

LA EVALUACIÓN CIENTÍFICA BASADA EN INDICADORES BIBLIOMÉTRICOS:
CREENCIAS Y CERTEZAS, MITOS Y REALIDADES, LUCES Y SOMBRAS
*SCIENTIFIC EVALUATION BASED ON BIBLIOMETRIC INDICATORS: BELIEFS AND
CERTAINTIES, MYTHS AND REALITIES, LIGHTS AND SHADOWS*

por EMILIO DELGADO LÓPEZ-CÓZAR

FACULTAD DE COMUNICACIÓN Y DOCUMENTACIÓN UNIVERSIDAD DE GRANADA

Evaluación, ciencia, investigación, científicos, bibliometría, indicadores bibliométricos.

Keywords: Evaluation, science, research, scientists, bibliometrics, bibliometric indicators.

Resumen: Los indicadores bibliométricos han adquirido un protagonismo creciente no sólo en la evaluación del rendimiento científico de los investigadores y las instituciones en las que estos trabajan sino en su reconocimiento profesional y social. Tras reflexionar acerca del porqué del uso de indicadores bibliométricos en la evaluación científica se exponen críticamente los principios y premisas sobre los que éstos se asientan. Se plantean preguntas y dudas acerca de la confiabilidad y validez de los indicadores bibliométricos con el fin de que los lectores sustituyan creencias por certezas, mitos por realidades y realicen un consumo inteligente, crítico y responsable de los mismos. Se especula sobre la orientación futura de la evaluación bibliométrica y se finaliza repasando los efectos positivos y negativos, las luces y las sombras del empleo de la bibliometría en la evaluación científica.

Abstract: *Bibliometric indicators have gained increasing prominence not only in evaluating the scientific performance of researchers and the institutions in which they work, but also in shaping their professional and social recognition. After reflecting on the reasons behind the use of bibliometric indicators in scientific evaluation, this work critically examines the principles and assumptions on which they are based. Questions and doubts are raised regarding the reliability and validity of these indicators, with the aim of encouraging readers to replace beliefs with certainties, myths with realities, and to make intelligent, critical, and responsible use of them. The paper also speculates on the future direction of bibliometric evaluation and concludes by reviewing the positive and negative effects—the lights and shadows—of employing bibliometrics in the assessment of scientific activity.*

¿Cómo es posible que investigadores ocupados en medir la naturaleza física y social con métodos y técnicas sofisticadas, insatisfechos casi siempre por su insuficiente precisión y certidumbre, hayan aceptado como dogma de fe que toda su actividad científica sea valorada por un único indicador tan tosco, grosero y falaz como JIF (*Journal Impact Factor*) de una revista? ¿Cómo es posible inferir el valor de un artículo concreto del valor promedio de su contenedor o trasladar dicha valoración a la ejecutoria de un investigador?

El problema es que en evaluación de la ciencia con indicadores bibliométricos lo habitual es primero actuar y después pensar. Así, los científicos consumen ávida, superficial y acríticamente las métricas y los *rankings* contruidos sobre ellas como forofos: importa el resultado no la forma de alcanzarlo. ¿Por qué no comportarse como un científico y examinar la metodología y la fuente empleada para capturar y tratar los datos? Pues bien, propongo en este breve texto que pensemos sobre lo que hacemos, planteando preguntas más que certezas, socavando mitos y prejuicios; en fin, mostrando las luces y las sombras

que proyecta el uso de los indicadores bibliométricos.

Desde el nacimiento de la ciencia moderna en el siglo XVII los científicos han venido usando su juicio (evaluación por pares) para valorar la relevancia y rigor de las ideas y descubrimientos que generaban. Y es que cualquier científico competente podía y puede distinguir los trabajos, los proyectos o los investigadores buenos de los malos y clasificar a unos como mejores que otros. Sólo cuando la ciencia se convirtió en una actividad cultivada por miles científicos en todo el mundo (la llamada *big science*) y los recursos eran insuficientes para sufragar los costes de la investigación surgió la necesidad de usar métodos cuantitativos que valoraran a muchos con el mínimo de medios, esto es, de forma eficiente e incontrovertible.

Contar publicaciones y su impacto a través de las citas bibliográficas recibidas de otros científicos emergió en la década de los ochenta del siglo pasado como la forma más aceptada por los científicos para medir su rendimiento. Al fin y al cabo, la publicación y la citación son ingredientes esenciales de la actividad científica pues apelan a las bases psicológicas

de su comportamiento. El método científico exige contrastación y aquiescencia de los pares; los descubrimientos alcanzarán valor cuando la comunidad científica los reconoce como tales. Mientras que la publicación actúa como registro público certificado de la autoría y la propiedad intelectual la citación actúa como moneda de pago del reconocimiento intelectual. La creación de los índices de citas, del Factor de Impacto y de los rankings de revistas (*Journal Citation Reports*) al permitir la cuantificación de publicaciones y citas hizo factible la evaluación con indicadores bibliométricos.

A fin de evitar el uso irreflexivo de los indicadores bibliométricos resulta imprescindible pensar en cuales son las premisas o presupuestos que asumimos cuando decidimos usarlos con fines evaluativos. Sobre cada uno de ellos formularemos preguntas que no son meramente retóricas:

A. Los resultados de investigación se manifiestan fundamentalmente a través de las publicaciones científicas.

¿Podemos afirmar hoy que todos los descubrimientos e ideas científicas cristalizan en publicaciones? Dada la enorme competitividad industrial existente y los intereses económicos que andan en juego ¿estamos seguros que hoy los avances tecnológicos pasan previamente por la publicación? ¿Qué papel juegan otros géneros como las patentes, las aplicaciones informáticas, los informes o memorias técnicas? ¿Tienen las publicaciones el mismo rol en todos campos científicos? ¿Podría ocultarse o minusvalorarse la actividad científica practicada en especialidades donde la publicación no sea un canal de comunicación preferente?

B. Al contar publicaciones se asume implícitamente que todas ellas portan la misma cantidad de conocimiento científico certificado.

¿Todas las publicaciones de un autor tienen el mismo contenido, relevancia, originalidad, rigor? ¿El contador de publicaciones resume la aportación real de conocimiento de sus publicaciones? Si dos autores A y B han publicado 10 artículos ¿podemos asegurar que esos documentos portan la misma cantidad de ideas, hechos, datos, resultados, procedimientos? ¿podemos afirmar que su productividad es la misma? ¿Podemos comparar sin más la productividad de un autor, institución revista a partir de un simple recuento de publicaciones?

C. Todos los firmantes de un trabajo son responsables de la investigación expuesta y han contribuido de igual manera a la misma.

Hoy, en casi todas las disciplinas, a excepción de algunas especialidades humanísticas, jurídicas y sociales las publicaciones son obras colectivas pues son firmadas por varios autores. Se atribuye el notable incremento de la coautoría a que la ciencia es cada vez más compleja (necesidad de colaboración de especialistas de distintas disciplinas y habilidades técnicas) pero también a la presión por publicar que sienten los investigadores al saber que su progresión en la carrera científica depende de las publicaciones que atesore (Figura 1).

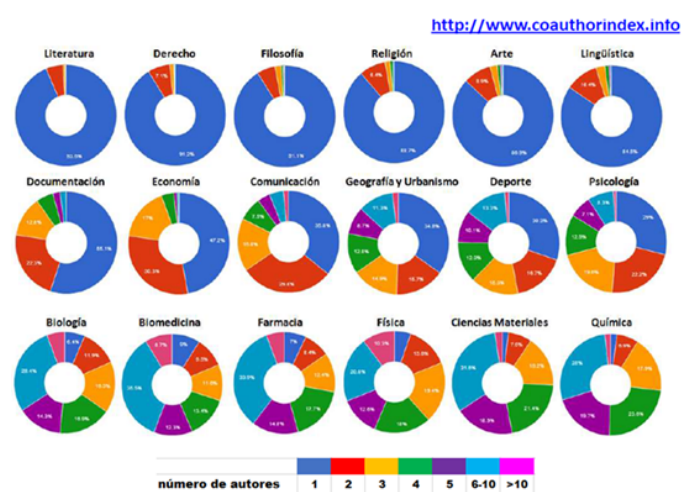


Figura 1. Número de autores por artículo de investigadores españoles en revistas nacionales e internacionales según disciplinas (1999-2015). Fuente <https://web.archive.org/web/20210415084601/http://www.coauthorindex.info/layout.php?subject=Biomedicina>.

Ante esta situación cabe preguntarse ¿Han participado realmente todos los autores que han firmado un trabajo en su confección? ¿Todos pueden ser legítimamente denominados como autores? ¿Todos han contribuido de similar manera a la idea, diseño, recogida, análisis e interpretación de los datos y redacción o revisión crítica del borrador? ¿Todos tienen la misma responsabilidad en cada una de estas tareas? Si asumimos lógicamente la disímil aportación de los autores ¿no sería injusto atribuir a todos los firmantes el mismo nivel de productividad (número de publicaciones) e impacto (número de citas) del trabajo como se viene haciendo? ¿Cómo atribuir el crédito justo a cada autor? ¿dividiendo por el número de autores?, ¿atribuyendo más a aquellos que figuran en posiciones preferentes en la cadena de autoría?

Desde finales del siglo pasado en las revistas biomédicas se empezó a implantar la costumbre de declarar las funciones desempeñadas por cada autor en la publicación (CREDIT <https://credit.niso.org/>) ¿De verdad estas declaraciones son confiables y fidedignas?

¿Si la autoría no es plenamente fiable y válida qué sentido tienen los indicadores bibliométricos que se construyen sobre ella?

D. Las citas son una medida fiable y válida de la calidad de una publicación.

Si un documento tiene 30 citas y otro 15 ¿quiere decir esto que el primero es el doble de bueno que el segundo? La pregunta puede parecer burda pero esta es la lógica implícitamente asumida por la comunidad científica y que se aplica tanto a autores como a instituciones y revistas en los distintos rankings bibliométricos. Pues bien, cuando un científico cita un trabajo anterior lo que formalmente hace es reconocer la utilidad que ha tenido dicho trabajo en la confección del suyo propio y, por ende, es un indicador de influencia. Si se acepta esta premisa los trabajos más citados tienen una influencia mayor en las actividades investigadoras que los que tienen menos citas.

Ahora bien, que un autor, institución, revista o editorial acopie más citas no quiere decir mecánicamente que produzca trabajos de mejor calidad, porque la calidad es una propiedad que se refiere a las cualidades intrínsecas de una publicación. Además, esta asunción es más que discutible puesto que la calidad es una propiedad compleja con múltiples dimensiones. Al menos tres: la calidad cognitiva (significación, relevancia y originalidad del conocimiento contenido); la calidad metodológica (rigor, sofisticación y novedad en el empleo de los métodos y técnicas para recoger, procesar y producir datos); la calidad comunicativa (calidad en la redacción y transmisión de los mensajes). En este sentido, ¿se puede afirmar rotundamente que citación equivale a calidad? ¿Qué noción de calidad es capturada por cada cita expresada como un simple numeral?

De otro lado, el número de citas a secas no nos dice nada del significado de la cita, en tanto en cuanto no nos señala el contexto en que se produce la citación ni las cualidades del documento, ni los propósitos e intencionalidad del autor citante. Entonces cabe preguntarse: ¿Todas las citas bibliográficas valen lo mismo? ¿Todas portan el mismo sentido y significado? ¿Todas se componen con la misma intencionalidad? ¿Se emiten todas las citas con rectos propósitos, es decir, reconocer trabajos anteriores, apoyar o confirmar ideas, identificar antecedentes, citar métodos y técnicas en las que basar la investigación propia, comparar y discutir sobre resultados? ¿Cuántas citas se realizan por motivos espúreos amistad, enemistad, adulación, servilismo, autobombo, manipulación de indicadores bibliométricos?

Por último, y no es una pregunta banal ¿Estamos

seguros que todos los trabajos citados han sido leídos o se copian y pegan sin más?

Sólo después de haber realizado estas cavilaciones se puede comprender el despropósito e irracional uso del JIF (un promedio de las citas recibidas por los artículos publicados en las revistas en un período de dos años) o el índice h (un simple y arbitrario cálculo de publicaciones citas aplicado a los investigadores) como vara para medir el rendimiento de los científicos y, por elevación, el de sus instituciones. Tímidamente salimos hoy día de lo que puede llamarse la prehistoria de la bibliometría tras la denuncia de DORA y las actuaciones de COARA y otros manifestos (Leiden, Hong Kong).

Ahora empiezan a usarse indicadores que miden el impacto directo de los artículos individualmente considerados (CNCI: Category Normalized Citation Impact, FWCI: Field-Weighted Citation Impact, FCR: Field Citation Ratio entre otros) contabilizando el número de citas de un documento y comparándolo con todos los documentos publicados en un mismo campo temático, del mismo tipo (artículo, revisión...) y en la misma fecha de publicación. Aunque ahora sí contamos citas de un trabajo no se olvide que seguimos tienen sólo números desnudos.

Es fácil advertir que el factor decisivo en estos nuevos indicadores, aunque también lo era en los antiguos, es la clasificación temática de los artículos. ¿Están agrupados coherentemente todos los artículos que versan del mismo tema? ¿No se mezclarán peras con manzanas, churras con merinas? Dado que la potencialidad de citación de un documento depende de la densidad de información que, a su vez, es dependiente de la temática de los artículos, resulta absolutamente clave tanto el esquema de clasificación adoptado y su nivel de agregación como la correcta asignación a la categoría temática de los documentos. Aún así siempre surgirá una incómoda pregunta ¿existen dos artículos iguales? Parecidos sí, pero...

En la última década al abrigo de la implantación universal de Internet y las telecomunicaciones, especialmente de los servicios que permiten la interacción de los usuarios entre sí y en diálogo con los documentos que estos producen, se ha generado una multiplicación en el número de métricas que se utilizarán para captar impactos no científicos (educativos, profesionales, económicos, mediáticos, sociales). Los documentos (almacenados en la Web), pueden ser visitados, visualizados, descargados, enlazados, compartidos, citados, reseñados, mencionados, comentados, discutidos, referenciados, etiquetados, valorados, seguidos, difundidos. Muchas de estas acciones se contabilizan y se traducen en métricas que están siendo ya explotadas por dos empresas que pugnan por

hacerse con el mercado: [Altmetric.com](https://altmetric.com) y [PlumAnalytics](https://plumanalytics.com). Mientras discutimos sobre las propiedades y el alcance de estas métricas, no debe perderse de vista la naturaleza de estas acciones, muchas de ellas, tan evanescentes como los medios en las que navegan. ¿Todo lo que se visita, descarga realmente se lee? ¿Qué atención genera un retuit, un me gusta, un etiquetado...? ¿Dejan una huella perenne o son de estado gaseoso?

E. Las bases de datos bibliográficas donde se indexa la literatura científica y, a partir, de las cuales se calculan los indicadores bibliométricos son una representación más o menos cabal de los territorios científicos.

La calidad, fiabilidad y validez de los resultados de los indicadores bibliométricos son proporcionales a la calidad, fiabilidad y cobertura de las bases de datos en que estos se sustentan. ¿Existe alguna base de datos capaz de reunir toda la literatura científica producida por todos los investigadores del mundo? ¿Es similar la cobertura de la *Web of Science* a la de *Scopus*, *Google Scholar*, *Dimensions*, *Semantic Scholar* o *Lens*? ¿Cubren todas por igual todos los campos científicos? ¿Hay diferencias de cobertura según el origen geográfico o la lengua de las publicaciones? Los investigadores deben saber que estas bases de datos no son estadísticamente representativas en el sentido literal del término. Por tanto, deben ser conscientes de sus fortalezas y limitaciones.

¿Hacia dónde vamos? La nueva bibliometría

No decimos nada nuevo cuando afirmamos que la nueva bibliometría se levanta a hombros tanto de las tecnologías de procesamiento de datos como de procesamiento del lenguaje (inteligencia artificial). Todavía sin abandonar del todo los viejos indicadores (JIF, índice h...) se empiezan a generar métricas que intentan capturar algo del sentido de las citas a partir de la frecuencia en que se repiten en un texto (intensidad de citación), de la sección en la que se ubican (Introducción, Metodología, Resultados, Discusión), de la inclinación que transmite (positiva, negativa, neutra), de las funciones retóricas e intenciones discursivas (de apoyo, referencia, técnico-metodológica, discusión...). Con las técnicas de minería de datos, de aprendizaje profundo y automático no está lejos el día en que se midan las entrañas de los documentos captando los significados o en que se cuenten las identidades de autores y (sus huellas dactilares digitales) con todas sus interacciones sobre los textos científicos. La bibliometría será sustituida por una

suerte de Allmetrics que no es más que la medición de todo y de todos, por todos, para todos, a todas horas y en todos los lugares.

El futuro no es menos métricas sino más métricas y más sofisticadas. Vivimos en el reino de la medida y eso no tiene vuelta atrás; la cuantificación de nuestra vida es una realidad. La ciencia no es una excepción. Ahora bien, la vida no es número, ni la vida científica puede ni debe reducirse a un número como se revela en el primer episodio de la temporada tercera de la serie *Black Mirror*.

¡¡¡Medir, Medir, Medir!!! ¿Alguien ha pensado en sus efectos?

El uso de indicadores bibliométricos para evaluar a la ciencia y los científicos no es neutro, tiene consecuencias, unas positivas y otras negativas; genera luces y sombras que dependen de los ojos que los miren o con que se los miren. Los gestores de agencias de financiación o evaluación, de universidades y centros de investigación se sienten atraídos a la par que aliviados por el uso de indicadores bibliométricos en sus tareas de asignación de fondos, selección, promoción y evaluación de las carreras científicas. Las calificaciones bibliométricas no sólo ahorran tiempo y recursos sino que, por su carácter objetivo, son más creíbles. Proporcionan una especie de coartada para evitar las discusiones, controversias y polémicas que las evaluaciones cualitativas (juicio por pares) plantean.

Esta búsqueda de la comodidad del evaluador conduce a la burocratización de la evaluación científica. Las evaluaciones se convierten en meros actos burocráticos que eluden dilucidar y pronunciarse sobre la relevancia, originalidad y novedad, condimentos imprescindibles de la creatividad y el riesgo que son realmente los factores que hacen cambiar la ciencia.

Los científicos, en general, en la medida en que desean valoraciones que sean precisas, sencillas, predecibles y estables estiman los indicadores bibliométricos. Los que se oponen a ellos -normalmente científicos adscritos a las Humanidades, Ciencias Sociales y determinadas especialidades aplicadas- no lo hacen por su naturaleza numérica o por principios ideológicos sino porque no se han construido fuentes, bases de datos y herramientas que representen adecuadamente sus obras y capten sus especificidades.

El verdadero poder del empleo de indicadores bibliométricos es el cambio de comportamiento inducido en las vidas de los investigadores. Se corre el riesgo de que la actividad científica se convierta en un juego estratégico y táctico para rendir más y mejor en los podios bibliométricos. El científico puede

olvidarse del qué o el porqué de su quehacer para centrarse exclusivamente en el cómo. Y cuando ocurre esto es posible que aparezcan fenómenos como la modificación de las líneas de investigación o los tipos de publicaciones hacia temas o géneros más rentables bibliométricamente hablando, como comportamientos deshonestos en la publicación (autoría falsaria) o manipulación de los indicadores bibliométricos (autocitación, carteles de citación, compra de citas...). Sin ignorar que jugar con métricas es muy adictivo porque apelan al narciso que habita en el científico, gratificando su yo, comparándolo competitivamente. Ello genera un juego que proporciona placer y adicción y que puede hacernos pasar de la impactitis (el uso indiscriminado del factor de impacto para valorar todo) a lo que llamo la ludopatía bibliométrica, enfermedad que convierte a los científicos en adictos a la bibliometría pero al mismo tiempo en personas dóciles, sumisas, esclavos de cadenas que anulan su capacidad de análisis y juicio crítico y, lo peor, de su libertad intelectual.

En fin, el problema de fondo en la evaluación con indicadores bibliométricos es el abuso o el mal uso de los mismos para medir propiedades que no pueden ser medidas con dichos indicadores. El uso responsable de la bibliometría pasa por usar las métricas como indicadores, esto es, indicios de acciones y actividades que ayudan a tomar decisiones.

Y para terminar un simple consejo de un bibliómetra: por principio desconfíe de los indicadores bibliométricos y rankings. Active sus mecanismos de defensa intelectual: escepticismo, duda metódica y espíritu crítico. Después, solo después, úselos responsablemente.

Bibliografía

Delgado López-Cózar, E. (2012). Cómo se cocinan los rankings universitarios. *Dendra médica. Revista de humanidades*, 11(1): 43-58.

Delgado-López-Cózar, E., Martín-Martín, A. (2019). El Factor de Impacto de las revistas científicas sigue siendo ese número que devora la ciencia española:¿ hasta cuándo? *Anuario ThinkEPI*, 13.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060): 471-479.

Garfield, E. (1979). Citation Indexing - Its Theory and Application in Science, Technology, and Humanities. New York: Wiley,

Glänzel, W., Moed, H. F., Schmoch, U., Thelwall, M. (eds.). (2019). Springer handbook of science and technology indicators (p. 850). Dordrecht: Springer.

Moed, H.F. (2006). Citation analysis in research evaluation (Vol. 9). Springer Science & Business Media.

Moed, H.F. (2017). Applied evaluative informetrics (p. 312). Berlin: Springer International Publishing.

Narin, F. (1976). Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity. Cherry Hill, NJ: Computer Horizons.

Orduña-Malea, E., Martín-Martín, A., & Delgado López-Cózar, E. (2017). Metrics in academic profiles: a new addictive game for researchers? *Revista española de salud pública*, 90, e20006.