

La evaluación de la ciencia: de lo cuantitativo a lo cualitativo

José-Manuel Estrada Lorenzo^a

^aBiblioteca del Hospital Universitario 12 de octubre. Instituto de Investigación imas12 Madrid (España). Miembro del Comité Editorial de la *Revista Clínica de Medicina de Familia*. ORCID: 0000-0002-1767-586X

La producción científica ha de ser evaluada, esta es una certeza indiscutible. Las razones son muchas y variadas, pero la cuestión no es tanto el porqué, sino el cómo. Aunque en sus orígenes las contribuciones a la ciencia estuvieron sujetas a los juicios de pares (*peer review*) o expertos como un cierto control de calidad¹, en las últimas décadas la evaluación científica ha estado anclada o sometida² a un «factorimpactismo» excesivo, en el que ha primado la asignación de unos indicadores cuantitativos —propios de las revistas y no de los artículos— en detrimento de la evaluación por expertos. Ya no es solo que prácticamente se contemple como único criterio evaluativo el peso de las citas, con los defectos que este indicador arrastra (diferencia de citación entre disciplinas, entre ciencias básicas y experimentales, entre revisiones y originales, entre publicaciones en inglés o en castellano, entre el norte y el sur, etc.)^{1,3-5}, sino que las instituciones académicas o sanitarias que han de evaluar a sus profesionales tienen que suscribir «obligatoriamente» estos recursos (léase, por ejemplo, *Web of Science*⁶ o *Scopus*) si quieren tener acceso a esos indicadores bibliométricos (factores de impacto, cuartiles, número de citas, etc.), con el consiguiente desembolso económico anual.

Durante mucho tiempo se ha considerado que el factor de impacto era el mejor indicador posible⁷ (idea beneficiada de la correlación entre revistas más citadas y revistas de mayor prestigio/indexadas en bases de datos internacionales) al no existir otros indicadores que lo sustituyeran y porque ofrecía algunas ventajas (por ejemplo, facilidad de comprensión y rapidez de cálculo)⁸. Aunque se han elaborado otras propuestas complementarias⁹, estas no han abandonado el específico ámbito cuantitativo de las citas, y este status de predominio del valor de las citas, durante décadas exclusivo del *Journal Citation Reports (JCR)*, ha persistido incluso ante la aparición de nuevos recursos de acceso libre como *Google Scholar*¹⁰ o *Scimago Journal & Country Rank (SJR)*, pero ha empezado a resentirse tras la publicación de la *Declaration on Research Assessment (DORA)*¹¹ o el *Manifiesto de Leiden*¹², la creación de la *Coalition for Advancing Research Assessment (coARA)*¹³ y el avance del movimiento *Open Science*.

Un grupo de editores de publicaciones académicas se reunieron en San Francisco en diciembre de 2012, en el marco de la reunión anual de la *American Society for Cell Biology*, y redactaron esta declaración (*DORA*), cuyo principal interés estriba en explorar nuevos caminos a la evaluación de la ciencia sin sometimiento al cuantitativo factor de impacto. Sus autores señalaban la «necesidad

apremiante de mejorar la forma en que las agencias de financiación, las instituciones académicas y otros grupos evalúan la investigación científica»¹¹, y concluían que la producción científica debía medirse con precisión y evaluarse con prudencia. Para ello propusieron tres principales recomendaciones:

1. La necesidad de eliminar el uso de métricas basadas en revistas, tales como el factor de impacto, en consideraciones de financiamiento, nombramiento y promoción.
2. La necesidad de evaluar la investigación por sus propios méritos, en lugar de basarse en la revista en la que se publica la investigación.
3. La necesidad de capitalizar las oportunidades que ofrece la publicación en línea (como flexibilizar los límites innecesarios en el número de palabras, figuras y referencias en los artículos, y explorar nuevos indicadores de importancia e impacto).

La *DORA*, que ha tardado en implantarse más años de lo esperado (el Instituto de Salud Carlos III [ISCIII], por ejemplo, no ha incluido los criterios *DORA* hasta este año 2024 en sus solicitudes de becas y proyectos de investigación), ha supuesto una pequeña revolución al solicitar el abandono del «factorimpactismo» e invitar a explorar nuevos indicadores implícitos en la calidad del artículo y no en la calidad bibliométrica de la revista donde se ha publicado. Ello ha abierto el camino a otros indicadores relacionados con la divulgación, el impacto social o el contexto de la investigación y ha permitido la aparición de nuevas «métricas narrativas»⁸, que deberían irse normalizando, estandarizando y generalizando entre el colectivo de profesionales que investigan y evalúan, a través de la divulgación y la formación⁵. No se trata, como también avanza el *Manifiesto de Leiden*¹², de abandonar taxativamente los indicadores bibliométricos cuantitativos, sino de que estos apoyen y refuercen la evaluación cualitativa realizada por personas expertas⁵.

Regresando al ISCIII y basándose en los fundamentos *DORA*, con buen criterio esta institución evaluadora ha incluido los siguientes parámetros para evaluar la producción científica declarada en las becas y proyectos de investigación:

4. Relevancia del artículo en el área de conocimiento: esto es, enmarcar el trabajo en su ámbito científico, donde se siguen



El contenido de la Revista Clínica de Medicina de Familia está sujeto a las condiciones de la licencia de Creative Commons Reconocimiento-NoComercial-SinObrasDerivadas 4.0

publicando infinidad de trabajos redundantes con escasa o nula aportación al avance de la ciencia.

5. Posible aplicación de los resultados en innovación o aplicación clínica: esto es, beneficiar la aplicabilidad y utilidad de la ciencia frente a otras investigaciones de corto alcance.
6. Lugar que ocupa el/la jefe/a de grupo en la publicación (se valorará especialmente la autoría principal): esto es, y ya era hora, se otorga más peso a la autoría principal que a la autoría anecdótica (¿cuántos artículos se han publicado en los últimos años cuya autoría se ha atribuido a cientos y cientos de autores?).
7. Papel desempeñado por el/la jefe/a de grupo en el trabajo del manuscrito: esto es, evitar la aparición de personas pseudoautoras que no han contribuido al manuscrito y revalorizar en su justa medida a cada autor y autora del trabajo considerando su contribución real.
8. Coherencia entre las publicaciones seleccionadas y la trayectoria investigadora, en relación con la propuesta: esto es, encuadrar las publicaciones en la línea de trabajo de la persona investigadora, su grupo y su institución.
9. Adecuación del artículo respecto al lugar relativo que ocupa la revista: esto es, contextualizar los indicadores bibliométricos y aportar otros además del consabido factor de impacto.

Para evaluar las 10 publicaciones que los investigadores deben entregar, donde a cada una de ellas se le puede asignar un máximo de 4 puntos, el ISCIII ha establecido que ninguna relevancia equivale a 0 puntos, y alta relevancia, a 4. Pero ¿cómo hacer esta evaluación con criterios objetivos? Es importante que, una vez que se ha hecho este significativo avance de evaluar de forma cualitativa la investigación, se establezcan parámetros objetivos a los que las personas que evalúan puedan asirse y asignar con solvencia esas puntuaciones de 0 a 4, estableciendo unas *checklists* o rubricas donde se consideren todas las facetas que tener en cuenta en estos indicadores evaluables. Si no, la evaluación más positiva puede quedar a cierto arbitrio, y podrían salir más beneficiadas aquellas personas investigadoras que mejor «vendan» los méritos de sus investigaciones a través de una redacción ampulosa y entusiasta aplicando los «nuevos» criterios narrativos.

Al mismo tiempo que en el ámbito de la salud se han difundido los criterios DORA, la Agencia Nacional de la Evaluación de la Calidad y Acreditación (ANECA) ha actualizado los criterios¹⁴ para la evaluación de méritos y competencias para el acceso a los cuerpos docentes universitarios al amparo del Real Decreto 678/2023 de 18 de julio¹⁵, introduciendo indicadores cualitativos que suponen también un cambio significativo en la evaluación académica del profesorado. Entre estos nuevos criterios, la ANECA valorará «principios como el del impulso a la ciencia abierta, el reconocimiento de la multidisciplinariedad e interdisciplinariedad, la ampliación del concepto de transferencia al de intercambio del conocimiento, o la valoración de la capacidad de liderazgo y de transformación e innovación»¹⁴.

En este contexto, y al amparo de estas novedades que han proporcionado las agencias de calificación, se hace necesario evaluar la investigación ya no solo desde la cuantificación de las citas y la supuesta calidad de la revista desde una óptica bibliométrica (como señalan algunos autores, «el hecho de publicar un artículo en una revista de alto impacto no significa en absoluto que dicho artículo sea una contribución científica de alta calidad»¹⁶), sino también de acuerdo a criterios de impacto social^{7,8}. Si de verdad queremos hablar de divulgación de la ciencia, y no solo de crear currículum, deben considerarse como parámetros evaluadores las menciones en redes sociales o en medios de comunicación, así como los criterios de uso y visibilidad de la investigación⁸ (número de descargas o enlaces). Para ello, es importante diversificar el almacenamiento de los trabajos, que ya no deben ser exclusivos de las webs de las editoriales, en repositorios institucionales o especializados, y otorgar mayor protagonismo a otros indicadores, como métricas de colaboración y métricas interdisciplinarias⁵ o como las anteriormente conocidas como «altmétricas», a las que ya se debería despojar de los adjetivos «otras» o «complementarias».

Es fundamental, en un futuro inmediato, obtener indicadores multifactoriales para las nuevas evaluaciones narrativas, asignar más recursos a las agencias de evaluación para que puedan asumir estas nuevas formas de calificación, impulsar la ciencia abierta (donde tanto artículos como datos de investigación estén disponibles para todos los científicos), desligar la evaluación de la ciencia de herramientas suscritas «obligatoriamente» (*Web of Science* o *Scopus*) y, más importante aún, regresar a los fundamentos de la publicación científica, donde se publicaba para divulgar y difundir la ciencia y no para «hacer» currículum.

BIBLIOGRAFÍA

1. Sanz Menéndez L. La evaluación de la ciencia y la investigación. REIS. [Internet]. 2014; 21:137-48. Disponible en: <https://digital.csic.es/handle/10261/184800>
2. Werner R. La tiranía del factor de impacto: al someterse al poder de la bibliometría, los artículos científicos pierden utilidad. *Investig Cienc*. 2015; 462:46.
3. Ares A. Uso, mal uso y abuso del factor de impacto. *Rev Esp Fis*. [Internet]. 2017;31(3):49-50. Disponible en: <https://revistadefisica.es/index.php/ref/article/view/2328>
4. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ*. [Internet]. 1997; 314(7079):498-502. Disponible en: <https://doi.org/10.1136/bmj.314.7079.497>
5. López-Borrull A, Ollé C, Abadal E, Melero R, Boté-Vericad JJ. La reforma de la evaluación de la investigación en España: un camino por emprender. *Hipertext.net*. [Internet]. 2023;27:5-17. Disponible en: <https://raco.cat/index.php/Hipertext/article/view/419906>
6. Gregorio-Chaviano O, López Mesa E, Limaymanta C. Web of Science como herramienta de investigación y apoyo a la actividad científica: luces y sombras de sus colecciones, productos e indicadores. *e-Cienc Inform*. [Internet]. 2022;12(1). Disponible en: <https://doi.org/10.15517/eci.v12i1.46660>
7. De Granda-Orive JJ. Algunas reflexiones y consideraciones sobre el factor de impacto. *Arch Bronconeumol*. [Internet]. 2003;39(9):409-17. Disponible en: <https://www.arch-bronconeumol.org/es-algunas-reflexiones-consideraciones-sobre-el-articulo-S0300289603754179>
8. Torres-Salinas D, Orduña-Malea E, Delgado-Vázquez A, Arroyo-Machado W. Fundamentos de bibliometría narrativa [Internet]. (V. 1. Publicado el 15 de enero de 2024). Disponible en: <https://doi.org/10.5281/zenodo.10512837>.

9. Alexandre-Benavent R, Valderrama-Zurián JC, González-Alcaide G. El factor de impacto de las revistas científicas: limitaciones e indicadores alternativos. *Prof Inf.* [Internet]. 2007;16(1):4-11. Disponible en: <https://doi.org/10.3145/epi.2007.ene.01>
10. Vine R. Google Scholar. *J Med Libr Assoc.* [Internet]. 2006;94(1):97-9. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1324783/pdf/i0025-7338-094-01-0097b.pdf>
11. Declaración de San Francisco de evaluación de la ciencia [web]. Disponible en: <https://sfedora.org/read/read-the-declaration-espanol/>
12. Hicks D, Wouters P, Waltman L, de Rijcke S, Ràfols I. Bibliometrics: The Leiden Manifesto for research metrics. *Nature.* [Internet]. 2015;(520):429-4. Disponible en: <https://www.nature.com/articles/520429a>
13. Coalition for Advancing Research Assessment (coARA) [web]. Disponible en: <https://coara.eu/>
14. Agencia Nacional de la Evaluación de la Calidad y Acreditación. Criterios de evaluación y requisitos mínimos de referencia de los méritos y competencias requeridos para obtener la acreditación. [Internet]. Madrid: ANECA; 2024. Disponible en: https://www.aneca.es/documents/20123/53669/Criterios200324_anexos.pdf/470f9b03-fd7a-85b6-4f3a-1925c0f82c50?t=1711094067735
15. Real Decreto 678/2023, de 18 de julio, por el que se regula la acreditación estatal para el acceso a los cuerpos docentes universitarios y el régimen de los concursos de acceso a plazas de dichos cuerpos. [Internet]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2023-19027>
16. Bahamonde A, Larrañaga P, López de Mántaras R. Contra la perversión del sistema de evaluación de la ciencia. *El País.* [Internet]. 3 de mayo de 2024. Disponible en: <https://elpais.com/ciencia/2024-05-03/contra-la-perversion-del-sistema-de-evaluacion-de-la-ciencia.html>