuántas restricciones hay que ponerle a un sistema de sorteo para que ofrezca garantías suficientes de competencia?
 8. Superar un sistema en el que las garantías vienen dadas por el sistema de indicadores y, si no sustituirlo, al menos combinarlo con un sistema en que las garantías radiquen en la forma de elección de las personas que llevan a cabo la evaluación, que deben pertenecer a la comunidad científica y contar con conocimientos suficientemente reconocidos. Además, deben *ejercer* como evaluadores, y no profesionalizarse como tales, esto es, deben estar siempre en contacto directo con la ciencia y la investigación en su especialidad.
 9. En este sentido garantista, añadir mucha más transparencia. La comunidad científica ha de poder constatar quién está evaluando y cómo lo está haciendo, pero no solo: cualquier ciudadano ha de poder consultar información veraz sobre los evaluadores y los evaluados. En la era de Internet, esto es cada día más fácil y nada debería impedir superar la actual carencia de información pública sobre la actividad académica (no solo investigadora, sino también docente) del profesorado y el personal investigador de universidades, CSIC y OPIs, y de sus grupos de investigación, que hoy se escuda tras una errónea interpretación del derecho a la privacidad y a la protección de datos que en ningún caso son de carácter personal.
 10. En este mismo sentido garantista, y en el supuesto de que en el futuro sigamos teniendo agencias, u otras entidades específicas encargadas de la evaluación, garantizar que estas ejerzan su competencia técnica de forma independiente. Su competencia *estrictamente* técnica, para que quede más claro. Y que lo puedan hacer de forma independiente de las personas evaluadas y también de los poderes públicos. Para ello es imprescindible mucho más control democrático sobre su funcionamiento. Los *aparatos* de las agencias no deben suplantar en ningún caso ni a los poderes políticos, que son los únicos legitimados para orientar las políticas científicas, ni a la comunidad científica, que es la única con los conocimientos necesarios para llevar a cabo las evaluaciones.

Los defectos del sistema de evaluación de la investigación no son exclusivos de nuestro país. Aunque en este —como en todos— tenga sus peculiaridades, la mayor parte de sus características están extendidas universalmente. Por otra parte, dudo de que exista o pueda existir un sistema de evaluación ideal. Cada sistema corresponde a un momento y a unos objetivos. En este momento, los objetivos que deberíamos perseguir no acaban de encajar con algunos de los efectos que el sistema de evaluación promueve. Algunas de las ideas

presentadas más arriba o bien están ya implantadas en otros lugares (por ejemplo, en la mayoría de países, los currículos tienen formato libre) o han empezado ya a ponerse en práctica (por ejemplo, el sistema de evaluación de proyectos de investigación para su financiación en Alemania ha sido modificado recientemente y ha pasado de usar informes enumerativos exhaustivos a requerir que se destaquen tan solo los resultados más relevantes). No estamos, pues, solos en este viaje.

Vera Sacristán Adinolfi (Barcelona, 1958) es profesora titular de Matemática Aplicada de la Universitat Politècnica de Catalunya (UPC), y su especialidad es la geometría computacional. Ha sido vicedirectora de personal de la UPC (2002-2006), es autora de diversos artículos sobre política universitaria, coordinadora de un libro sobre el mismo tema, y miembro del Observatorio del Sistema Universitario. Más información, <http://www-ma2.upc.es/vera/politica-universitaria/pub/>