
Evaluación de la ciencia: ante un escenario de desafíos e incertidumbres

Research evaluation: facing a scenario of challenges and uncertainties

María Bordons

(maria.bordons@cchs.csic.es)

Daniela De Filippo

(daniela.defilippo@cchs.csic.es)

*Grupo de Análisis Cuantitativo en Ciencia y Tecnología (ACUTE)
Instituto de Filosofía (IFS), CSIC*

Recibido: 18-05-2023; Revisado: 07-06-2023; Publicado: 29-06-2023

Resumen: La evaluación de la ciencia es necesaria para favorecer la investigación de calidad y optimizar las inversiones en investigación. La revisión por pares o *peer review* es el principal sistema de evaluación de la calidad de la ciencia, llevado a cabo por expertos en las materias a evaluar. La participación de otros actores, sociales o políticos, adquiere especial interés para valorar la relevancia social de la investigación. Los indicadores bibliométricos, aplicados con conocimiento de sus limitaciones, pueden constituir un apoyo importante para los expertos porque aportan objetividad y transparencia. Asistimos hoy a un movimiento de reforma del sistema de evaluación tradicional, propiciado por el uso abusivo y reduccionista de los indicadores cuantitativos en el pasado, y por la necesidad de instaurar aproximaciones más holísticas, aplicar las métricas de forma responsable y apoyar el desarrollo de una ciencia más abierta y colaborativa.

Palabras clave: evaluación de la ciencia; *peer review*; indicadores bibliométricos; métricas responsables

Abstract: The evaluation of science is necessary to promote high quality research and optimize research investments. Peer review is the main system for evaluating the quality of science, carried out by experts in the subjects to be assessed. The participation of other actors, whether social or political, is of particular interest for assessing the social relevance of research. Bibliometric indicators, applied with knowledge of their limitations, can be an important support for experts because they provide objectivity and transparency. Today we are witnessing a movement to reform the traditional evaluation system, fostered by the abusive and reductionist use of quantitative indicators in the past, and by the need to establish more holistic approaches, to apply metrics responsibly and to support the development of a more open and collaborative science.

Keywords: research evaluation; peer review; bibliometric indicators; responsible metrics

Como citar este artículo/Citation: Bordons, M. y De Filippo, D. (2023). Evaluación de la ciencia: ante un escenario de desafíos e incertidumbres. *Enredadera: revista de la Red de Bibliotecas y Archivos del CSIC*, (39), 13-20. <https://doi.org/10.20350/digitalCSIC/15375>



1. ¿POR QUÉ EVALUAR LA ACTIVIDAD CIENTÍFICA?

La ciencia, a través de la generación de nuevo conocimiento, contribuye al progreso social y económico de los países y a mejorar la salud y la calidad de vida de las personas. Sin embargo, su desarrollo requiere importantes inversiones económicas y el mantenimiento de infraestructuras científicas, cada vez más sofisticadas y costosas. Dado que los recursos destinados a la investigación son limitados, existe una creciente necesidad por instaurar mecanismos de evaluación de la ciencia, orientados a favorecer la investigación de alta calidad y relevancia social, e intentando optimizar las inversiones en investigación.

Las prácticas de evaluación operan en distintos ámbitos del quehacer científico. Desempeñan un papel fundamental en el sistema de comunicación científica, en el que los artículos y otras contribuciones de investigación son evaluadas de forma previa a su publicación. Pero sus aplicaciones se extienden también al campo de la política científica. Se pueden evaluar investigadores individuales, grupos, centros, proyectos o incluso programas de investigación. Así, la evaluación juega un papel esencial en los procesos de contratación, adjudicación de plazas o promoción de personal investigador, donde se valoran sus méritos curriculares y científicos; en la concesión de premios y distinciones de academias o sociedades científicas; o en la asignación de recursos económicos para la investigación, por ejemplo, a través de la evaluación de propuestas de proyectos de investigación.

La revisión por pares o *peer review*, constituye el principal sistema de evaluación de la calidad de la investigación, que se caracteriza por ser “interno” a la ciencia, en cuanto que lo llevan a cabo los propios científicos. Sin embargo, con frecuencia interesa valorar no sólo la calidad de la investigación sino también otros aspectos como su relevancia social, lo que puede llevar a la participación de otros actores, por ejemplo, de la esfera social o política. Asimismo, la inclusión de otras metodologías, como los indicadores bibliométricos, también conlleva cierta externalización del proceso de evaluación, que deja de descansar solamente sobre la valoración cualitativa de los pares.

2. EL SISTEMA DE REVISIÓN POR PARES O JUICIO DE EXPERTOS (“PEER REVIEW”)

La evaluación por expertos o pares es el principal mecanismo de control de calidad de la ciencia en todos los campos del conocimiento. Su amplio uso se basa en la premisa de que los expertos de un campo científico son los más cualificados para valorar la calidad de la investigación que se realiza en su área de especialización y juzgar aspectos como la originalidad, validez y relevancia de dicha investigación.

La práctica de la evaluación por pares se remonta a la segunda mitad del siglo XVIII, cuando la Royal Society of London asumió la responsabilidad oficial de las *Philosophical Transactions* (1752) y se instauró la existencia de comités (“*committee of papers*”) para evaluar los artículos y garantizar la calidad de

aquellos finalmente publicados (Bornman, 2011). El sistema de revisión por expertos se fue incorporando poco a poco a la mayor parte de las revistas académicas a partir de la segunda guerra mundial y es actualmente una parte esencial del proceso de publicación científica. Las revistas científicas, a través de sus editores y revisores (también denominados “*gate-keepers*” de la ciencia) (Merton, 1979), son responsables de la validación del nuevo conocimiento, que se hace público y se difunde a la comunidad científica.

El sistema de *peer-review* es la piedra angular de la ciencia, ya que vela por mantener sus estándares de calidad y la integridad de la investigación. En un entorno científico cada vez más competitivo, no sólo es necesario salvaguardar la calidad de la investigación, sino también detectar posibles malas prácticas científicas. Así, en el contexto de la publicación científica, editores y revisores necesitan estar cada vez más alerta antes posibles situaciones de fraude (p.e. fabricación de datos), plagio, duplicidad de artículos, fragmentación de resultados, problemas en la asignación de autoría o conflictos de interés no declarados.

Aunque existe un amplio consenso sobre el papel crucial de la revisión por pares como mecanismo de control de calidad de la ciencia, el sistema no está exento de limitaciones, ampliamente recogidas en la literatura (ver p.e., Lee *et al.*, 2013). Entre ellas, se puede mencionar la parcialidad de los expertos, cuyos juicios pueden estar sesgados e influidos por aspectos ajenos al mérito científico de los artículos o de las propuestas de proyectos (p.e. sesgos personales o sociales), la incapacidad de los expertos para detectar algunos errores o fraudes, su dificultad para evaluar investigación interdisciplinar y la actitud conservadora de ciertos evaluadores que, en algunos casos, puede limitar la publicación o financiación de ideas nuevas, innovadoras y poco convencionales (Brezis y Birukou, 2020). A pesar de estas limitaciones, hoy generalmente se acepta que el sistema de revisión por expertos es necesario y beneficioso, aunque también imperfecto y mejorable, sobre todo en lo que respecta a su transparencia y susceptibilidad a diversos tipos de sesgos.

Existen en la literatura numerosos estudios que ponen de manifiesto la preocupación por aumentar la calidad de los procesos de evaluación llevados a cabo por expertos, incrementar su credibilidad, reducir sus sesgos, incorporar innovaciones en sus procedimientos y mantener un equilibrio entre coste y eficiencia de las prácticas de evaluación, entre otros aspectos. Una buena muestra del interés que despiertan estos temas es la celebración periódica de un congreso monográfico internacional sobre la revisión por pares a nivel editorial (ver p.e. el *International Congress on Peer Review and Scientific Publication*, en 2022)¹, o los informes de la organización *Science Europe* sobre las prácticas de evaluación para la financiación de programas y promoción científica².

¹ <https://peerreviewcongress.org/>

² <https://doi.org/10.5281/zenodo.4915999>

3. VENTAJAS Y LIMITACIONES DE LOS INDICADORES BIBLIOMÉTRICOS

La bibliometría y, en particular, la bibliometría evaluativa, que es la aplicación de las técnicas bibliométricas a la evaluación de la actividad científica (Narin, 1976), ha experimentado un gran auge desde los años 60 del siglo pasado hasta la actualidad. A ello ha contribuido la creciente demanda de indicadores cuantitativos por parte de los gestores de la investigación y, por otro lado, la proliferación de fuentes de información cada vez más diversas, así como los avances en las tecnologías de la información y en el tratamiento de grandes volúmenes de datos.

Entre las ventajas de los indicadores bibliométricos hay que señalar que, por su carácter cuantitativo, aportan objetividad y transparencia a las evaluaciones. Por otro lado, pueden aplicarse al estudio de unidades de muy distinto tamaño: desde investigadores o grupos de investigación (nivel micro) hasta instituciones o áreas (nivel meso) y países (nivel macro), siendo este último nivel difícilmente abordable a través del sistema de expertos. Sin embargo, los indicadores bibliométricos no pueden sustituir a las valoraciones cualitativas realizadas por los expertos, sino que pueden complementar y apoyar sus tomas de decisión en lo que se ha denominado “*informed peer review*” (Moed, 2007).

Las bases de datos de publicaciones científicas *Science Citation Index*, *Social Sciences Citation Index* y *Arts & Humanities Citation Index* -elaboradas por el *Institute for Scientific Information* (ISI) de Filadelfia (EEUU), creado en 1963 por Eugene Garfield- fueron las primeras en recoger las citas recibidas por los artículos y otros indicadores derivados, como el factor de impacto de las revistas. Actualmente estos índices están integrados en la *Web of Science* (WoS), producida por *Clarivate Analytics*. Posteriormente han surgido otros recursos útiles para la investigación bibliométrica como la base de datos *Scopus*, elaborada por Elsevier desde 2004, o *Google Scholar* (GS), buscador de publicaciones científicas en la web, también creado en 2004. GS presenta algunas ventajas como su amplia cobertura y libre acceso, pero su uso en evaluación se ve limitado por la menor calidad de su información, ya que se ha descrito falta de transparencia en su cobertura, falta de normalización y presencia de duplicados, entre otros inconvenientes (Torres-Salinas *et al.*, 2009).

En la actualidad existe una gran variedad de fuentes de información, algunas de las cuales agregan recursos de diversa naturaleza (no sólo publicaciones) (p.e. *Dimensions*, *Lens*, *OpenAlex* o *Crossref*) e incluyen numerosos indicadores, tanto del ámbito de la bibliometría tradicional como indicadores de uso de las publicaciones e indicadores altmétricos. A pesar de todo ello, *WoS* y *Scopus* siguen siendo las fuentes más utilizadas en evaluación científica por su mejor normalización, claros criterios de cobertura y existencia de indicadores y productos derivados (p.e. *JCR*, *In-Cites*, *Scival*...).

Contamos en la actualidad con una gran variedad de indicadores bibliométricos, que permiten analizar aspectos como la actividad científica, el impacto y las prácticas de colaboración de los investigadores, grupos, centros o países, y enmarcar estas características en un contexto institucional, nacional o mundial. Aunque la bibliometría no aporta herramientas para medir un concepto tan

El uso de los indicadores bibliométricos en evaluación requiere considerar algunas premisas básicas que no siempre son bien conocidas fuera de la comunidad bibliométrica como, por ejemplo, las diferencias en los hábitos de publicación y citación según disciplinas o la importancia de contextualizar los datos. De lo anterior deriva que las comparaciones inherentes a los procesos de evaluación no deben realizarse entre investigadores o centros de distintas áreas a no ser que se utilicen indicadores normalizados; o la importancia de contextualizar los datos y realizar comparaciones con alguna unidad de referencia (p.e. el país o el mundo). Especial cuidado hay que tener con el uso de los indicadores en las ciencias sociales y, sobre todo, en las humanidades, porque sus prácticas de producción y difusión del conocimiento difieren mucho de las de las ciencias experimentales y naturales. En este sentido, es fundamental reconocer en este campo científico aspectos como el rol protagónico de los libros y las publicaciones en lenguas nacionales (no siempre adecuadamente recogidos en las bases de datos bibliográficas), la lenta obsolescencia de la literatura y sus bajas tasas de citación. Asimismo, en todas las áreas del conocimiento es recomendable usar múltiples indicadores, para captar las distintas dimensiones de la actividad investigadora (Martin, 1996). Los indicadores bibliométricos aportan interesante información sobre la actividad de los investigadores que da lugar a publicaciones, pero es conveniente su aplicación complementando al juicio de expertos, y junto a otro tipo de indicadores para valorar las diversas facetas de la investigación.

4. HACIA LA CONSOLIDACIÓN DE “MÉTRICAS RESPONSABLES”

A pesar de las consideraciones anteriores, en los últimos años se ha descrito un uso inadecuado y/o abusivo de los indicadores bibliométricos en los procesos de evaluación, bien porque se les otorga un peso excesivo o porque se usan sin conocer sus limitaciones y condiciones adecuadas de aplicación, en especial si no se cuenta con asesoramiento experto. Como ejemplo de prácticas inadecuadas, se puede citar el uso del índice h como criterio de promoción de investigadores o la asignación de “recompensas” (gratificaciones o incentivos económicos) a los investigadores que publican en revistas de alto factor de impacto. Estas prácticas repercuten en el comportamiento de los investigadores y, en última instancia, afectan a la integridad del proceso científico. Así, la implementación de indicadores bibliométricos en las evaluaciones ha trasladado a los investigadores una gran presión por publicar, y ha conducido a prácticas no deseadas como la publicación de resultados fragmentados entre diversas publicaciones (“*least publishable unit*”), la excesiva auto-citación, la autoría fantasma, los círculos de co-citación, o un incremento de retractaciones de artículos que se asocian a errores derivados de la prisa por publicar. Por otro lado, se ha señalado que asignar un excesivo peso a los indicadores bibliométricos en la evaluación puede desincentivar la diversidad e interdisciplinariedad en la investigación y fomentar la investigación más convencional y con resultados a corto plazo frente a aproximaciones más innovadoras, de resultados más inciertos y, por tanto, incierta publicación (Weingart, 2005). Asimismo, promover en exceso la publicación entre los investigadores puede ir en detrimento de su dedicación a otras actividades importantes como la docencia (Abramo y D’Angelo, 2021).

En este contexto ha aparecido el concepto de “métricas responsables” (Wilsdon *et al.*, 2015), que se refiere al uso apropiado de los indicadores cuantitativos en la evaluación de la ciencia. Los indicadores apoyan, pero no sustituyen, al juicio de expertos y los efectos de su aplicación deben vigilarse e instaurar cambios si se detectan repercusiones negativas de su uso. En esta línea, han surgido distintas iniciativas y manifiestos que proporcionan recomendaciones sobre el buen uso de los indicadores en evaluación científica, como son *The Metric Tide* (Wilsdon *et al.*, 2015), la Declaración de San Francisco (DORA)³ o el Manifiesto de Leiden (Hicks *et al.*, 2015). Más recientemente, los Principios de Hong Kong (Moher *et al.*, 2020) enfatizan la importancia de considerar también la integridad científica en los procesos de evaluación. Todas estas iniciativas han tenido un gran efecto a nivel mundial, y en el contexto europeo han propulsado la formación de la *Coalition on Advancing Research Assessment* (CoARA)⁴, formada por organizaciones científicas comprometidas con el desarrollo de nuevas prácticas de evaluación de la ciencia en el marco europeo. Esta iniciativa propone buscar aproximaciones más holísticas, reconocer las distintas dimensiones de la actividad investigadora, que va más allá de la publicación, y recuperar el papel de los métodos cualitativos, en los que los expertos juegan un papel central, con un uso responsable de los indicadores cuantitativos.

Sin duda, la evaluación de la ciencia es fundamental para asegurar el adecuado desarrollo de la investigación y velar por su calidad y buen funcionamiento. Sin embargo, asistimos a un momento de amplia discusión y crítica a los modelos de evaluación tradicionales, en concreto a la aplicación reduccionista y al uso abusivo de los indicadores cuantitativos en los procesos de evaluación. Este reclamo por un cambio en el modelo de evaluación está en consonancia con una época en que las prácticas de la ciencia abierta van ganando espacio y, con ellas, la necesidad de abogar por una ciencia más colaborativa, accesible, responsable e inclusiva. Es evidente que el tipo de evaluación que se implemente repercutirá directamente en las prácticas científicas, por lo que, adoptar el modelo de evaluación más adecuado requerirá responder primero a la pregunta ¿qué tipo de ciencia queremos? Este será, probablemente, uno de los grandes desafíos a los que nos enfrentaremos en los próximos años.

5. REFERENCIAS

- Abramo, G., y D'Angelo, C.A. (2021). The different responses of universities to introduction of performance-based research funding. *Research Evaluation*, 30(4), 514-528. <https://doi.org/10.1093/reseval/rvab022>
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45(1), 197-245. <https://doi.org/10.1002/aris.2011.1440450112>
- Brezis, E.S. y Birukou, A. (2020). Arbitrariness in the peer review process. *Scientometrics*, 123, 393–411. <https://doi.org/10.1007/s11192-020-03348-1>
- Garfield, E. (1955). Citation Indexes for Science. *Science*, 122(3159), 108–111. <https://doi.org/10.1126/science.122.3159.108>
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S. y Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429–431. <https://doi.org/10.1038/520429a>

³ <https://sfdora.org>

⁴ <https://coara.eu>

- Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569-6572. <https://doi.org/10.1073/pnas.0507655102>
- Lee, C.J., Sugimoto, C.R., Zhang, G. y Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2-17. <https://doi.org/10.1002/asi.22784>
- Martin, B.R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36(3), 343-362. <https://doi.org/10.1007/BF02129599>
- Merton, R.K. (1979). *The sociology of science: an episodic memoir*. Southern Illinois University Press.
- Moed, H.F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy*, 34(8), 575-583. <https://doi.org/10.3152/030234207X255179>
- Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M.H., Barbour, V., Coriat, A.-M., Foeger, N. y Dirnagl, U. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology*, 18(7), e3000737. <https://doi.org/10.1371/journal.pbio.3000737>
- Narin, F. (1976). Evaluative bibliometrics. The use of publications and citation analysis in the evaluation of scientific activity. *Computer Horizons*
- Priem, J. y Hemminger, M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social web. *First Monday*, 15(7). <https://doi.org/10.5210/fm.v15i7.2874>
- Torres-Salinas, D., Ruiz-Pérez, R. y Delgado-López-Cózar, E. (2009). Google Scholar como herramienta para la evaluación científica. *El profesional de la información*, 18(5), 501-510. <https://doi.org/10.3145/epi.2009.sep.03>
- Waltman, L. (2016). A review of the literature of citation impact indicators. *Journal of Informetrics*, 10(2), 365-391. <http://dx.doi.org/10.1016/j.joi.2016.02.007>
- Weingart, P. (2005). Impact of bibliometric upon the science system: inadvertent consequences? *Scientometrics*, 62(1), 117-131. <https://doi.org/10.1007/s11192-005-0007-7>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. y Johnson, B. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. HEFCE. <http://doi.org/10.13140/RG.2.1.4929.1363>