

Informe I
**La evaluación pública de la
investigación científica en el
contexto internacional:
posibilidades y límites**

  **Carmen Pérez-Esparrells**

  **Núria Bautista-Puig**

  **Enrique Orduña-Malea**

Esta publicación está financiada por la ACCUEE con el objetivo de mejorar la calidad del diálogo sobre las universidades en general, y más particularmente en Canarias. La ACCUEE se limitó a proponer el tema de la investigación dejando el contenido del estudio en manos de quienes lo dirigen. Por lo tanto, la ACCUEE no comparte necesariamente las opiniones y conclusiones de este estudio.

Elaborado en Madrid y Valencia, entre marzo y noviembre de 2022.

© Carmen Pérez-Esparrells, Núria Bautista-Puig y Enrique Orduña-Malea.

ISBN: 978-84-09-45275-0

TABLA DE CONTENIDOS

Preámbulo	4
1. Introducción	15
2. La evaluación de la investigación: nuevos horizontes en el contexto internacional	16
3. Revisión de la literatura sobre evaluación de la ciencia	33
3.1. Metodología	33
3.2. Resultados	37
4. La Bibliometría en la evaluación de la investigación	52
4.1. Indicadores bibliométricos a nivel de fuente	54
4.2. Indicadores bibliométricos a nivel de individuo	61
4.3. Indicadores bibliométricos a nivel de publicación	67
4.4. Nuevas métricas y métricas alternativas	71
5. Efectos adversos del uso de indicadores en procesos evaluativos	76
6. Conclusiones	86
6.1. Posibilidades y límites de la evaluación	86
6.2. Observaciones finales	88
6.3. Aspectos a considerar	90
7. Referencias	80

PREÁMBULO

Son muchos los fenómenos que han acontecido desde finales del siglo XX a raíz de la revolución tecnológica facilitada primero por Internet (en tanto que infraestructura), y posteriormente potenciada por la llegada de la Web y de los dispositivos móviles inteligentes. Entre estos fenómenos destacan el ritmo de crecimiento de datos y contenidos, la acelerada transformación de estos contenidos (y su demanda por la Sociedad), la necesidad de compartir conocimiento de forma instantánea y global (especialmente acentuada por la pandemia de COVID-19, al ser sensible a los ojos de toda la Sociedad), la reducción del coste de comunicar a distancia de un modo muy efectivo, la transformación digital de las organizaciones, el reto demográfico y de la inmigración a escala global, la atención singular a los temas de género y diversidad, el aumento del consumo energético en un contexto de cambio climático, o incluso el incremento del fraude y la amenaza de la ciberseguridad, entre otros muchos asuntos de preocupación contemporánea, en los que la actividad académica e investigadora juega un rol fundamental.

Todo ello ha ocasionado multitud de cambios y transformaciones en el sector de la educación superior que afectan a la visión estratégica, propiciando o requiriendo cambios de política y financiación de las instituciones de educación superior (*Higher Education Institutions*, HEIs), y afectando muy especialmente al futuro de la innovación e investigación en las universidades europeas (*European Commission*, 2020).

En el caso de la actividad docente, la inclusión de métodos docentes más creativos y centrados en el aprendizaje, la comunicación asíncrona avanzada con el estudiantado o el uso de nuevas tecnologías en la generación de material docente han modificado drásticamente esta actividad formativa. Como señalan recientes informes de organizaciones internacionales, empresas, consultoras y *think tanks* sobre el futuro de las universidades, tras las sucesivas crisis surgen nuevos retos sobre el valor colectivo e individual que ofrece la educación universitaria, que incluyen no sólo la formación tradicional del estudiante, sino el establecimiento de redes y oportunidades sociales, así como nuevos contenidos educativos (Schleicher, 2020). Con la reinención de los entornos de enseñanza-aprendizaje tras la intensificación de la digitalización, se han ampliado las relaciones docente-discente, por lo que la evaluación de la actividad docente debe adaptarse a los nuevos desafíos globales, valorando, por ejemplo, la

integración del cambio climático y los Objetivos de Desarrollo Sostenible (ODS) en el aprendizaje (UNESCO, 2021).

En cuanto a las actividades de transferencia, éstas se han ampliado en cantidad y variedad, buscando un mayor valor añadido y dejando atrás el restrictivo concepto de asociar transferencia a patentes, protección industrial y otros indicadores de mercado, para ampliarlo a un servicio global a la sociedad, que es la aproximación al concepto de “transferencia” que se tiene en otras culturas académicas. Nuevas metodologías permiten además cuantificar y valorar el impacto de estas nuevas actividades relacionadas con la transferencia social (De la Torre et al. 2018).

En el caso de la actividad investigadora, se puede hablar de un movimiento o reforma en la manera de comunicar y compartir conocimiento (*open science*) con ineludibles consecuencias en la evaluación de la ciencia y de la investigación científica a todos los niveles: sistemas universitarios, instituciones, grupos, centros e institutos de investigación, personal investigador (el Personal Docente e Investigador, en adelante PDI) y proyectos de investigación, que termina implosionando a nivel mundial a partir de la publicación de la llamada *San Francisco Declaration on Research Assessment* (popularmente conocida como la *Declaración de DORA*) (ver sección 2).

Considerando el rol social atribuido a las universidades (especialmente a las públicas), así como el fin último de la investigación científica (ampliar nuestro conocimiento), los derechos y obligaciones del PDI (*Ministerio de Universidades, 2022*) y la creciente demanda de responsabilidad social a la Ciencia (utilizar el nuevo conocimiento adquirido para solucionar problemas tanto de la Sociedad en su conjunto como de sus ciudadanos a nivel individual), se asume que la evaluación de los resultados de investigación así como de los actores que producen estos resultados (PDI) debería ser considerada como una actividad de carácter eminentemente público y de enorme repercusión económica, científica y social. Por tanto, una actividad que no debería quedar exclusivamente en manos de empresas, organizaciones o personas no cualificadas o con distintos conflictos de intereses, más teniendo en cuenta la alta sensibilidad de los métodos y métricas empleadas para su evaluación y de la alta repercusión de los resultados de una evaluación en las personas e instituciones evaluadas.

Por ese motivo, desde las políticas de educación superior hasta las estrategias universitarias, resulta crucial establecer nuevas formas de evaluación y valoración de la investigación y del personal de investigación que se adapte a los nuevos tiempos y se

alinee con los resultados obtenidos por las disciplinas y campos vinculados al estudio de la ciencia en sí misma (entre ellas: *research evaluation, science policy, science communication, scientometrics, meta-research, science studies, research on research*). Una evaluación que considere las nuevas exigencias y demandas de la sociedad y las nuevas tecnologías y métricas, tanto cuantitativas como cualitativas. Además, en todo este proceso se debe garantizar la equidad, diversidad e igualdad de género.

Los nuevos modelos de evaluación deben ser capaces de traspasar la simple valoración de los méritos y llegar a valorar las actitudes y habilidades de las personas candidatas a la hora de generar y transferir el conocimiento, y de reconocer la dedicación del personal de investigación a las actividades de liderazgo, comunicación y transferencia social (*LERU, 2022*). Ello implica no sólo un nuevo concepto de PDI, sino un nuevo concepto de universidad.

En concordancia con la filosofía europea, liderada por la *Comisión Europea* como facilitadora, se pone de manifiesto la relevancia del valor científico, económico y social de toda la actividad del PDI, en el que las distintas actividades de transmisión y difusión de conocimiento (diseminación, explotación, comunicación) juegan un rol cada vez más relevante. En ese sentido, los modelos de evaluación actuales suponen una barrera al cambio natural, pues las personas se ven obligadas a adaptar su modo de trabajo para superar unas evaluaciones que están fundamentadas en la simplificación numérica del mérito académico. Una decisión pragmática que facilita y agiliza el proceso evaluativo, pero que modifica artificialmente el funcionamiento y progreso del sistema.

Contexto del sistema universitario español

Desde la década de los noventa del siglo XX, y bajo el paraguas de la *Ley de Reforma Universitaria* en 1983 (Ley Orgánica 11/1983), con la creación de la Comisión Nacional Evaluadora de la Actividad Investigadora (CNEAI) en 1989 (Real Decreto 1086/1989) encargada de evaluar los nuevos complementos retributivos del profesorado universitario mediante sexenios, el Gobierno de España introdujo un sistema de incentivos individual a los profesores funcionarios para evaluar su actividad investigadora por su medición temporal en periodos de seis años. Con esta iniciativa, puesta en marcha en 1989 y que premia la investigación realizada a nivel individual, se consiguió dar un impulso sin precedentes a la investigación en nuestro país a finales del siglo XX.

En los inicios de la implantación del sexenio de investigación, sólo un número de profesores muy minoritario optaba por presentarse a la evaluación de su actividad científica de manera voluntaria. Aun así, en la década de los noventa este sistema de incentivos provocó un considerable crecimiento de las publicaciones científicas (se partía de una producción científica baja en España) y una mayor internacionalización. Sin embargo, ya a inicios del siglo XXI, el comportamiento del PDI (con contrato indefinido) cambió; de este modo, un mayor número de PDI ha evaluado periódicamente su investigación, buscando no sólo la mayor retribución económica por el complemento retributivo correspondiente concedido, sino también un mayor prestigio y reputación entre los pares, la obtención de proyectos de investigación, la dirección de tesis y la participación en programas de Doctorado, la posibilidad de estar en Tribunales (se recomendaban dos sexenios para las plazas de profesor titular y tres sexenios para las de catedrático), así como una valoración positiva de cara al acceso a plazas. Se puede decir, por tanto, que en sus primeras etapas el sexenio de investigación evitó la designación “a dedo” de los miembros de los Tribunales y, en cierta medida, profesionalizó la carrera académica. Aun cuando esta medida no logró evitar totalmente la endogamia académica, se podría decir que ayudó a moderarla.

A partir de la *Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa* (LOMCE) se produjo además una reducción de la docencia en relación con la investigación acreditada mediante sexenios, lo que originó que un mayor número de profesores solicitara la evaluación individual de su investigación. Aunque esto podría ser interpretado como un primer efecto perverso, es comprensible dado que el puesto de Catedrático/a supone una responsabilidad especialmente elevada en su labor científica y de liderazgo, que puede llegar en algunos casos a ser incompatible con la alta dedicación docente o incluso con la gestión universitaria.

Sin embargo, los beneficios de la evaluación de la actividad investigadora mediante sexenios trajeron otros efectos negativos. Algunos profesores “abandonaron” la progresión en sus carreras investigadoras por problemas de motivación y autoestima. Son tan buenas las recompensas que se obtienen por obtener un sexenio como los castigos por no obtenerlos. Esa presión generó una reacción en una parte del profesorado universitario, con consecuencias tanto deseadas como indeseadas, como se verá más adelante. Además, el sistema actual (apto, no apto) genera el llamado “resentimiento de frontera”, por el cual las personas solicitantes con un recorrido investigador reseñable, pero no suficiente, pasa a ser considerada igual que aquellas personas que no hacen nada,

En este sentido, por un lado, se distinguen las actividades oportunistas. Estas actividades provienen de parte del profesorado capaz de investigar, pero que desea usar atajos o acelerar el proceso. Un ejemplo es la *ingeniería de currículum* (modificar el comportamiento natural de la carrera académica para centrarse solo en los aspectos que son evaluables, y dar menos importancia a qué se investiga, y más importancia a minimizar los esfuerzos para lograr estos méritos) o en focalizar la investigación en materias de moda o que se espera que sea más fácil publicar. Por otro lado, surgen actividades manifiestamente fraudulentas. Estas actividades vienen generalmente por parte del profesorado incapaz de investigar, por distintos motivos (personales o profesionales). Un ejemplo de ello es la falsa autoría (firmar sin participar en los trabajos). Esto se suele lograr mediante la presión a doctorandos/as o profesorado joven a cambio de facilitar su promoción o el mutuo acuerdo con otras personas a cambio de otro tipo de favores académicos o económicos, lo que origina todo tipo de deudas académicas.

Por tanto, como suele ocurrir en los sistemas de incentivos al personal que se mantienen en el tiempo, las medidas van perdiendo su eficacia, sus objetivos se distorsionan hasta generar consecuencias imprevistas e indeseadas, dando síntomas de agotamiento. Actualmente ya se empiezan a detectar comportamientos anómalos o no naturales (Oviedo-García, Casillas Bueno; González Rodríguez, 2021; Delgado López-Cózar y Martín-Martín, 2022) que comienzan a pervertir el sentido original del sexenio de investigación. Estamos ante casos concretos de lo que de forma genérica se conoce como ley de Campbell: cuando se emplea un indicador cuantitativo, cuanto más recompensas se asignan según los resultados del indicador, más se corrompe el proceso social evaluado (Campbell, 1976; Sidorkin, 2016).

En este contexto, la creación y evaluación de Transferencia del Conocimiento e Innovación en 2018 (Resolución de 28 de noviembre de 2018), única convocatoria de momento, del denominado sexenio de transferencia supone un paso más en la evaluación individual del personal investigador. Mediante este nuevo incentivo se reconoce de forma explícita (en caso de evaluación positiva) el impacto social realizado por el PDI de las universidades españolas, implicando igualmente una retribución económica a nivel individual. Por tanto, aunque de manera piloto, la recompensa no proviene únicamente de la actividad investigadora sino del reconocimiento a otras actividades. Sin embargo, y como señalan Repiso et al. (2020), el sexenio de

transferencia debe eliminar o reducir importantes defectos, como las incongruencias en la puntuación o algunos sesgos significativos, como el sexo, entre otros.

La filosofía que está detrás de estos dos tipos de sistemas de incentivos regulados a nivel nacional es distribuir los exiguos fondos complementarios a la investigación, la transferencia de conocimiento e innovación a quienes mejor los utilizan. De hecho, para poder optar de manera voluntaria al sexenio de transferencia es necesario contar con, al menos, un sexenio de investigación. Otra cuestión bien diferente será la necesidad de establecer un tercer sistema de incentivos individual que premie la excelencia en la labor docente, el bautizado como sexenio de docencia. Este nuevo complemento vendría para diferenciarse del llamado quinquenio de docencia, que es otro sistema de incentivos individual que premia la docencia desempeñada cada cinco años. No obstante, esta evaluación interna se realiza de manera bastante automática por parte de las propias universidades, y con claros signos de ausencia de evaluación real del rendimiento docente del profesorado (Delgado López-Cózar y Martín-Martín, 2022).

No obstante, a pesar de que la creación de sexenios de docencia y transferencia refleja síntomas de apertura en los procesos evaluativos a nivel individual, el sistema de acreditaciones a los cuerpos de funcionarios (profesores/as titulares y catedráticos/as) sigue siendo muy dependiente de la obtención de los incentivos individuales a la investigación. Aunque la priorización de la rendición de cuentas de los recursos públicos invertidos y el intento de disuasión de la endogamia académica han provocado que otros criterios e indicadores comiencen a tomar cierta relevancia a la hora de evaluar el trabajo académico de cara a los procesos de acreditación a las distintas figuras de profesorado, todavía queda mucho camino por recorrer para lograr un modelo más flexible, justo, inclusivo y adaptado a los nuevos tiempos y metodologías de evaluación individual, que sea capaz de recoger todas las facetas del PDI.

La propuesta del *Anteproyecto de Ley Orgánica del Sistema Universitario* (APLOSU) (Ministerio de Universidades, 2022) de figuras equiparables a profesor titular y catedrático a nivel autonómico abre una ventana de oportunidad para establecer criterios que se adapten a las necesidades del PDI en el *Sistema Universitario Español* (SUE) así como en los diecisiete sistemas universitarios de las diferentes Comunidades Autónomas. La perspectiva del contexto local en aquellas Comunidades Autónomas que cuentan con una Agencia de evaluación regional¹, como es el caso de la *Agencia*

¹ <http://www.aneca.es/Agencias-de-las-Comunidades-Autonomas>

Canaria de Calidad Universitaria y Evaluación Educativa (ACCUEE), siempre debe ser diseñar las herramientas, métodos y criterios que, respetando los criterios a nivel nacional por lealtad institucional, potencien los efectos de cercanía positivos con un análisis de prospectiva de cara a la necesidad de acreditar estas nuevas figuras, no sólo en el contexto académico y científico, sino también en el contexto económico y social de cada territorio.

Por último, pero no por ello menos importante, se observa cómo una gran mayoría de criterios e instrumentos de evaluación de la investigación incluyen indicadores, fuentes o análisis que han sido profundamente cuestionados por la literatura científica, o simplemente que no siguen los principios y recomendaciones proporcionados por expertos en la materia. En este sentido, se aprecia un efecto desmedido del Factor de Impacto o de la cita desde una publicación indexada, una ausencia de indicadores normalizados, la inclusión de indicadores inservibles (medias de Factores de Impacto o de índices H sin normalizar), una falta de comprensión de las diferencias disciplinarias en cuanto a tradiciones científicas reflejadas en la producción e impacto basado en citas), etc.

El interés de este estudio (formado por dos informes complementarios) radica en un doble reto. Por un lado, se precisa de una reevaluación de los procesos de acreditación para adaptarlos a la diversidad de las contribuciones que se esperan del personal investigador universitario tanto en los nuevos tiempos como en los nuevos entornos académicos siguiendo las tendencias internacionales. Por otro lado, es necesario implementar nuevos modelos de evaluación que combinen lo mejor de las diferentes aproximaciones a la evaluación de la ciencia (tradicionales y modernas) para evitar los comportamientos anómalos (o artificiales) que pervierten los procesos de acreditación. Y todo ello debe realizarse considerando los contextos locales, tanto del caso español en general como canario en particular, así como el marco epistemológico de la ciencia.

Se pretende, por tanto, proporcionar ideas y recomendaciones para limar aquellas ambigüedades o lagunas en las valoraciones y para señalar las carencias en los procesos actuales de evaluación en el nuevo contexto de la ciencia abierta y el actual debate de reforma de los sistemas de evaluación, todo ello desde el marco de los resultados que la propia Ciencia está proporcionando. Es decir, de los resultados empíricos de aquellos campos científicos relacionados con el estudio de la Ciencia y de la evaluación de la actividad científica.

Para ello, se describirá cómo las acreditaciones (evaluaciones de docencia e investigación estandarizadas) han cambiado la naturaleza de la actividad académica en sus diferentes intentos de cuantificar de manera objetiva el valor de la investigación y de la docencia. Del mismo modo, se valorará si se debe mantener la organización de la evaluación de las disciplinas sin tener en cuenta los campos “frontera” o si, por el contrario, se deberían introducir mecanismos que incentiven o premien la interdisciplinariedad, multidisciplinariedad y transdisciplinariedad.

Por último, el estudio pretende, a través de una propuesta de modelo de evaluación holístico, flexible, inclusivo y personalizable, evitar los efectos de la “homogeneización” de la actividad investigadora, docente y de transferencia en detrimento de la diversidad.

Objetivos

Los objetivos principales del estudio (compuesto por dos informes) se describen a continuación:

- **REALIZAR una contextualización de las nuevas tendencias internacionales en evaluación de la Ciencia**, realizando una revisión de los Manifiestos y Declaraciones más relevantes a lo largo del tiempo.
- **REALIZAR una revisión de la literatura científica sobre evaluación de la actividad investigadora** que recoja los principales polos de discusión en evaluación de la investigación, con el fin de conocer si las prácticas de evaluación actuales están o no alineadas con los resultados científicos y con las nuevas directrices y recomendaciones más innovadoras.
- **IDENTIFICAR las prácticas evaluativas llevadas a cabo por Agencias de Evaluación españolas**, con el fin de conocer cuáles son los requerimientos y requisitos que están solicitando, las fuentes e indicadores utilizados, así como los criterios generales aplicados por cada una de ellas (nivel nacional y regional).
- **IDENTIFICAR, DESCRIBIR y DISCUTIR los efectos de los procesos de evaluación en el personal de investigación**, que cristalizan en prácticas artificiales orientadas a optimizar el esfuerzo para superar un proceso evaluativo u obtener más prestigio, distinguiendo entre aquellas prácticas artificiales pero que respetan las reglas de juego y la legalidad (ingeniería curricular) y aquellas que no (fraude).
- **PROPONER un nuevo marco de modelo de evaluación del profesorado universitario holístico, flexible, inclusivo y personalizable** adaptado a las nuevas tendencias de cara a su aplicación futura con el fin de corregir las

consecuencias y efectos no deseados e imprevistos del modelo actual en España.

El estudio se compone de dos Informes, que se describen a continuación:

El primer informe se centra en estudiar las prácticas de la evaluación científica² como disciplina científica en sí misma. Para ello, se realiza un cuádruple análisis. En primer lugar, se lleva a cabo un estado de la cuestión sobre la evaluación de la actividad investigadora, especialmente en lo que se refiere a la elaboración de informes técnicos y propuestas por parte de distintas asociaciones, así como de la reacción de la comunidad científica a los cambios en los modelos de evaluación. En segundo lugar, se realiza un análisis de la literatura científica con el fin de detectar los principales núcleos y temas de discusión. En tercer lugar, se identifican los malos usos en evaluación, incluyendo los indicadores y métricas que se usan de forma inadecuada. En cuarto y último lugar, se identifican y describen los efectos negativos más comunes derivados de la práctica de evaluación, potenciados por los modelos actuales de evaluación.

El segundo informe (Orduña-Malea, Bautista-Puig y Perez-Esparrells, 2022) se centra en el diseño de una propuesta de evaluación del PDI en la universidad que trate de integrar los nuevos hallazgos de la literatura científica, así como las directrices de las asociaciones y organizaciones internacionales referentes en la materia, intentando evitar aquellos efectos adversos que se pueden producir como consecuencia de un modelo de evaluación del profesorado sostenido en el tiempo. Para ello se lleva a cabo un doble análisis. En primer lugar, se realiza un análisis comparativo de las prácticas de evaluación de las distintas Agencias de evaluación en España (tanto ANECA como las distintas agencias regionales). En segundo lugar, se propone un novedoso marco de evaluación del profesorado que pretende incorporar todas las buenas prácticas surgidas de los distintos grupos de trabajo que han elaborado recomendaciones para la evaluación, así como las evidencias surgidas de la literatura científica. Este modelo³ se

² Es preciso matizar que, en ocasiones, se mezcla la evaluación de la ciencia con la evaluación de la investigación, como actividad específica de la evaluación del profesorado universitario, sabiendo que se trata de dos asuntos diferentes pero interrelacionados.

³ Una cuestión es cómo se mide la actividad investigadora, por ejemplo, aplicando técnicas e indicadores bibliométricos, ofreciendo una mayor valoración tanto de los inputs (esfuerzo realizado en muchas actividades que no suponen publicaciones con factor de impacto) como de los outputs (sociales y económicos) que han generado la actividad investigadora (transferencia) y, otra diferente, aunque relacionada con el segundo informe, es el peso que tendría que tener la investigación en la evaluación integral del PDI, que realiza otras muchas actividades, (docencia, gestión, opinión experta, etc.).

construye de forma genérica y conceptual a nivel macro, siendo completamente adaptable a las exigencias de cada Agencia y Figura mediante la definición de una serie de ítems dentro de cada categoría de evaluación y un sistema de pesos.

Cada informe finalizará, a modo de conclusión, con una serie de recomendaciones y propuestas de acciones concretas que se podrían implementar en un futuro para diseñar enfoques de evaluación más adecuados a los nuevos tiempos. Se procura, por un lado, ampliar las maneras de llevar a cabo una evaluación de los investigadores desde una perspectiva multidimensional y, por otro, minimizar las malas praxis y adecuarse a los estándares y recomendaciones procedentes de la literatura científica.

De esta forma, se pretende, finalmente, diseñar las líneas generales de una propuesta integradora de máximos que recoja criterios cuantitativos y cualitativos con el fin de desafiar la “desigualdad” arraigada en algunos criterios entre los científicos de las ciencias 'duras' (ciencias naturales y físicas) y 'blandas' (ciencias sociales y humanidades) y que tenga en consideración desde los aspectos más globales (*open Science*) a los aspectos más locales, propios del Sistema Universitario Canario (SUC) (cuestiones de género, inmigración, exiliados, etc.) necesarios para servir al entorno en el cual se realiza la evaluación por parte de la ACCUEE, sin perder de vista la posible extrapolación a otros entornos evaluativos y a otras figuras nuevas que surjan de la futura *Ley Orgánica del Sistema Universitario*.

Proponer una nueva forma de evaluación individual de la docencia, investigación y transferencia no es tarea fácil ni única, porque afecta a personas e implica, en última instancia, un juicio. Por ello, es necesario que, tras estos Informes sobre la evaluación pública de la investigación y una propuesta de un modelo de evaluación del profesorado universitario, la ACCUEE reflexione estratégicamente junto con la *Dirección General de Universidades* y las universidades canarias, y desarrolle para un futuro su nuevo marco de evaluación de la docencia, investigación y transferencia del conocimiento de los jóvenes académicos. Este marco evolucionará en los años venideros con las nuevas figuras que se incorporen al marco de contratación en las universidades públicas españolas y, por ende, de Canarias. Esto requerirá del tiempo, espacio y recursos para experimentar.

Este Informe espera asimismo animar a los equipos de gobierno de las universidades públicas canarias y a los *policymakers* en el gobierno de Canarias que apoyen este proceso y faciliten espacio y recursos para la experimentación (prueba y error) con el fin

de maximizar el rendimiento del profesorado durante el proceso de evaluación individual, ya que las universidades canarias necesitan nutrirse de los mejores docentes e investigadores en el inicio de sus vidas académicas.

1. Introducción

El objeto de este primer informe es realizar un estudio sobre la evaluación de la investigación resaltando sus posibilidades y limitaciones en la actualidad, con el fin último de detectar los principales polos de discusión (tanto científicos como profesionales) y, con ello, ayudar a sentar las bases para la mejora de la evaluación pública de la investigación en general, y del SUC en particular.

Este objetivo está en línea con la transformación cultural de la evaluación científica en la que el movimiento *open science* es una parte fundamental, así como con distintas iniciativas tanto supranacionales (como la impulsada por la *Comisión Europea*) como nacionales, relativas a recompensas y reconocimientos de la actividad investigadora que potencien su repercusión social o su impacto en el entorno, fortaleciendo el valor de lo intangible.

Para ello, este informe se estructura en los siguientes apartados:

- Identificación de los **hitos e iniciativas más importantes acerca de la evaluación de la investigación** a nivel internacional, especialmente en Europa.
- **Revisión de la literatura científica sobre evaluación de la investigación**, para contextualizar y señalar aquellos aspectos conceptuales más importantes o ideas teóricas más relevantes.
- Recopilación de **indicadores bibliométricos utilizados en evaluación de la investigación**, incidiendo en sus ventajas e inconvenientes.
- Identificación y descripción de los principales **efectos adversos en la comunicación de la ciencia** propiciados por los sistemas de evaluación.
- A modo de conclusión, exposición de las **posibilidades y limitaciones de la evaluación de la investigación**, algunas observaciones finales y aspectos a tener en cuenta en la evaluación en futuro.

2. La evaluación de la investigación: nuevos horizontes en el contexto internacional

Los sistemas de evaluación de la investigación tienen como objetivo medir la excelencia académica de la investigación realizada tanto por investigadores como por diferentes organizaciones de investigación, afectando por tanto al reclutamiento, promoción y reconocimiento del personal de investigación.

Los indicadores bibliométricos se popularizaron en los años sesenta como una opción para evaluar el desempeño de la investigación, basados en datos cuantitativos de publicaciones y citas (Garfield, 1963). Aun cuando los índices de citas (*citation indexes*) no se elaboraron con el fin de evaluar personas sino para servir de ayuda a las bibliotecas en sus procesos de selección de fondos, estos productos fueron poco a poco adquiriendo un papel central en la evaluación del rendimiento científico, efecto que se acentuó cuando los distintos productos creados por Eugene Garfield (especialmente el *Science Citation Index*, *Social Science Citation Index* y los *Journal Citation Reports*) pasaron de imprimirse en papel a ofrecerse de forma interactiva en la Web (de ahí, por cierto, el nombre de *Web of Science*), favoreciendo su acceso y consulta a nivel global.

Estas bases de datos comenzaron a ser utilizadas para fines diferentes a los inicialmente previstos, a pesar de las advertencias del propio Garfield (1979) acerca de los efectos perjudiciales que tenía el uso de estos productos en términos evaluativos. Por ejemplo, el uso de indicadores inapropiados con un mayor énfasis en la cantidad de resultados de investigación en vez de en su calidad, potenciando así la cultura del *publish or perish*, término ampliamente utilizado que describe la presión sobre los académicos para publicar artículos científicos como condición para el empleo, promoción o éxito a lo largo de su carrera académica.

El uso de estas bases de datos por parte de las agencias de evaluación como base en sus modelos de evaluación, junto al surgimiento de otras bases de datos similares (*Scopus*), propició la creación de un ecosistema fundamentado en la aplicación de ciertos indicadores bibliométricos a la hora de evaluar a personas. La rapidez (los datos vienen precocinados en la propia base de datos) y la supuesta objetividad de estos datos (más fácilmente comparables a nivel internacional que otras evidencias, como la docencia) potenciaron su uso hasta convertirlos, en algunos casos, en los únicos

parámetros considerados a efectos de evaluación, especialmente en algunas disciplinas y en algunas Agencias y centros e institutos de investigación. La oferta y demanda de indicadores bibliométricos llevó al sistema hacia una *impactolatría* (Camí, 1997), potenciada por la existencia de plataformas online intermediadoras capaces de conectar investigadores y métricas a gran escala (Ma, 2022), propiciando de ese modo todo tipo de “adicciones” entre el personal académico (Orduña-Malea, Martín-Martín y Delgado López-Cózar, 2017).

Sin embargo, durante la última década han surgido múltiples iniciativas (por ejemplo, DORA o *Leiden Manifesto*) que consideran que las métricas tradicionales (y las fuentes que las proporcionan) no son instrumentos apropiados (o suficientes) para cuantificar y evaluar el desempeño de los académicos, poniendo de manifiesto la necesidad de reformular estos sistemas. Estas iniciativas han venido a mostrar el creciente descontento de la comunidad científica por los efectos perniciosos del sistema de evaluación, que ha fomentado no sólo conductas inapropiadas (publicar a toda costa) sino todo un negocio editorial basado en el pago por la publicación, en lugar del pago por la suscripción (Delgado López-Cózar, 2018). Además, ha potenciado el “efecto expulsión” de personal científico con prácticas, hábitos y culturas de publicación diferentes, pero no por ello de menor calidad. En otros casos ha forzado a los investigadores a modificar sus hábitos de publicación, especialmente en ciertas disciplinas sociales, técnicas y humanísticas, así como en ciertos colectivos, especialmente el de las personas en etapas tempranas de su carrera investigadora.

Mediante la reformulación de estos sistemas se busca evaluar la calidad, el rendimiento e impacto de sus investigadores, así como un juicio más cualitativo (que no subjetivo) junto a un uso más responsable de los indicadores bibliométricos (relacionado con el movimiento *Responsible Research Metrics*). Esto requiere de una serie de cambios culturales y sistémicos hacia un sistema en el que se promueva una valoración tanto cualitativa como cuantitativa. Un sistema el que se reconozca además la multidimensionalidad del sistema de evaluación a través de méritos tanto intrínsecos como de rendimiento (por ejemplo, considerando las actividades de mecenazgo, los roles de liderazgo, la divulgación de los hallazgos científicos a la comunidad, etc.) para conseguir la excelencia y el impacto. Un sistema que permita reforzar aún más la confianza de la sociedad en el sistema de investigación e innovación y en sus resultados, lo que se ha convenido en denominar la “transferencia social” o “transferencia y sociedad”.

La necesidad de una evaluación multidimensional ya se destacaba en la CAM (*Career Assessment Matrix*) (O'Carroll et al., 2017); también lo refleja la iniciativa holandesa “*Room for everyone's talent*” o el Marco para la Evaluación de los Investigadores de la red universitaria LERU (*League of European Research Universities*) en enero de 2022, a través del informe *A Pathway towards Multidimensional Academic Careers. A LERU Framework for the Assessment of Researchers* (LERU, 2022).

La intensificación de los procesos de evaluación de la investigación (especialmente a inicios del siglo XXI) empezó a suscitar un amplio debate sobre sus posibles efectos a nivel individual, como el aumento artificial de las publicaciones (Jiménez-Contreras et al. 2003) y de la internacionalización de la investigación (Den Besselaar, Heyman y Sandström, 2017), así como la intensificación de malas prácticas derivadas de la necesidad de publicar (por ejemplo, el *salami slice*). A pesar de ello, algunos estudios cuestionan si esto es realmente una consecuencia directa del sistema de evaluación (Osuna et al, 2011).

Como resultado, se están realizando múltiples esfuerzos en torno a la idea de las Evaluaciones de Investigación Responsable (*Responsible Research Assessment*) por parte de actores clave, como la *Comisión Europea*, que apuestan por “enfoques de evaluación que incentiven, reflejen y recompensen las características plurales de la investigación de alta calidad, en apoyo de investigaciones diversas e inclusivas” (*Comisión Europea*, 2021).

Ejemplo de este interés es la reforma de la evaluación de la investigación que está liderando la *Comisión Europea* (2021), que está incluida asimismo como una prioridad en la agenda política de la *European Research Area* (ERA). En esta línea, la *Comisión Europea* publicó el informe *Towards a Reform of the Research Assessment System* (2021) con el objetivo de establecer un marco de referencia y un proceso de cocreación para reformar el sistema actual de evaluación de la investigación. El documento presenta las conclusiones de una serie de consultas en profundidad realizadas de marzo a noviembre de 2021 con 130 agentes involucrados (tanto europeos como internacionales), en los que se propone un enfoque coordinado mediante una coalición de organizaciones investigadoras y/o financiadoras comprometidas para poder facilitar y acelerar esta transición. El documento incluye no sólo el compromiso de estos agentes sino también identifica objetivos y remarca principios básicos (el *European Code of Conduct of Research Integrity* y recomendaciones como las de DORA), así como un

plan de implementación con acciones entre organizaciones investigadoras y/o financiadoras para transformar estos conocimientos en cambios efectivos.

En un nuevo *momentum* en el que hay un reconocimiento en la comunidad científica cada vez mayor a las prácticas de investigación enmarcadas en la ciencia abierta (publicación de los datos de investigación, métodos, software, etc.), estas prácticas se ven obstaculizadas por los sistemas actuales de evaluación que, en general, no tienen en cuenta sus logros en este sentido. Para acelerar el proceso de ciencia abierta y los nuevos procedimientos de evaluación, también se necesita un compromiso práctico de implementación por parte de todos los agentes involucrados (Méndez, 2021).

Los sistemas de evaluación de la investigación (*Research Evaluation Systems*, RES) basados fundamentalmente en indicadores bibliométricos se implementaron hace décadas en muchos países como por ejemplo España, Italia, Reino Unido o China, entre otros. Además, en algunos casos, existen agencias de evaluación de acreditación externa a las propias universidades u organizaciones (como Francia, Italia o España). Sin embargo, existen algunos países donde la evaluación de la investigación se está abordando de forma diferente, como Holanda, donde todas las Universidades han establecido la *Strategy Evaluation Protocol* (SEP 2021-2027)⁴.

Dada la necesidad urgente de mejorar la forma mediante la cual los organismos de financiación, las instituciones académicas y otras partes evalúan el resultado de la investigación científica, un grupo de editores de revistas académicas se reunieron durante la *Reunión Anual de la Sociedad Americana de Biología Celular* (*American Society for Cell Biology*, ASCB) en San Francisco, California, el 16 de diciembre de 2012. Este grupo desarrolló un conjunto de recomendaciones, conocidas como la *Declaración de San Francisco sobre Evaluación de la Investigación* (DORA, 2012), que ha sido el inicio de una serie de iniciativas, declaraciones, manifiestos, y posicionamientos que tienen como objetivo principal reclamar una evaluación de la investigación responsable, así como premiar la integridad de la evaluación en los procesos de evaluación de los investigadores.

Esta postura también está respaldada asimismo por redes de universidades, por ejemplo la *League of European Research Universities* (LERU, 2022) o la *Young*

⁴ https://www.universiteitenvannederland.nl/files/documenten/Domeinen/Onderzoek/SEP_2021-2027.pdf

European Research Universities Network (YERUN, 2021), que destacan la necesidad de considerar otros elementos en la evaluación de la investigación más allá del Factor de Impacto de las revistas u otros indicadores bibliométricos. Estas redes universitarias han contribuido a este debate desde una perspectiva mucho más positiva, mostrando que el futuro de la evaluación del personal investigador podría abrirse hacia un marco general de evaluación de las carreras académicas multidimensionales, combinando criterios cuantitativos y cualitativos.

En los últimos años han surgido además distintos actores (“*movers and shapers*”) como *International Network of Research Management Societies* (INORMS), *Global Young Academy*, *Human Metrics Initiative* (HuMetricsHSS), *European University Association* o *Science Europe*, que promueven todas aquellas iniciativas que han respondido a esos desafíos, para reducir los efectos negativos de la evaluación de la investigación y mejorar la robustez de los sistemas de evaluación.

Desde la publicación de DORA se han ido sucediendo una serie de manifiestos muy importantes en la misma línea, pero destaca la alta concentración de declaraciones que se ha producido entre 2021 y 2022, donde destaca especialmente el reciente *Agreement on Reforming Research Assessment*, impulsado por la *Coalition for Advancing Research Assessment* (COARA) (European Commission et al., 2022) y firmado recientemente por la Comisión Europea⁵, en el que se apuesta por una evaluación cualitativa, lo que supone un choque con las prácticas evaluativas llevadas a cabo en España (Ràfols y Molas-Gallart, 2022).

Se trata por tanto una conversación que va a más a nivel internacional, creciendo en intensidad. Algunas declaraciones tienen menos recorrido, otras son más estructurales y otras son más ideológicas, pero todo el movimiento va encaminado a cambiar la tendencia.

En la Tabla 1 se ofrece un listado de todas aquellas iniciativas relacionadas con los modelos y sistemas de evaluación de la investigación que se han publicado desde el lanzamiento de la *Declaración de DORA* (2012) y hasta la más reciente en 2022 de Coalición para la Reforma de la Evaluación de la Investigación (*Coalition on Reforming Research Assessment*, COARA) impulsada por *Science Europe* y *European University*

⁵ https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/commission-signs-agreement-reforming-research-assessment-and-endorses-san-francisco-declaration-2022-11-08_en

Association (EUA), junto con el apoyo de la *Comisión Europea*. A continuación, se recogen tanto el año de publicación y la fuente (páginas web), así como los principales elementos de discusión y la relevancia de cada uno de estos documentos.

Tabla 1.

Iniciativas de evaluación de la ciencia alternativas y sus implicaciones sobre los sistemas de evaluación en un futuro

Documento	Año	Relevancia
<i>San Francisco Declaration on Research Assessment (DORA)</i> https://sfdora.org	2012	<p>Primera iniciativa a nivel mundial desarrollada durante la reunión anual de la <i>Sociedad Americana de Biología Celular</i> (ASCB) en San Francisco. Está formada por un conjunto de recomendaciones para agencias de financiación, instituciones, editores y organizaciones que proporcionan métricas e investigadores, para cambiar el sistema de evaluación.</p> <p>Entre las recomendaciones destaca: la necesidad de eliminar el uso de métricas basadas en revistas (<i>Journal Impact Factor</i>, JIF) en convocatorias de financiación y promoción, evaluar la investigación por sus propios méritos (en lugar de la revista en la que se publica); y capitalizar las oportunidades que brindan las publicaciones online (p. ej., para explorar nuevos indicadores de impacto).</p> <p>En la fecha en la que se escribe este informe (septiembre, 2022), han firmado esta declaración 22.080 personas y organizaciones en 159 países.</p>

***LERU Report on Research
Universities and Research
Assessment***

<https://www.leru.org/files/Research-Universities-and-Research-Assessment-Full-paper.pdf>

2012

Este informe pone de relieve la explosión de los tipos de evaluaciones de la investigación para una gran variedad de usuarios y propósitos, alertando de la creciente obsesión con la medición y la monitorización (cultura de “contar”) en contraposición a la cultura de la calidad de la investigación, y de la búsqueda ilimitada de nuevo conocimiento.

Surge como respuesta dada por las universidades de la *LERU* dado que la investigación universitaria está financiada en gran medida por dinero público, a menudo asignado competitivamente en Europa, con lo cual es normal que se evalúe regularmente como una cuestión de responsabilidad.

Los objetivos, procesos y criterios utilizados deben definirse de manera clara y transparente. Se pone el foco de atención en que la Bibliometría proporciona una alternativa coste-eficaz a la revisión por pares, pero también tiene inconvenientes considerables. Por ese motivo se recomienda que la Bibliometría sea utilizada como complemento de la revisión por pares. Ambos métodos deben usarse con sabiduría, discreción y la aplicación rigurosa del juicio humano. Una buena evaluación probablemente requiera de un conjunto de metodologías y, sin duda, de una buena gestión de datos.

Las universidades deben mantenerse firmes en la defensa del valor a largo plazo de su actividad investigadora, que no es fácil de evaluar en una cultura donde el retorno de la inversión se mide en períodos de tiempo muy cortos.

Independent Peer Review

2014

Manifiesto

<https://www.openscholar.org.uk/independent-peer-review-manifesto>

Manifiesto elaborado por *Open Scholar C.I.C.* (organización sin fines de lucro) en el que se remarca que la evaluación de la investigación está controlada por revistas académicas que, a su vez, son responsables del acceso al conocimiento científico. Se propone una evaluación de la investigación más objetiva mediante diferentes recomendaciones como acceso público gratuito inmediato; revisión por pares independientes; evaluación en abierto, entre otras..

En el momento de escribir este informe, esta Declaración ha sido firmada por 340 personas.

**Leiden Manifesto for Research
Metrics**

<http://www.leidenmanifesto.org>

<https://www.nature.com/articles/5204>

[29a](#)

2015

El *Manifiesto de Leiden* es una lista de “diez principios para guiar la evaluación de la investigación”, publicado como comentario en *Nature* el 22 de abril de 2015.

En concreto, el Manifiesto defiende que (1) la evaluación cuantitativa debe respaldar la evaluación cualitativa de expertos; (2) se debe medir el desempeño frente a las misiones de investigación de la institución, grupo o investigador; (3) proteger la excelencia en la investigación relevante a nivel local; (4) mantener la recopilación de datos y los procesos analíticos abiertos, transparentes y simples; (5) permitir que los evaluados verifiquen los datos y el análisis; (6) tener en cuenta la variación por campo en la práctica de publicación y citación; (7) basar la evaluación de los investigadores individuales en un juicio cualitativo de su trayectoria; (8) evitar la concreción fuera de lugar y la falsa precisión; (9) reconocer los efectos sistémicos de la evaluación y los indicadores; y (10) examinar los indicadores periódicamente y actualizarlos. Este informe ha sido implementado en muchas universidades en Países Bajos y se recoge en el protocolo de la Estrategia de Evaluación (*Strategy Evaluation Protocol*, SEP 2021-2027).

<p>The Metric Tide</p> <p>https://www.ukri.org/publications/review-of-metrics-in-research-assessment-and-management/</p> <p>https://responsiblemetrics.org/2022/08/11/the-metric-tide-revisited/</p>	<p>2015</p>	<p>Informe independiente coordinado por James Wilsdon en el que, a través de una revisión exhaustiva de literatura, propone un marco para métricas responsables así como veinte recomendaciones específicas. Este informe ha sido empleado para aplicar métricas responsables en la gestión de la investigación, por ejemplo, en el REF (<i>Research Excellence Framework</i>), el sistema de evaluación de la calidad de la investigación en el Reino Unido). El informe está actualmente en proceso de actualización⁶.</p>
<p>Next Generation Metrics</p> <p>https://data.europa.eu/doi/10.2777/337729</p>	<p>2017</p>	<p>Informe elaborado por un grupo de expertos en Altmetrics en el que se describe el uso, potencial y limitaciones de las conocidas como '<i>Next Generation Metrics</i>' (NGM). Como recomendación se afirma lo siguiente: "La Comisión Europea debería fomentar el desarrollo de nuevos indicadores, y evaluar la idoneidad de los existentes, para medir y apoyar el desarrollo de la ciencia abierta". En el ámbito de la evaluación, los autores justifican el uso de las NGM afirmando que éstas tienen el potencial de complementar algunos indicadores como los de impacto (número de citas). Entre las recomendaciones, se apuesta por un sistema de ciencia abierta que combine expertos y medidas cuantitativas y cualitativas para fines de evaluación. Finalmente, se propone un listado de potenciales indicadores.</p>

⁶ <https://responsiblemetrics.org/2022/08/11/the-metric-tide-revisited/>

<p>Open Science Career Assessment Matrix</p> <p>2017</p> <p>https://data.europa.eu/doi/10.2777/75255</p>	<p>Informe elaborado por Grupo de Trabajo sobre 'Rewards under Open Science', en el que se presenta una propuesta para la evaluación de carreras científicas, reconociendo las prácticas de ciencia abierta. A pesar de que el informe se centra más en este tipo de prácticas, se reconoce la necesidad de nuevas medidas para medir la calidad de la evaluación y se propone una <i>Open Science Career Matrix</i>, que incluye diferentes indicadores para evaluar la investigación.</p>
<p>6 Principles for Assessing Scientists for Hiring, Promotion and Tenure</p> <p>2018</p> <p>https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2004089</p> <p>https://blogs.lse.ac.uk/impactofsocialsciences/2018/06/04/six-principles-for-assessing-scientists-for-hiring-promotion-and-tenure/</p> <p>http://eprints.lse.ac.uk/90753/</p>	<p>Entrada de blog basada en un artículo en el que varios autores (entre los que se incluyen David Moher y John Ioannidis) proponen seis principios para la evaluación de investigadores en procesos de contratación y promoción, tomando como base las interacciones y <i>feedback</i> obtenidos durante un <i>workshop</i> celebrado en enero de 2017 en Washington DC para discutir y proponer estrategias a la hora de contratar, promocionar y estabilizar a personal de investigación.</p> <p>Entre los principios propuestos se destacan los tres siguientes, más relacionadas con evaluación:</p> <ol style="list-style-type: none"> 1) La contribución a las necesidades de la sociedad es un objetivo importante en el mundo académico (principio 1); 2) La evaluación de los investigadores y docentes debe basarse en indicadores responsables que reflejen más plenamente la contribución científica (principio 2); 3) Las ideas innovadoras (<i>out of the box</i>) para financiación deben valorarse en las decisiones de promoción (principio 6).

<p><i>The Hong Kong Principles for Assessing Researchers: Fostering Research Integrity</i></p> <p>https://doi.org/10.1371/journal.pbio.3000737</p>	<p>2019</p>	<p>Estos principios se formularon y aprobaron en la 6ª Conferencia Mundial sobre integridad de la investigación en Hong Kong. Se enfocan en premiar la integridad de la evaluación en los procesos de evaluación de los investigadores.</p> <p>A pesar de que también pone el foco sobre la evaluación de las prácticas de ciencia abierta, en cuanto a evaluación se destacan los siguientes principios:</p> <ol style="list-style-type: none"> 1) Evaluar a los investigadores sobre prácticas responsables desde la idea hasta la finalización y posterior difusión de la investigación (principio 1); 2) Valorar una amplia gama de investigaciones y estudios, como la replicación, la innovación, la traducción, la síntesis y la meta-investigación (principio 4); 3) Valorar otras contribuciones a la investigación, como revisión por pares para evaluación de proyectos y publicaciones, tutoría, divulgación e intercambio de conocimiento (principio 5).
<p><i>The EU's Open Science Policy</i></p>	<p>2021</p>	<p>Es la política lanzada por la <i>Comisión Europea</i> con relación a la -ciencia abierta. En cuanto a evaluación, se destacan el uso de '<i>Next-Generation Metrics</i>', es decir, nuevos indicadores para complementar los indicadores convencionales de calidad e impacto de la investigación, con el fin de recompensar las prácticas de ciencia abierta de los investigadores.</p>

<p>Unesco Recommendation on Open Science (UNESCO, 2021b)</p> <p>2021</p> <p>https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa</p>	<p>Esta recomendación fue adoptada por la <i>Conferencia General de la UNESCO</i> en noviembre de 2021. Sobre evaluación de los investigadores, se hace especial énfasis en revisar los sistemas de evaluación de la investigación y la evaluación de las carreras científicas para alinearlos con los principios de la ciencia abierta.</p> <p>Teniendo en cuenta que un compromiso con la ciencia abierta requiere tiempo, recursos y esfuerzos, los sistemas de evaluación deben tener en cuenta la amplia variedad de formas de creación y comunicación de conocimiento, que van más allá de los formatos tradicionales (es decir, publicaciones en revistas científicas).</p>
<p>UKRN Statement on Responsible Research Evaluation</p> <p>2021</p> <p>https://doi.org/10.31219/osf.io/4pqwv</p>	<p>Declaración elaborada por el <i>UKRN Steering Group</i> en el que se llama a un uso responsable de las métricas para la evaluación, dirigida principalmente a instituciones. En este sentido, se proponen once principios en los que se hace una crítica a conceptos como la calidad de la investigación, los indicadores (por ejemplo, evitar aplicar ciertas métricas a investigadores individuales, asegurarse de que no se evalúe los artículos sólo por el Factor de Impacto de la revista, etc.).</p> <p>Sobre evaluación y el uso de las métricas, se afirma lo siguiente: “Cuando sea necesario, como en la evaluación de investigadores individuales, elegir una fuente que permita verificar y tratar los registros para garantizar que los registros sean completos y precisos, o comparar las listas de publicaciones con los datos de los sistemas de la institución”.</p>

<p><i>YERUN Position Paper</i></p> <p>https://yerun.eu/wp-content/uploads/2021/11/YERUN-Position-Paper_Research-assessment.pdf</p>	2021	<p>Declaración (o <i>position paper</i>) elaborada por la red de universidades <i>YERUN</i> en la que se proponen siete recomendaciones para reformar la evaluación de la investigación en las universidades, haciéndola más inclusiva, y atractiva para el sector universitario.</p> <p>Entre las recomendaciones se destacan las siguientes:</p> <ol style="list-style-type: none">1) Comprender qué significa “evaluación de la investigación” para las universidades.2) Fomentar la circulación de las buenas prácticas existentes, abrazando la diversidad y respetando la autonomía (por ejemplo, incluyendo a más actores externos en la evaluación de la investigación).3) Distinguir entre la discusión sobre la reforma de la evaluación de la investigación y la discusión sobre la carrera precaria de los investigadores.4) Permitir el diálogo constructivo entre una amplia gama de actores.
--	------	--

***LERU Framework for the
Assessment of Researchers***

https://www.leru.org/files/Publications/LERU_PositionPaper_Framework-for-the-Assessment-of-Researchers.pdf

2022

Informe elaborado por el Profesor Bert Overlaet, de la red universitaria *LERU*, en el que se plantea un marco multidimensional para la evaluación de los investigadores. Se traza un conjunto de dimensiones tradicionales (investigación, servicio a la institución, compromiso público y divulgación y educación) y una serie de indicadores, algunos de ellos novedosos, para su evaluación. Por ejemplo, en cuanto a la evaluación de la investigación, se listan criterios como la contribución en el área, la colaboración o la interdisciplinariedad, entre otros. Además, se añaden otras dimensiones nuevas como liderazgo (en redes, agrupaciones y asociaciones científicas, responsabilidad de personas), colaboración e innovación (incluyendo las lecciones aprendidas de los fracasos) en la perspectiva de desarrollo y atención al contexto profesional (balance entre disciplinariedad y multidisciplinariedad, situación particular de la universidad) y personal (edad académica, género y cargas familiares y docentes) del investigador en la perspectiva de contexto.

**Paris Call on Research
Assessment**

<https://osec2022.eu/paris-call/>

2022

Este texto fue preparado por el *Comité de Ciencia Abierta de Francia* y presentado en la *Conferencia Europea de Ciencia Abierta de París (OSEC)*, que se celebró en París los días 4 y 5 de febrero de 2022.

Es una llamada a la reforma del sistema de evaluación. Para ello, propone una serie de recomendaciones entre las que destaca la necesidad de diseñar un sistema que premie la calidad y los diferentes tipos de impacto; se asegure que la investigación cumpla con los más altos estándares de ética e integridad; se valore la diversidad de actividades y resultados de investigación; utilice criterios y procesos de evaluación que respeten la variedad de disciplinas de investigación o que valore el trabajo colaborativo, así como la interdisciplinariedad y la ciencia ciudadana, cuando sea aplicable.

También se pide la creación de una coalición de organizaciones comprometidas a reformar el actual sistema de evaluación de la investigación en torno a objetivos, principios y acciones acordados en común.

<p><i>Agreement on Reforming Research Assessment (COARA)</i></p> <p><i>https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf</i></p>	<p>2022</p>	<p>Este Acuerdo es el resultado de un proceso de cocreación iniciado en enero de 2022 para establecer una coalición entre varios grupos (investigadores, organizaciones científicas, etc.) para cambiar el sistema de evaluación de la investigación. Fue impulsado por <i>Science Europe</i>, <i>European University Association</i> (EUA), junto con el apoyo de la <i>Comisión Europea</i>.</p> <p>Incluye principios, compromisos y plazos para las reformas y establece una serie de acciones para trabajar de forma conjunta en la implementación de cambios.</p> <p>En este proceso, han participado más de 350 organizaciones de investigación de más de 40 países.</p> <p>La implantación del acuerdo será llevado a cabo a través de una Coalición voluntaria creada por la <i>Comisión Europea</i> y denominada <i>Coalition on Reforming Research Assessment (COARA)</i>.</p> <p>A fecha de 8 de noviembre, el acuerdo ha sido firmado por 183 instituciones, incluyendo la propia <i>Comisión Europea</i>⁷.</p>
--	-------------	---

⁷ https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/commission-signs-agreement-reforming-research-assessment-and-endorses-san-francisco-declaration-2022-11-08_en

3. Revisión de la literatura sobre evaluación de la ciencia

En este apartado se presenta una revisión de la literatura científica sobre evaluación de la ciencia, cuyo propósito principal es identificar las principales escuelas y polos de discusión en la literatura a partir del análisis bibliográfico descriptivo de los documentos más relevantes en el ámbito. Este análisis pretende en última instancia confrontar lo que la literatura científica estudia, analiza y discute (dimensión científica) con lo que las Agencias de evaluación implementan a la hora de diseñar modelos de evaluación reales, aplicados a la comunidad científica (dimensión profesional). Por tanto, el objetivo perseguido no es el de recopilar toda la producción científica sobre el tema analizado, sino identificar los trabajos más relevantes y, a partir de éstos, identificar los principales temas tratados y que aspectos han despertado más discusión en la literatura internacional.

3.1. Metodología

Los pasos seguidos para esta revisión sistemática se detallan a continuación.

- a) **Elaboración de la estrategia de búsqueda.** En primer lugar, se diseñó una estrategia adecuada para identificar la literatura científica sobre este tema. Para ello se definieron una serie de conceptos clave como:

'individual scientific performance', 'research evaluation system', 'evaluation of performance', 'effects of evaluation', 'responsible research assessment', 'evaluative bibliometrics', 'research evaluation', 'research assessment', 'author-level metrics'.

Una vez identificados los términos más relevantes, se determinó la idoneidad de estos términos a través de distintas búsquedas en un base de datos bibliográfica multidisciplinar (en este caso, *Scopus*). Finalmente se decidió utilizar una estrategia de búsqueda formada únicamente por tres de las palabras clave identificadas previamente (*evaluative bibliometrics; research evaluation; research assessment*), pues resultaban ser las más precisas y las que recopilaban una mayor cantidad de resultados relevantes. Con el fin de eliminar ruido documental (documentos no relacionados con el tema) de la estrategia, se procedió a restringir la búsqueda de esos términos a su presencia en el título y las palabras clave de los documentos.

b) Descarga de los datos. Para esta fase del análisis, se utilizó la base de datos multidisciplinar de *Scopus (Elsevier)*, a partir de la siguiente ecuación de búsqueda:

(KEY ("evaluative bibliometric*" OR "research* evaluation*" OR "research assessment*") AND TITLE ("evaluative bibliometric*" OR "research* evaluation*" OR "research assessment*"))

Adicionalmente, se utilizó la base de datos de *Dimensions (Digital Science)*, pues presenta una mayor cobertura de documentos (Singh et al., 2021). Dado que *Dimensions* no permite la búsqueda delimitada por palabras clave (sólo permite las opciones de búsqueda a texto completo, título y resumen, y por DOI), se utilizó la búsqueda a texto completo, mediante el uso de las palabras clave identificadas anteriormente.

En *Scopus* se identificaron 260 documentos (incluyendo todas las tipologías documentales), mientras que en *Dimensions* se identificaron 5.633 documentos, debido al criterio de búsqueda usado más general. Dado el gran volumen de documentos no relacionados con el tema (por ejemplo, la evaluación de algunas enfermedades), se decidió realizar una revisión manual en ambos *datasets* para descartar los documentos no pertinentes. En el caso de *Dimensions*, este filtrado se restringió a aquellos documentos con al menos 50 citas recibidas (463 documentos, 8,22% del *dataset*). Aun cuando es una decisión pragmática, y dado que el fin no es realizar un *scoping review* de la literatura sino identificar los documentos más relevantes sobre el tema, este filtrado resultó conveniente para los objetivos del análisis. Una vez depurados los datos, se seleccionaron 212 documentos (81,53%) en *Scopus* y 166 documentos (36%) en *Dimensions*.

Finalmente se combinaron los registros obtenidos por *Scopus* y *Dimensions* y se eliminaron los duplicados, obteniendo 356 registros únicos, con un solape sólo de 22 documentos, que constituyen el *dataset* analizado (Tabla 2 y Figura 1). La descarga de los datos se realizó en Julio 2022. Este conjunto de registros pasó a la siguiente etapa de filtrado temático.

Tabla 2.

Resumen de los documentos seleccionados

Datos	Total documentos	Documentos no relacionados	Documentos seleccionados
Scopus	260	48	212
Dimensions	5633 (463*)	297	166
Solapamiento	203**		22
Total	5690	345	356

* Documentos con, al menos, 50 citas

** Solape con el *dataset* entero de Dimensions (n=5633)

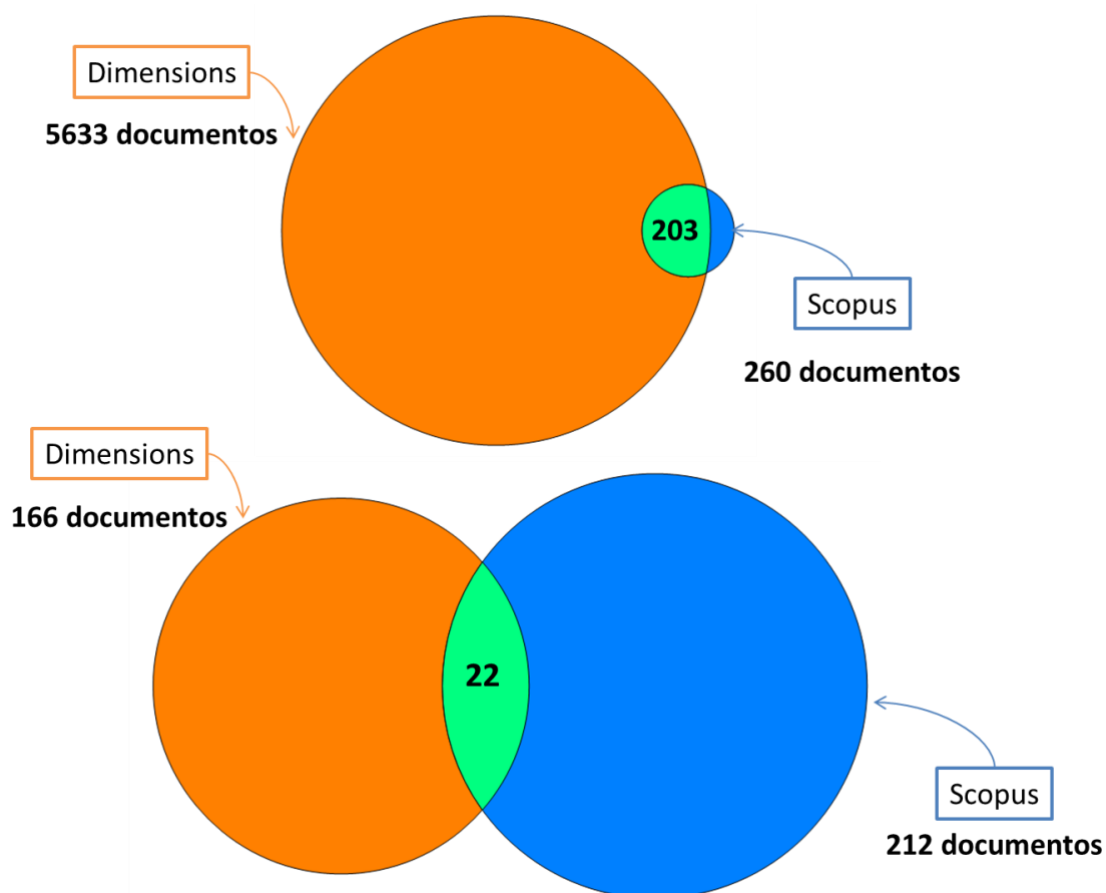


Figura 1.

Solapamiento de los documentos identificados entre las dos bases de datos, antes de filtrado (arriba) y tras el filtrado (abajo)

Nota: el solape se ha calculado a través del DOI

c) Identificación de temas. Una vez recopilados todos los registros, se procedió a clasificar cada uno de los documentos identificados en función de la temática

que abordaba mediante la lectura del resumen, título y palabras clave. Debido a la estrategia de búsqueda utilizada, aquellos estudios con implicaciones directas o indirectas en evaluación de la investigación, pero que no mencionan explícitamente este concepto (por ejemplo, trabajos que hablan de fuentes bibliográficas o de métricas alternativas) pueden haber sido obviados. Existe asimismo una revista de gran relevancia en el área, titulada *Research evaluation*, en la que obviamente todos los trabajos publicados están relacionados en mayor o menor medida con la evaluación de la ciencia, pero que no son identificados si no mencionan explícitamente los conceptos de búsqueda.

Para cada publicación se indicó asimismo la localización (por ejemplo, de Reino Unido si hacía referencia al sistema *Research Assessment Exercise*, RAE) y el campo de análisis (ciencias sociales, humanidades, ingenierías, etc.). En este sentido, se han identificado temas que abarcan desde la revisión a los diferentes indicadores utilizados en evaluación hasta la descripción de los diferentes sistemas de evaluación (por ejemplo, Italia o Reino Unido), así como distintas propuestas alternativas. Mediante este procedimiento adicional de filtrado, se eliminaron aquellos documentos que no tenían una relación directa con el tema. En concreto, se descartaron un total de 48 documentos (18,46%) en *Scopus* y 319 en *Dimensions* (68,9%), dando como resultado un total de 356 documentos únicos, que constituyen el *dataset* final. Para estas publicaciones, se muestran los siguientes indicadores:

- a. Evolución del número de documentos evaluación de la investigación durante el período temporal de 1987 a 2022.
- b. Identificación de la localización de los sistemas de evaluación analizados, así como las disciplinas que se evalúan.
- c. Fuentes en las que se ha publicado los documentos clave sobre evaluación de la ciencia.
- d. Identificación de áreas clave. Para ello, se llevó a cabo un análisis de coocurrencia de las palabras clave para identificar los temas de interés, mediante el software VOSviewer. Los diferentes nodos corresponden con palabras clave, cuyo tamaño tiene relación con el número de documentos (a mayor tamaño, mayor número de documentos). Las palabras clave están unidas por enlaces en función de la coocurrencia (el número de veces que los términos aparecen juntos), mientras que el grosor de la línea indica la frecuencia (un mayor grosor implica que los dos términos aparecen más veces mencionados conjuntamente). Finalmente, el color indica el grupo (*cluster*) al que el término está vinculado temáticamente. Dado que *Dimensions* no dispone de palabras

clave, se asignaron unas palabras clave de forma manual en base a la clasificación del análisis de contenido utilizado.

- e. Identificación de polos de discusión en base al análisis de contenido. Una vez obtenida la perspectiva general con el análisis de coocurrencia, este apartado se basa en la clasificación manual más exhaustiva de los contenidos en base a la información de título, resumen, palabras clave (y, en algunos casos, texto completo) de cada uno de los textos identificados por grandes agrupaciones de temas.

3.2. Resultados

Análisis descriptivo del dataset

La Figura 2 muestra el número y evolución de documentos por año, mientras que la Figura 3 muestra esta evolución desglosada por las dos bases de datos utilizadas (*Scopus* y *Dimensions*). La publicación más antigua localizada corresponde con el artículo “*A review of bibliometric and other science indicators and their role in research evaluation*”, de King Jean (1987), en el que se realiza una revisión de los indicadores bibliométricos. Con un 47,47%, la producción se concentra de forma reciente, desde el año 2015, lo que va alineado también con los diferentes debates surgidos sobre el tema (por ejemplo, por parte de la *Comisión Europea*).

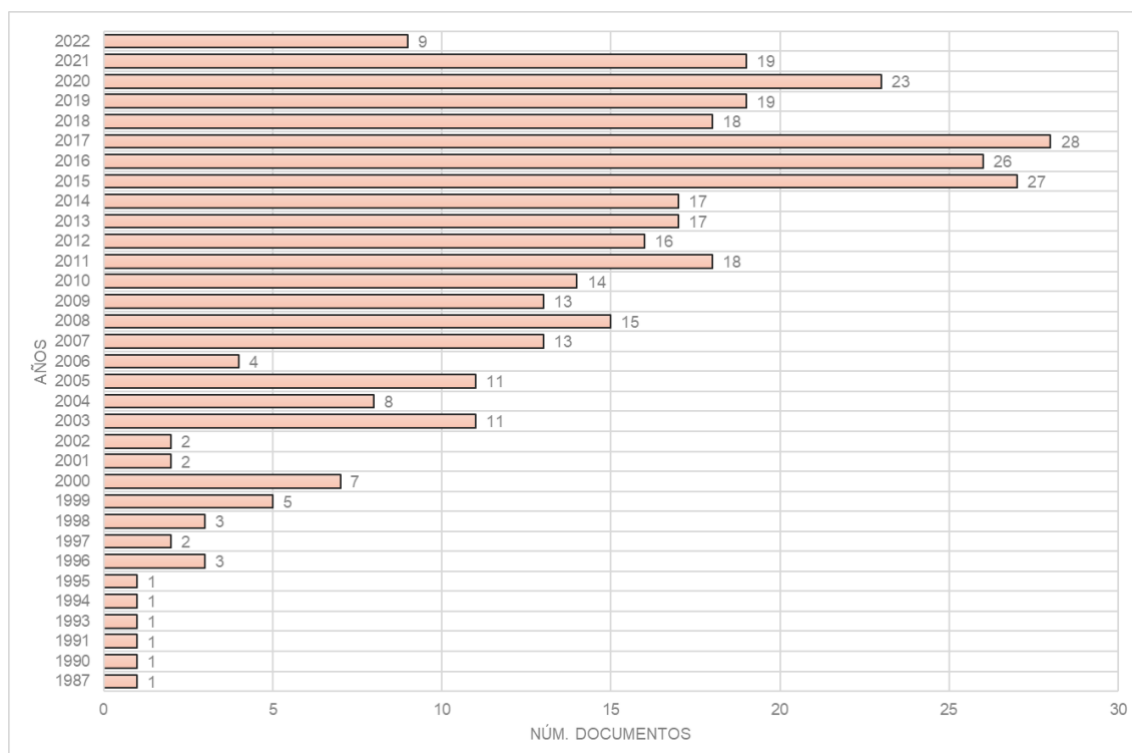


Figura 2.

Número de documentos totales por año (1987-2022)



Figura 3.

Evolución del número de documentos (1987-2022)

La Figura 4 muestra por su parte la distribución de estas publicaciones por revistas. *Scientometrics* es la revista que ha publicado un mayor número de artículos en los que se menciona explícitamente la evaluación de la investigación, según los criterios de búsqueda empleados (53; 14,89%), seguido de *Journal of Informetrics* (22; 6,18%), *Journal of the Association for Information Science and Technology* (16; 4,49%), *Research Policy* (14; 3,93%) y *Research Evaluation* (11; 3,09%). Es especialmente destacable que la producción se encuentra publicada en revistas con un carácter bibliométrico (*Scientometrics* y *Journal of Informetrics*), en vez de otras fuentes con un enfoque más relacionado a otros ámbitos de la actividad científica más allá de los aspectos cuantitativos (por ejemplo, *Research Policy* o *Research Evaluation*), lo que refuerza la influencia de la Bibliometría en el estudio de la evaluación de la investigación. También destaca el hecho de que haya dos revistas relacionadas con el ámbito de educación superior (*Higher Education* y *Studies in Higher Education*) y, sin embargo, no aparezcan revistas de estudios de la comunicación y divulgación de la ciencia (*science communication*).

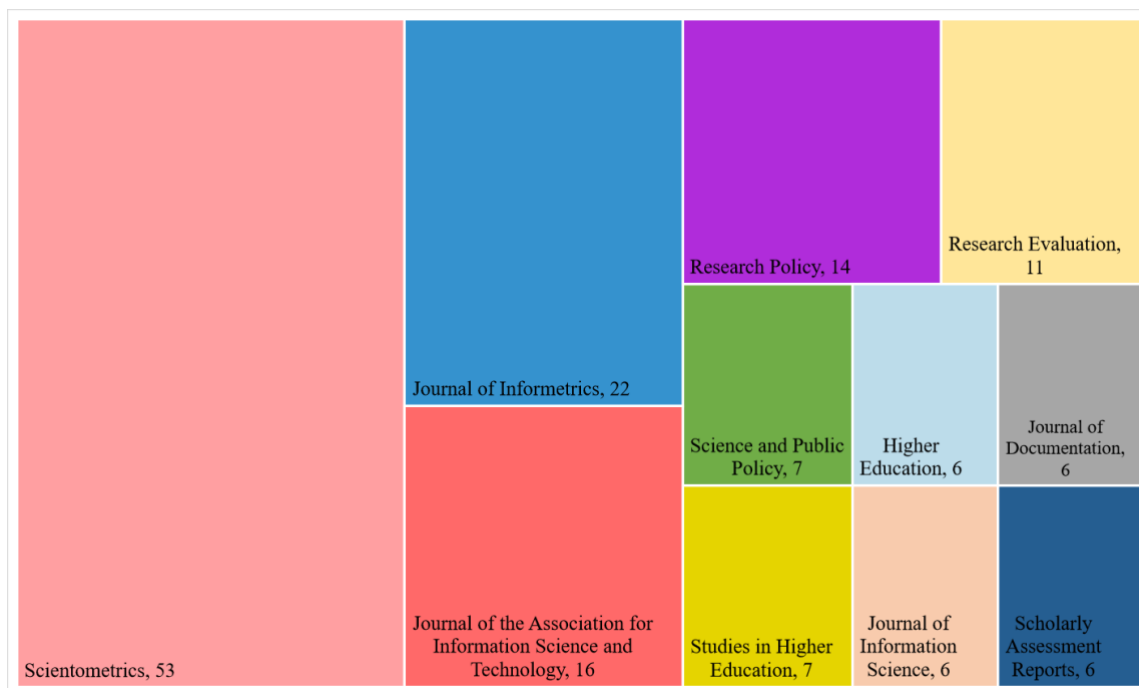


Figura 4.

Distribución por revistas de la producción científica en evaluación científica.

La Figura 5 muestra el mapa de coocurrencia (con una frecuencia mínima de 3) de la producción científica sobre evaluación científica, basado en las palabras clave de *Scopus* y las asignadas (de forma manual) en los artículos seleccionados de *Dimensions*.

Para este *dataset* se han identificado un total de seis grupos que se resumen a continuación:

- **Cluster #1** (rojo): “Evaluación científica y discusión de los indicadores”. Este grupo lo conforman aquellos nodos que tratan sobre evaluación, junto a indicadores bibliométricos (por ejemplo, aparece el *h-index* como uno de los nodos), de citas (por ejemplo, el nodo *citation analysis* o *citation counts*), indicadores alternativos (*altmetrics*), así como la discusión acerca del impacto social y el sistema de China. Constituye, por tanto, un *cluster* sobre la discusión del uso de indicadores para evaluación.
- **Cluster #2** (azul marino): “Evaluación de la investigación en diferentes contextos”. En este conjunto encontramos nodos sobre los sistemas de evaluación (por ejemplo, el sistema italiano), así como el uso de metodologías cuantitativas (por ejemplo, *metrics*) y cualitativas (por ejemplo, *peer review*) usadas en estos sistemas.

- **Cluster #3** (verde): "Evaluación en Reino Unido". Gran parte de los artículos identificados tratan sobre el sistema de evaluación británico. Dentro de este grupo encontramos los sistemas de evaluación del Reino Unido (*Research Excellence Framework, REF* y *Research Assessment Exercise, RAE*), así como las principales discusiones en torno a estos sistemas (por ejemplo, el concepto de *research excellence*).
- **Cluster #4** (amarillo): "Bases de datos y áreas". Dentro de este *cluster* encontramos la discusión de las diferentes bases de datos (*Google Scholar, Web of Science, Scopus*), especialmente su problemática en áreas como las Ciencias sociales y las Humanidades.
- **Cluster #5** (morado): "Responsabilidad y calidad de la investigación". Conceptos como responsabilidad, así como medición del desempeño (*performance measurement*) en la evaluación de la investigación conforman este *cluster*.
- **Cluster #6** (azul claro): "Ciencia abierta". En este grupo aparecen nodos como "ciencia abierta", "acceso abierto" o "comunicación científica", aspectos vinculados directa e indirectamente en procesos de evaluación de la investigación.

De cada uno de los artículos identificados, se procedió a identificar su localización (por ejemplo, si hacía referencia al sistema de evaluación británico, italiano, chino, etc.), así como el área (ciencias sociales, humanidades, etc.). La Figura 6 muestra la localización de los sistemas de evaluación analizados. Reino Unido (76 artículos) e Italia (43 artículos) lideran la clasificación, seguidos de Australia (14) y China (11). En el caso de este último, es importante resaltar que el *Ministerio de Educación* emitió una regularización para evitar la dependencia excesiva del uso de bases de datos como la *Web of Science (WoS)* y su evaluación, resaltando las contribuciones nacionales. Entre las recomendaciones indicadas por el ministerio chino se remarca divulgar una menor cantidad de artículos pero de mayor calidad, así como publicar un tercio de éstos en revistas nacionales prestigiosas según el *Ministerio de Educación de la República Popular China (2020)*. Por tanto, podemos hablar de una política proteccionista también en la ciencia producida en este país, que apunta una tendencia diferente a la globalización e internacionalización que está ocurriendo en la producción científica de otras naciones.

A nivel de regiones se localizan cinco artículos en el *dataset* que hacen referencia a sistemas de evaluación europeos, tres a sistemas americanos y uno al sistema africano y asiático, respectivamente.

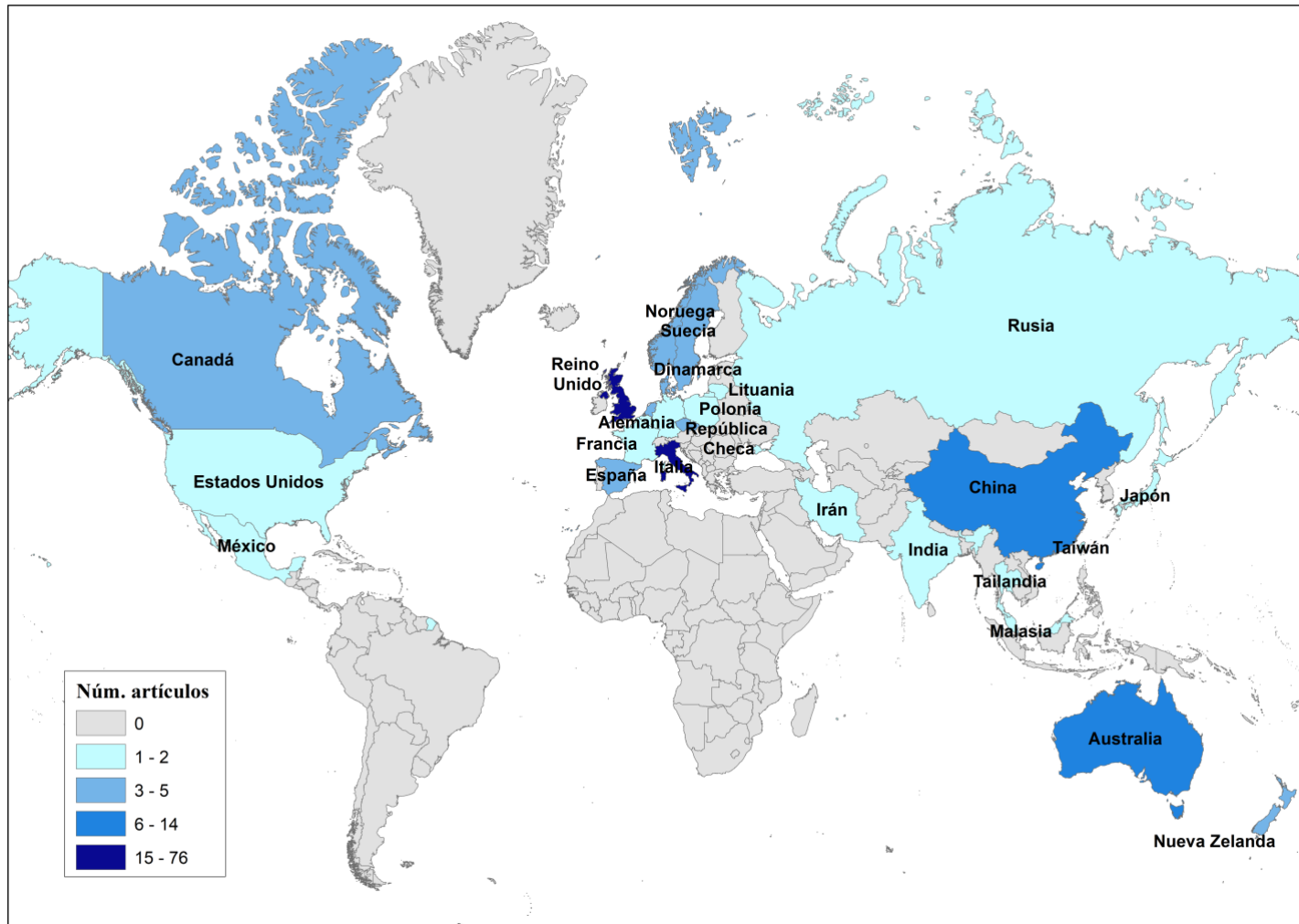


Figura 6.
Ubicación de los sistemas de evaluación analizados en este estudio

La figura 7 incluye las principales áreas en las que se enfocan los artículos de evaluación científica. La mayoría de estos trabajos analizan los sistemas de evaluación y sus resultados en áreas multidisciplinares (42 artículos, 11,8% del *dataset*), seguida de economía y negocios (27 artículos, 7,6%), ciencias naturales y aplicadas (19; 5,3%), ciencias sociales (16; 4,5%) y humanidades (9; 2,5%). Los artículos que abordan temas multidisciplinares se publican principalmente en revistas como *Scientometrics* (5 artículos), *Journal of Informetrics* (2 artículos), *Journal of the Association for Information Science and Technology* (2 artículos) y *Studies in Higher Education* (2 artículos).

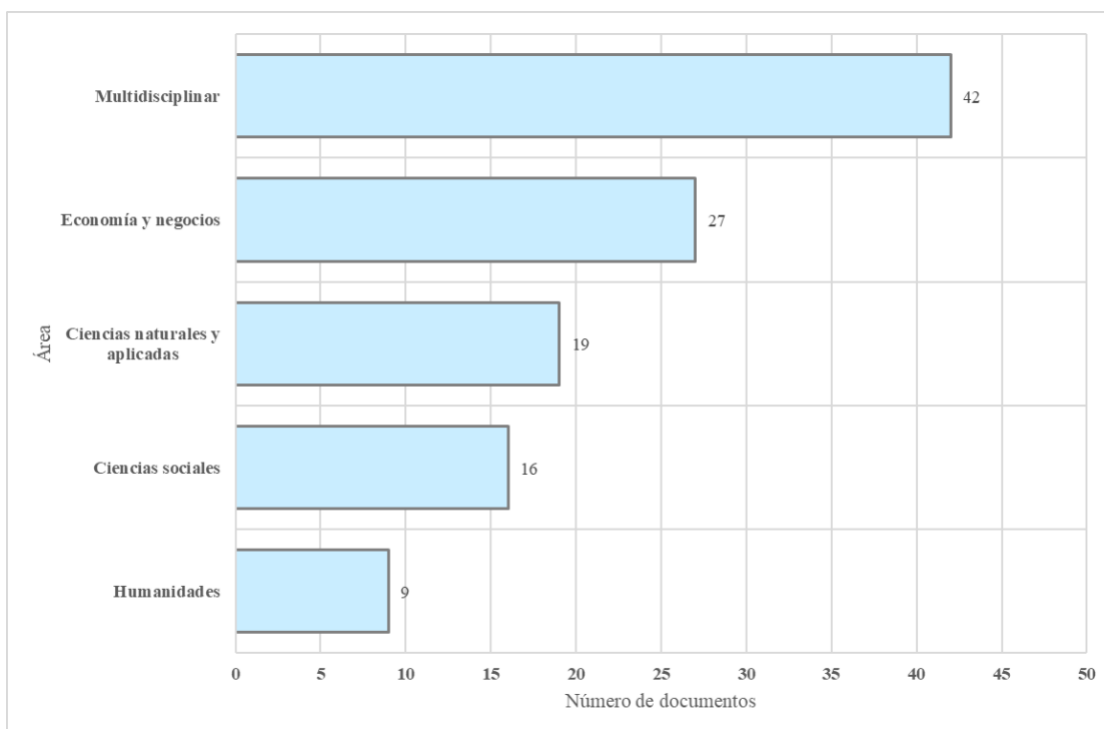


Figura 7.

Distribución de publicaciones por áreas temáticas

Nota: Las áreas hacen referencia a las analizadas por los artículos, no a la categorización de *Scopus* y *Dimensions* de las revistas.

Análisis de contenido (temáticas)

En función del análisis del contenido de los resúmenes, títulos y palabras clave, se procedió a identificar una serie de temas clave que se pueden agrupar en los siguientes puntos, y que complementan el análisis de corte más descriptivo ofrecido anteriormente.

1. **Recopilación de indicadores utilizados en evaluación de la investigación**

El principal objetivo de este grupo de publicaciones es realizar un análisis de los principales indicadores utilizados para evaluación científica, así como de sus ventajas e inconvenientes. En primer lugar, se encuentran estudios que evalúan el uso de métricas a nivel individual como, por ejemplo, Wildgaard (2015), quien analiza 17 indicadores a nivel de autor y afirma que los resultados de estos indicadores están basados en función de la base de datos utilizada (y de su cobertura y número de citas), la forma en la que se calculan los indicadores, la disciplina evaluada y la antigüedad. Kostoff (1997) destaca la complejidad de la interpretación de las métricas a nivel individual y que deben conceptualizarse y, en la medida de lo posible, complementarse.

En la literatura se encuentran muchos artículos de crítica a los indicadores, especialmente aquellos basados en el número de citas y el *Journal Impact Factor* (JIF) de las revistas. Abramo, Cicero y D'Angelo (2012) critican que el uso de las citas se vea como un *proxy* de calidad y proponen alternativas (normalización de citas). Kostoff (1998) también crítica el uso de las citas como medida de impacto, analizando el propósito de su uso (por ejemplo, como marcador, herencia intelectual, rastreador de impacto y propósito egoísta). Además, introduce el concepto del '*Pied Piper Effect*' (flautista de Hamelín), que se produce cuando en una disciplina hay un trabajo muy citado. Con el tiempo, se descubre que la aproximación o hipótesis de este trabajo no es correcta, pero los autores siguen citando este trabajo. Rodríguez-Narrarro y Brito (2019) critican el uso de las publicaciones y las citas para la evaluación científica, en la que se perjudica el rendimiento de algunos países (por ejemplo, con la indexación tardía de los artículos). Parra et al. (2011) también afirman que hay una falta de idoneidad de algunos indicadores para la evaluación, como es el caso del *h-index*. Sin embargo, también hay algunos autores que apoyan el uso de estos indicadores, como es el caso del estudio de Mingers y Yang (2017), quienes afirman que el *h-index* es el indicador más efectivo para evaluación. En una línea similar, Oppenheim (1995; 1997), a través de un análisis de correlaciones con los resultados obtenidos en el sistema de evaluación del Reino Unido, considera que las citas son un buen indicador.

Por otro lado, el uso del JIF constituye una de las controversias más discutidas en la literatura científica (Rushfort y De Rijcke, 2015; Lariviere et al., 2016; Williams y Morrone, 2018). Por ejemplo, Bordons, Fernández y Gómez (2002) destacan la

problemática de su uso en el contexto de países periféricos, la dificultad de las comparaciones entre campos, y que su uso puede afectar a la estrategia de publicación de los científicos. Sin embargo, estas mismas autoras destacan la utilidad de este indicador en análisis macro, meso y micro, su gran accesibilidad, en la misma línea que otros autores que destacan su simplicidad (Haustein y Larivière, 2014).

Alguno de los artículos identificados en este apartado también habla de los efectos del uso de las citas en los sistemas de evaluación (Abramo, D'Angelo y Grilli, 2021), especialmente en el sistema de evaluación italiano o en el británico. Moed (2007) señala como efectos a largo plazo de los sistemas de evaluación un aumento de las citas, un incremento de las publicaciones en revistas de impacto o un incremento de la colaboración. También hay otros estudios que se centran más en las motivaciones de los investigadores para citar (Vaughan y Shaw, 2007), así como las percepciones (negativas) de los propios investigadores sobre ellas (Aksnes y Rip, 2009).

2. Propuestas de mejora de indicadores (o desarrollo de nuevos indicadores)

Los indicadores con más propuestas de alternativas de mejora en la literatura son el JIF y el *h-index*. En el caso del primero, por ejemplo, se proponen nuevos indicadores para solucionar sus problemáticas como el *Discounted Cumulated Impact* (Javrlein y Persson, 2008) o un indicador para calcular el envejecimiento de las revistas científicas (Moed, Van Leeuwen y Reedijk, 1998). En el caso del *h-index*, se proponen alternativas para el caso de la coautoría (Crespo y Simões, 2021) o simplemente mejoras o variantes, como el *hg-index* (Alonso et al., 2009) o el *h-idx* (Prathap, 2010). En la literatura también se encuentran artículos que proponen algunos nuevos indicadores para evaluar a los investigadores como el *Research Excellence Index* (Farid, 2021) o el *Fractional Scientific Strength* (Abramo y D'Angelo, 2013). Otro grupo de artículos identificados en este apartado se centra en proponer mejoras en los sistemas nacionales de evaluación a nivel individual (Abramo y D'Angelo, 2010; Kim et al., 2014).

3. Descripción de sistemas de evaluación

La literatura científica se centra ampliamente en la descripción del sistema de evaluación de Reino Unido (*Research Assessment Exercise*, RAE; *Research*

Excellence Framework, REF) y, en menor medida, del italiano (*Valutazione della Qualità della Ricerca*, VQR). Por ejemplo, una gran mayoría de los artículos describen el funcionamiento de estos sistemas (Paisey y Paisey, 2005; Ball y Butler, 2005) o su rendimiento en una disciplina específica (Traynor y Rafferty, 1999; Masrton y Ayub, 2000; Anthony, 2005). Incluso algún trabajo destaca el cambio de criterio en las diferentes ediciones de un mismo sistema de evaluación como, por ejemplo, el italiano (Abramo, D'Angelo y Di Costa, 2011). Otros estudios realizan comparativas entre varios sistemas de evaluación (Geuna y Piolatto, 2016; Rebora y Turri, 2013).

Parte de la literatura que se agrupa en esta sección hace referencia a los efectos que ha tenido el uso de los sistemas de evaluación, como el incremento de las publicaciones, la publicación concentrada en revistas de mayor impacto o el incremento de la colaboración (Doyle y Arthurs, 1998; Checchi et al., 2020; Demetrescu, Ribichini y Schaerf, 2020; Moed, 2007). Algunos estudios, como el realizado por Karlsson (2017), destacan que, en el caso de Suecia, el sistema de evaluación ha repercutido en un incremento de la gestión de la investigación, el liderazgo, la comunicación y el buen orden administrativo.

En la literatura también se identifican las problemáticas de los sistemas de evaluación para algunos países. Vasen (2018), a través de un informe cualitativo a investigadores sociales, determina que existe una asimetría entre el discurso de las instituciones y los comportamientos que se premian en las prácticas de evaluación, siendo estas últimas las que fomentan un perfil más clásico. Otros estudios presentan el contexto de otros países como Malasia, donde uno de los grandes retos de su sistema de evaluación es la organización de la información (Yassin et al., 2011).

Se identifican asimismo críticas al uso de las citas como indicadores en sistemas de evaluación (Warner, 2000), especialmente en relación a la tensión institucional que ocasionan (Dawson, Findlay y Sparks, 2004), al sesgo hacia algunos temas de interés (Dunne y Harney, 2008; Butler y Spoelstra, 2014) o alguno de sus efectos, como las prácticas de escritura (Mcculloch, 2017), las malas conductas de publicaciones (Sheikh, 2000) o las consecuencias no deseadas, como el espíritu competitivo, adversario y punitivo a largo plazo (Elton, 2000), contrario al sentido de comunidad, propio de la cooperación científica.

Por el contrario, otros trabajos muestran los efectos beneficiosos del uso de indicadores bibliométricos en algunos sistemas de evaluación específicos, como es el caso del Reino Unido (Taylor, 2010). Otros autores van más lejos indicando que sólo deberían usarse metodologías bibliométricas, tal y como indican Abramo y D'Angelo (2011) en relación al sistema italiano (Abramo y D'Angelo, 2011). En esta misma línea, Geuna y Martín (2013) afirman, a través del análisis de varios sistemas de evaluación del mundo, que los beneficios pueden superar los costes; sin embargo, con el tiempo producen rendimientos decrecientes.

4. Comportamientos oportunistas

En la literatura encontramos un amplio número de trabajos que describen los sistemas adversos ocasionados por los sistemas de evaluación en los investigadores. Así, por ejemplo, se menciona el regalo de autoría o comportamientos oportunistas (Abramo, D'Angelo, Di Costa, 2019); el *gaming* del sistema, el juego de excelencia que “domina a sus jugadores” (Vasen, 2018; Butler y Spolestra, 2014), el incremento de publicaciones en grupos elitistas (O'Connell et al, 2020), injusticias epistémicas como el idioma de las publicaciones (Rowlands y Wright, 2020), la competición de lo global-local (Vidovich, 2008), la concentración de la producción científica en algunos departamentos (Bonaccorsi y Cicero, 2016) o las respuestas “adaptativas” de los investigadores (Akbaritabar, Bravo y Squazzoni, 2021).

5. Crítica al uso de bases de datos selectivas (*Web of Science* y *Scopus*) en la evaluación

Muchos estudios comparan el uso de diferentes bases de datos bibliográficas y los efectos que éstas pueden tener de cara a prácticas evaluativas. Por ejemplo, se analizan aspectos como la cobertura (Amara y Landry, 2012; Norris, Oppenheim, 2007), las citas (de Winter et al., 2014), el idioma de las publicaciones (Vera-Baceta, Thelwall y Kousha, 2019) o los sesgos de *WoS* y *Scopus* (Mongeon y Paul-Hus, 2015). Otros estudios comparan la cobertura de *WoS* y *Scopus* en contraposición a bases de datos nacionales, como la *Norwegian Science Index Database*, en la que se aprecia una cobertura insuficiente en ciencias sociales y humanidades (Aksnes y Siversten, 2019).

Dentro de este grupo se encuentran estudios que proponen el uso de bases de datos alternativas como *Dimensions* o *Google Scholar* para la evaluación, no sujetas a suscripción (*Dimensions* dispone de versión gratuita y de pago). En cuanto a *Dimensions*, Thelwall (2018a) afirma que constituye una buena base de datos para analizar el impacto de las publicaciones y destaca su alta correlación con *Scopus*. Herzog, Hook y Konkiel (2020) detallan igualmente las ventajas de esta base de datos (diferentes tipologías de documentos, inclusión de *altmetrics*, información de publicaciones y citas). También existen muchos estudios que destacan el rol de *Google Scholar* como una base de datos complementaria para la evaluación científica (Bornmann et. al, 2016; Mingers y Meyer, 2017; Delgado López-Cózar, Orduña-Malea y Martín-Martín, 2019). Entre sus ventajas, se destaca su mayor cobertura de publicaciones (Baneyx, 2008), la cobertura de citas (Martín-Martín et al., 2018) así como su inmediatez (de Winter et al., 2014). En esta misma línea, se destaca el uso de *Google Books* en áreas como las ciencias sociales y humanidades (Kousha y Thelwall, 2009; 2011). Finalmente, otros estudios destacan el uso de *Microsoft Academic*, subrayando su mayor cobertura en citas (Harzing y Alakangas, 2016) así como la inclusión de otras tipologías documentales como libros y conferencias, a pesar de que esta base de datos (ya desactualizada) también posee una cobertura insuficiente para las humanidades (Hug, y Brandle, 2017).

6. Uso de las métricas alternativas para la evaluación de la investigación

Algunos trabajos identificados en la revisión de la literatura apuestan por el uso de métricas alternativas para la evaluación (Ronald y Fred, 2013; Thelwall, 2018c; Mohammed y Thelwall, 2014). Algunas de estas métricas están muy correlacionadas con el impacto medido en citas, como el número de *Readers* en *Mendeley* (Bornmann, 2015), aunque otras métricas no presentan apenas correlación, como los *tweets* (Bornmann y Haunschild, 2018). Sotudeh, Ravaei y Mirzabegi (2018) recalcan como fortalezas su apertura y facilidad de acceso, la no dependencia en bases de datos comerciales, la evaluación de impacto de *preprints*, la medición de diferentes tipos de impacto de investigación, la promoción de procesos de revisión por pares, la reducción de sesgos lingüísticos, la evaluación del impacto de jóvenes investigadores, la aceleración del proceso de evaluación, el cálculo de impactos en todo tipo de audiencias, las comparaciones interdisciplinarias y el análisis de temas poco citados, entre otros asuntos. Por otro lado, Haustein (2016) también identifica retos en su uso

como su heterogeneidad (que hace difícil tener un marco común), la falta de precisión, la consistencia o replicabilidad y las dependencias particulares (depende de APIs o DOIS, proveedores de datos y agregadores).

7. La medición cuantitativa frente a la medición cualitativa (revisión por pares)

Este apartado incluye el debate sobre el uso y la combinación de metodologías cuantitativas y cualitativas en la evaluación. En este grupo encontramos estudios que afirman que el uso de indicadores bibliométricos y de procesos de revisión por pares proporcionan resultados parecidos (Bertocchi et al., 2015; Baccini y De Nicolao, 2022). Por el contrario, otros estudios afirman que los resultados son muy dispares, como es el caso del sistema de evaluación italiano (Abramo y D'Angelo, 2011). Algunos otros trabajos consideran que combinar los indicadores y el proceso de revisión por pares es una buena solución (Martin, 1996) e incluso que la combinación de los indicadores es indispensable en el proceso de evaluación a nivel de grupo de investigación, departamentos o institutos (Van Raan, 2005). Por el contrario, autores como Abramo y D'Angelo (2011) afirman que sería más eficiente (con un menor coste) el uso exclusivo de los indicadores bibliométricos para los sistemas de evaluación.

8. Manifiestos que abogan por un cambio del sistema de evaluación

En este apartado encontramos estudios que examinan Manifiestos más generales, como la *Declaración de San Francisco sobre Evaluación de la Investigación* (DORA) (2014), en la que se hacía hincapié en dejar de usar el Factor de Impacto de la revista para evaluar la investigación científica de un investigador (al poder afectar a decisiones de contratación, promoción o financiación) (O'Connor, 2022). Otro de los grandes manifiestos que incluye el *dataset* es el *Manifiesto de Leiden* (Hicks et al., 2015), cuyo documento incluye una lista de diez principios para la evaluación de la investigación. Este manifiesto es de especial importancia porque es elaborado por personas de reconocido prestigio en Bibliometría y estudios de la Ciencia, resultando unos principios de corte más técnico, respecto a DORA, que es más ideológico. Por otro lado, aparece el *Metric Tide* (Wildson, 2015), informe independiente con una revisión del papel de las métricas en la evaluación y gestión de la investigación, así como una serie de recomendaciones (uso de métricas responsables) dirigidas a los líderes universitarios,

a quienes financian la investigación, a editores e investigadores individuales (ver listado de manifiestos en Tabla 1).

En el ámbito de las ciencias sociales, encontramos el '*Yes we should; Research assessment in the humanities*' (van den Akker, 2016), en el que se aboga por un cambio del sistema de evaluación para humanidades, ya que las prácticas científicas son muy diferentes a otras áreas (por ejemplo, la tipología de publicación o el idioma de las publicaciones). En el ámbito español, Delgado-López-Cózar, Ràfols y Abadal (2021) publican la carta '*Letter: A call for a radical change in research evaluation in Spain*', en el que se pide un cambio en las políticas de evaluación actuales, basadas en un uso indebido de los indicadores bibliométricos.

9. Aspectos transversales: interdisciplinariedad, género o diversidad

Finalmente, una serie de estudios critica aspectos más transversales, como la interdisciplinariedad, el género o la diversidad. Ràfols et al. (2012) critica el hecho de que los rankings de revistas pueden poner en desventaja la investigación interdisciplinar en la evaluación de la investigación. En cuanto al género, encontramos artículos que evalúan la brecha de género en el sistema británico (Jappelli, Nappi y Torrini, 2017) o la discriminación de género, con un enfoque mayoritariamente masculino para el éxito profesional (Knights y Richards, 2003). Por su parte, Northcott y Linacre (2010) también afirman que los sistemas de evaluación pueden afectar a la diversidad y la originalidad de la investigación realizada (en su caso, analizan el campo de la contabilidad).

4. La Bibliometría en la evaluación de la investigación

Como se ha podido apreciar a través de los resultados obtenidos en la revisión de la literatura científica en evaluación de la investigación, la Bibliometría ocupa un lugar significativo en la disciplina. El uso y aplicación de distintos índices e indicadores bibliométricos se ha extendido en los distintos modelos de evaluación de la actividad científica a distintos niveles (tanto institucionales como regionales o nacionales).

Aunque el uso de estos indicadores puede aportar indicios inestimables a la hora de determinar la calidad, impacto y relevancia del trabajo realizado por una persona (u otras agregaciones como grupos, centros, institutos, organizaciones, países o revistas), su uso indiscriminado y sin conocimiento puede dañar el sistema desde la base, generando no sólo evaluaciones injustas o erróneas, sino modificaciones en la conducta del personal investigador, algo que hasta incluso en áreas como la Ética y Filosofía, paradójicamente, está sucediendo (Feenstra, Delgado López-Cózar y Pallarés-Domínguez, 2021), consecuencia directa del sistema de incentivos y recompensas existente.

Otro de los problemas detectados en los sistemas de evaluación es el propio mecanismo de evaluación. En el proceso de acreditación externo a las instituciones efectuado por distintas agencias (regionales o nacionales), éstas delegan la tarea de evaluación al propio profesorado (lo que genera un conflicto de interés, todo sea dicho de paso). Aun cuando el profesorado experto en el área tiene la capacidad para determinar muchos de los aspectos relativos a la relevancia de la carrera científica de las personas solicitantes (dado su conocimiento de la disciplina y de su estructura: fuentes y autores relevantes, líneas de investigación, eventos, etc.), esto no implica necesariamente que tengan un conocimiento suficiente para aplicar los indicadores bibliométricos contemplados en los procesos evaluativos.

Este hecho provoca que tanto profesorado como personal técnico administrativo sin formación en Bibliometría esté aplicando estos indicadores (de forma generalmente poco adecuada) en procesos evaluativos, con la responsabilidad que esto implica para las personas evaluadas. Resulta manifiestamente incomprensible que técnicos en Bibliometría

no participen en los procesos de diseño de evaluación de la investigación, ni en su uso y aplicación en el proceso evaluativo, lo que da pie en algunas ocasiones a evaluaciones ridículas y arbitrarias (Orduña-Malea, 2021).

La consecuencia de ello es una desalineación entre lo que la comunidad científica (Bibliometría, Cienciometría, Evaluación de la investigación, etc.) discute, y lo que los modelos de evaluación nacionales y autonómicos diseñan y aplican. Mientras la comunidad científica aboga por utilizar—cautelosamente—indicadores normalizados, por evitar los indicadores a nivel de revista, por ampliar las tipologías documentales analizadas y las ventanas de citación consideradas, por reconocer las diferencias entre disciplinas, por valorar el uso de fuentes alternativas y abiertas en lugar de fuentes cerradas y de pago, o incluso por valorar aspectos más allá de las citas recibidas, todo ello con el fin de determinar otros aspectos de índole más social (como la transferencia a la sociedad de los resultados, la aplicabilidad profesional, la resolución de problemas locales o regionales), los servicios gubernamentales con competencias en la gestión de la investigación y las agencias de acreditación persisten en diseñar y utilizar indicadores de revista de forma determinante (principalmente el JIF y SJR), en obviar indicadores normalizados, en focalizar los análisis bibliométricos a los artículos de revista indexados en determinadas bases de datos elitistas (que son además cerradas y de pago, ofrecidas a través de licencias contractuales carísimas), y en generar falsas diferencias entre disciplinas, modificando subjetivamente la cantidad mínima de artículos publicados en revistas indexadas por fuentes selectivas a la hora de valorar positivamente una acreditación como medida de corrección.

El objetivo de este apartado es describir los principales indicadores bibliométricos usados en procesos evaluativos, distinguiendo aquellos aplicados a nivel de fuente (sección 4.1, Tabla 3), a nivel de persona (sección 4.2, Tabla 4), a nivel de publicación (sección 4.3., Tabla 5) y, finalmente, las métricas alternativas (sección 4.4., Tabla 6). Para cada métrica se describen sus principales ventajas e inconvenientes a la hora de ser utilizadas en procesos de evaluación.

4.1. Indicadores bibliométricos a nivel de fuente

En este apartado se incluyen las comúnmente llamadas “*journal-level metrics*”. No obstante, con el fin de evitar asociar la revista como única agregación posible, se ha decidido usar el término fuente. Algunas bases de datos como *Scopus* ya utilizan el concepto “*source*” para referirse precisamente al agregado en el que se incluye una publicación.

Tabla 3.

Listado de indicadores a nivel de fuente

Métricas a nivel de fuente	Características, ventajas e inconvenientes
Factor de Impacto (JIF)	<p>Probablemente se trata del indicador más polémico y analizado, y cuyas debilidades a la hora de ser aplicado a la evaluación de individuos han sido claramente estudiadas, analizadas y discutidas por la comunidad científica (Lariviere y Sugimoto, 2019).</p> <p>Su fórmula implica la evaluación de una revista mediante un promedio de citas “esperado”. Es esperado porque se trata de un valor obtenido considerando las publicaciones de los dos años precedentes dividido por las citas recibidas por esas publicaciones durante en el año de evaluación.</p> <p>Entre sus principales inconvenientes se hallan los siguientes:</p> <ul style="list-style-type: none">• Es un indicador basado en un promedio, cuando se conoce que la distribución de las citas no es normal. Esto implica que el Factor de Impacto se obtiene principalmente por el elevado nivel de citas recibidas por unos pocos artículos de la revista, mientras que la mayoría de los artículos publicados en el marco temporal analizado probablemente no ha recibido ni siquiera una cita. Por tanto, no es representativo de la revista en su conjunto, sino de unas pocas publicaciones exitosas.• Es un indicador orientado a medir el impacto (basado en citas) de una revista y, por tanto, no de la capacidad o calidad de un autor o autora. Las evaluaciones deben realizarse preferentemente a nivel de documento, no a nivel de revista. Un artículo en una revista muy prestigiosa puede recibir 0 citas, mientras que otro artículo en una revista

menos prestigiosa puede recibir muchas citas. Es cierto que el Factor de Impacto provoca que muchos autores tiendan a citar más artículos en revistas con alto impacto (sesgo por reputación), pero no existe causalidad.

- El indicador se calcula en función de la cobertura de revistas indexadas en *Science Citation Index* y *Social Science Citation Index* de la *Web of Science* (WoS). Por tanto, está limitado a la cobertura de estas bases de datos selectivas, sesgadas hacia ciertas tipologías documentales (artículos de revista), idiomas (inglés) y disciplinas (experimentales).
- Aunque existe el producto *Arts & Humanities Citation Index*, no existe un Factor de Impacto asignado a las revistas indexadas en esta base de datos, lo que suele llevar a confusión en algunos sistemas evaluativos y en algunas personas solicitantes.
- El Factor de Impacto es fácilmente manipulable por los editores de las revistas. Considerando las tipologías documentales admitidas como documentos citantes y documentos citados, el JIF puede crecer sustancialmente. Además, las editoriales pueden llegar a acuerdos para citarse mutuamente estableciendo carteles de citación (Mongeon, Waltman y Rijcke, 2016).
- El Factor de Impacto es solamente el indicador proporcionado por *Clarivate Analytics* y utilizado en los *Journal Citation Reports* para posicionar a las revistas dentro de sus disciplinas. Existen otros indicadores a nivel de fuente basados en citas, pero no son el *Journal Impact Factor*. En ocasiones se habla de este indicador de forma inexacta cuando se refiere a otras métricas, como el *Scimago Journal Rank* o el *Citescore*, cuando son completamente diferentes. Esto genera en ocasiones confusiones importantes en muchos modelos de evaluación, en los que se equiparan estos indicadores de impacto. Un Q1 en SCI no se puede comparar con un Q1 en JCR. Primero, porque las categorías no son exactamente las mismas. Segundo, porque la cantidad de revistas por categoría no son las mismas.

-
- Precisamente, la ordenación de revistas en función del JIF genera artefactos estadísticos, por la poca diferencia (estadísticamente no significativa) entre revistas al usar tres decimales, que añaden artificialmente diferencias irrelevantes, que pretenden transmitir una cierta precisión, inexistente a nivel matemático. Además, la división de revistas en cuatro tramos, aunque comúnmente llamados cuartiles, son en realidad “cuartos” (las revistas se ordenan según el Factor de Impacto obtenido y, posteriormente, se obtienen las zonas (cuartiles) dividiendo el número total de revistas indexadas en un campo científico por cuatro. Es decir, se obtienen cuatro zonas con un número idéntico de revistas en cada una. Sin embargo, no se tiene en cuenta que el Factor de Impacto no sigue una distribución normal), generando de ese modo una información engañosa a la hora de posicionar revistas, pues dos valores prácticamente idénticos pueden posicionar a dos revistas en distintos cuartiles de forma completamente artificial.
 - Las diferencias entre las disciplinas, cuando son cuantificadas mediante citas recibidas, son enormes. Las citas tienen sentido en áreas experimentales en las que se publica mucho y se cita mucho como parte de su herencia y cultura científica. El JIF genera problemas para disciplinas sociales, formales, humanas e incluso en ciertas ingenierías, en los que los Factores de Impacto obtenidos son muy pequeños, generando diferencias estadísticas no significativas y engañosas.
 - El Factor de Impacto trabaja con una ventana de citación de 2 años, manifiestamente insuficiente para muchas disciplinas en las que se necesitan años para acumular la cantidad de citas necesaria para que se generen diferencias significativas a la hora de ordenar las revistas. Esta limitación ha sido corregida mediante el diseño de otros indicadores con ventanas de citación superiores, pero los modelos de evaluación usan generalmente el Factor de Impacto. Aun cuando *Clarivate Analytics* proporciona el JIF con una ventana de 5 años (*5 year Journal Impact Factor*), este parámetro no es generalmente utilizado.
-

-
- Actualmente, el JIF es fundamentalmente un indicador que mide la capacidad de la revista para atraer manuscritos. Es valorable que una persona publique en una revista con un alto Factor de Impacto. Podemos presuponer que el proceso de revisión ha sido más duro y exigente (en todo caso, la exigencia no debería ir ligada al Factor de Impacto, sino a la rigurosidad científica, independiente de la revista), pero este hecho no es determinante en la valoración del personal investigador, pues la autoría/coautoría de la persona puede haberse debido a muchas circunstancias, y no sólo a la capacidad de la persona evaluada.

La principal crítica no es tanto al *Journal Impact Factor* como indicador bibliométrico, sino a su utilización a la hora de valorar a una persona, en cualquier disciplina, y además que este indicador sea determinante en el resultado de la evaluación.

**Scimago Journal
Rank (SJR)**

Este indicador (Guerrero-Bote y Moya-Anegón, 2012) pretende igualmente evaluar a las revistas en función de las citas recibidas, siendo ampliamente usado en modelos de evaluación de la investigación (por ejemplo, España) como complemento al JIF. A continuación, se indican algunas consideraciones a tener en cuenta respecto a este indicador:

- No todas las citas son iguales. Es decir, se parte del axioma de que no todas las citas tienen el mismo valor. Una cita procedente de un artículo publicado en una revista prestigiosa y de alto impacto tiene más valor que una cita procedente de una revista irrelevante. Bajo esta premisa, heredera de la filosofía del *PageRank*, se construye un índice por disciplinas.
 - Al incorporar el comportamiento de las citas en diferentes disciplinas, el SJR se puede utilizar para hacer comparaciones entre revistas de diferentes disciplinas. El efecto del SJR es aplanar las diferencias entre campos, es decir, las citas en campos de alto nivel de citación (por ejemplo, neurociencia, farmacología) valen menos que una cita en campos de bajo nivel de citación (matemáticas, humanidades).
 - En su cálculo no se consideran las autocitas, lo que supone un valor añadido. Aun cuando existe el JIF sin autocitas, este parámetro apenas
-

es utilizado o aportado por las personas solicitantes, probablemente debido a que ignoran su existencia, o porque aportarlo les perjudica.

- El corpus para generar el indicador es *Scopus*, por lo que este indicador tiene igualmente un sesgo hacia ciertas disciplinas y tipologías documentales, aunque su cobertura es mayor que *Web of Science* (*Science Citation Index* y *Social Sciences Citation Index*).
- Precisamente, el hecho de disponer de una mayor cobertura de revistas significa que los cuartiles (cuartos) de revistas por disciplina son mayores (sustancialmente mayores en algunas áreas). Por este motivo, equiparar el primer cuartil de SCI/SSCI y el primer cuartil de *Scopus* no tiene sentido, como ya se ha comentado anteriormente.
- La ventana de citación es de tres años, siendo más adecuada que la ventana de dos años del JIF. Con todo, puede ser insuficiente para muchas disciplinas.
- Es un indicador cuyo cálculo resulta computacionalmente complejo, y no es completamente reproducible.

Nuevamente, el principal problema de este indicador es su uso determinante a la hora de evaluar a una persona, no el indicador en sí, que aporta información valiosa acerca del prestigio de las revistas científicas.

Journal Citation Indicator (JCI)

Se trata de un indicador reciente creado por *Clarivate Analytics* en 2021, computado retrospectivamente hasta 2017 incluido, que se calcula para todas las revistas incluidas en *Web of Science Core Collection*, incluyendo aquellas que no tienen un JCI computado.

Formalmente, el JCI es la media del valor CNCI (*Category Normalized Citation Impact*) obtenido por todos los artículos y *reviews* publicados en los tres años precedentes al año del cómputo. Por tanto, el JCI es un indicador normalizado a nivel de fuente, de forma similar al CNCI a nivel de publicación. Esto supone una ventaja a la hora de realizar comparaciones entre disciplinas.

Es un indicador muy reciente y, por tanto, todavía no existen muchos trabajos que lo hayan analizado. Los primeros análisis empíricos apuntan a que es

un indicador que correlaciona fuertemente con el JCI y el 5-Year JCI, pero que puede ser muy útil para la evaluación de revistas en artes y humanidades (Torres-Salinas et al., 2022).

CiteScore

Se trata de un indicador calculado a partir de los indicadores de *Scopus*. Entre sus principales ventajas se encuentran las siguientes:

- Cubre el año en curso.
- Posee una amplia cobertura de revistas (43.685 revistas en 2022).
- Aparte de cuartiles (cuartos), trabaja con percentiles por disciplina, que ayudan a posicionar mejor la revista dentro de su área.
- Considera los datos en abierto.
- Las tipologías documentales consideradas tanto como documentos citantes y citados son las mismas.
- Trabaja con ventanas de citación de 4 años, ampliando las consideradas tanto por el JIF como el SJR.

No obstante, presenta igualmente algunos inconvenientes:

- No está normalizado por disciplinas
- Incorpora los sesgos propios procedentes de *Scopus*
- *Elsevier* se convierte en editor y evaluador de revistas, generando un conflicto de intereses.

**Source
Normalized
Impact per paper
(SNIP)**

El SNIP (impacto normalizado por publicación) es un indicador normalizado introducido por Henk F. Moed (2010), que se calcula a través del número de citas en el año en curso a publicaciones en los últimos tres años, dividido por el número total de publicaciones en los últimos tres años. El SNIP considera artículos revisados por pares, ponencias de conferencias y revisiones.

Las citas se normalizan para corregir las diferencias en las prácticas de citación entre campos científicos. Para ello, se aplica un criterio que básicamente se basa en que cuanto más larga sea la lista de referencias de

una publicación que cita, menor será el valor de una cita que se origine en esa publicación.

Una revista con un SNIP de 1 tiene la mediana (no la media) del número de citas de las revistas de ese campo.

A continuación, se muestran algunas consideraciones a tener en cuenta respecto a este indicador:

- Es un indicador que se puede consultar en abierto.
- Se calcula a partir de los datos de *Scopus*, con las limitaciones acerca de su cobertura ya indicadas anteriormente.
- Se centra en el “potencial de citación” (opera con los documentos citantes y no con los citados).
- Se excluyen los elementos no citables.
- No tiene en cuenta las revistas con un alto porcentaje de artículos de revisión.
- El campo objeto se define como la colección de documentos que citan una revista determinada. De esta manera, el campo temático se construye sobre la base de una matriz de citas de artículos, en lugar de revistas.
- Se incluyen las autocitas en su recuento.

Su uso como complemento a la evaluación es de interés. No obstante, el SNIP plantea algunos problemas y limitaciones (Mingers, 2014), que llevaron a realizar algunas modificaciones a la propuesta original (Waltmal et al., 2013). Los problemas aparecen especialmente cuando se aplica a áreas con un *dataset* de publicaciones, en el que existan *outliers* (publicaciones con muchísimas citas en un orden de magnitud superior al resto). En España no es muy utilizado por las Agencias de evaluación de profesorado, aunque aporta datos complementarios de interés acerca del impacto de las revistas a partir del impacto de sus publicaciones.

4.2. Indicadores bibliométricos a nivel de individuo

En este apartado se incluyen aquellos indicadores aplicados directamente a la persona autora de la contribución científica. En algunas ocasiones estos indicadores se calculan directamente y, en otras ocasiones, se calculan como consecuencia de la agregación de distintos valores asociados al individuo.

Tabla 4.

Listado de indicadores a nivel de individuo

Métricas a nivel de autoría	Características, ventajas e inconvenientes
Número de publicaciones	<p>En España se utiliza frecuentemente el número de publicaciones indexadas (generalmente en JCR y <i>Scopus</i>) como criterio de evaluación del profesorado. En concreto, este indicador se usa como un <i>threshold</i> o requisito de mínimos para poder ser acreditado a las distintas figuras de profesorado existentes tanto en ANECA como en las distintas Agencias de evaluación regionales. No obstante, la utilización de esta métrica con fines de establecer umbrales es problemática por distintos motivos:</p> <ul style="list-style-type: none">• Al ser un indicador dependiente del tamaño, debe usarse con precaución, pues las prácticas (productividad) de publicación son diferentes en distintas áreas. Las Agencias tratan de evitar esto estableciendo distintos umbrales. No obstante, el establecimiento de estos umbrales no deja de ser subjetivo y, en algunos casos, injustificado metodológicamente. Además, el concepto de autoría científica está siendo debatido debido a prácticas de hiperproductividad (Cronin, 2001), especialmente debido a las tradiciones de algunas disciplinas dentro de la llamada <i>big science</i>. Por este motivo, el concepto “autor/a” se diluye en algunos casos, replanteándose incluso su definición y sentido.• La cuantificación de publicaciones se basa fundamentalmente en ciertas tipologías documentales (artículo de revista) indexadas en ciertas bases de datos. No obstante, para comprender el significado de un indicador de conteo de artículos científicos, se debería tener claro qué es y qué significa un trabajo científico (Fernández Carro,

2020). Podemos asumir que es una unidad mínima de conocimiento validado, pero esto no es necesariamente cierto en todos los casos y disciplinas.

- Esta práctica obliga a los autores a centrarse en publicar artículos de revista en lugar de publicar otras tipologías documentales, propias en algunos campos (como los *conference papers* en ingeniería, los capítulos de libro en ciencias jurídicas, o los libros en historia o sociología).
- La contabilización de artículos se realiza en base a la indexación de esas revistas en las bases de datos consideradas (básicamente JCR y *Scopus*, tal y como se ha comentado anteriormente). Esto provoca que el profesorado trate de enviar sus trabajos a un número limitado de revistas. Esto es esencialmente peligroso en ciertas áreas en las que se dispone de muy pocas revistas en las que su ámbito (*scope*) sea el adecuado. El profesorado tiende en ocasiones a adaptar sus resultados para amoldarse a las revistas indexadas en las que pueden tener más oportunidades de ser publicados, aunque no sean las revistas de su ámbito.
- Finalmente, al valorar la cantidad de publicaciones por encima de la calidad, relevancia y contribución del trabajo a la disciplina, se genera que el profesorado se centre en llegar al límite de publicaciones establecido, en algunas ocasiones de cualquier forma.

Número de citas absolutas

Del mismo modo que contabilizar la cantidad total de publicaciones conlleva unos defectos en la valoración (además de unos efectos en la conducta), la contabilización de las citas recibidas no constituye por sí misma un indicador adecuado:

- El número de citas totales recibidas está delimitado por disciplina, tipo de publicación, número de coautores, año de publicación de los trabajos y edad académica. Estos parámetros deben controlarse o normalizarse a la hora de generar comparaciones, pues puede haber diferencias significativas incluso entre distintos campos dentro de una misma disciplina.

-
- La distribución de citas de un individuo está fuertemente sesgada (por ejemplo, una pequeña proporción de todas las publicaciones recibe la mayoría de las citas). Es decir, es un número engañoso del impacto y de la trayectoria de una persona.
 - Este parámetro no tiene en cuenta la participación de la persona en las publicaciones, un aspecto que solamente puede contemplarse con una valoración narrativa (escrita u oral) de un trabajo o una trayectoria.
 - Finalmente, al igual que el umbral de publicaciones, este parámetro promueve la obtención de citas recibidas, lo que puede llevar a conductas no éticas como la autocitación excesiva, las citas entre colaboradores y las citas coercitivas (cuando un individuo revisa textos de terceros y recomienda citas de artículos en las que es autor).
 - Los hábitos de citación pueden proporcionar información deformada. Un ejemplo de ello es el llamado efecto “Flautista de Hamelin” o “*Pied Piper Effect*”, descrito por Kostoff (1997). Se refiere a cuando existe una aproximación/método altamente citado en una disciplina pero que, con el tiempo, se descubre que no es correcta. A pesar de ello, el artículo sigue siendo citado por inercia, aunque sus resultados estén desfasados, obsoletos o se haya descubierto que son directamente falsos.

H-index

El *h-index* (Hirsch, 2005) es, junto al JIF, el indicador probablemente más criticado y discutido en Bibliometría, y a la vez más usado en prácticas evaluativas, aunque sus debilidades han sido ampliamente analizadas (Costas y Bordons, 2007).

Es el indicador con más variantes existentes (Alonso et al., 2009; Bornmann et al., 2011), cuyas diferentes versiones siguen apareciendo con cierta frecuencia. No obstante, prácticamente ninguna de ellas ha tenido una fuerte implantación en los análisis métricos, y menos en las evaluaciones del profesorado.

Entre los principales inconvenientes localizados por la literatura científica, se destacan los siguientes:

-
- Es un indicador que favorece la cantidad de publicaciones. Por tanto, se ve muy afectado por la edad académica, la actividad colaborativa y la disciplina del autor/a.
 - En el caso de la edad científica, claramente el *h-index* depende de la duración de la carrera de cada científico: con el tiempo, aumentan las publicaciones y las citas (Hirsch, 2005; Kelly y Jennions, 2006). Esto genera un sesgo hacia los investigadores más *senior* (European Commission, 2017). Por tanto, no podemos comparar los *h-index* de dos personas sin considerar los años que llevan publicando. Aun cuando este factor trató de controlarse dividiendo el *h-index* por el número de años en los que la persona está activa, este parámetro es raramente solicitado, y plantea a su vez otros problemas secundarios, como por ejemplo establecer la fecha de inicio de la carrera (tras el primer artículo publicado, tras la defensa de la tesis, etc.) o cómo obtener ese dato para todas las personas solicitantes.
 - En el caso de las disciplinas científicas, existen áreas en las que se publica mucho, por lo que un *h-index* de 20 podría ser incluso normal. En otras disciplinas, un *h-index* de 20 correspondería a una persona de relevancia internacional. Estas diferencias ya se perciben en el trabajo original de Hirsch, pero la forma en la que se implementan en las distintas comisiones de disciplinas en las Agencias de evaluación no es del todo conocida. La comunidad científica trató de elaborar un *h-index* universal normalizado (Radicchi, Fortunato y Castellano, 2008). No obstante, su cómputo es complicado y es prácticamente desconocido, por lo que su uso ha sido prácticamente nulo.
 - Al ser un indicador que combina dos dimensiones (publicaciones y citas), ambos parámetros quedan condicionados uno al otro. Por tanto, la publicación de artículos altamente citados puede verse atenuada o poco recompensada por el *h-index*. Para solucionar este problema, surgió el *g-index* (Egghe, 2006), probablemente la variante del *h-index* más utilizada. Sin embargo, este parámetro, complementario al *h-index*, es raramente utilizado en procesos de evaluación.
-

-
- Dada la forma de computación del *h-index*, dos investigadores en el mismo campo con carreras científicas muy diferentes podrían obtener exactamente el mismo *h-index*. Esto provoca fuertes discrepancias y constituye el germen de muchas evaluaciones deficientes, especialmente cuando el personal evaluador toma el *h-index* sin contemplar otras variables de contexto o trayectoria.
 - El *h-index* provoca adicionalmente cambios en el comportamiento editorial del personal científico, como por ejemplo el aumento de las autocitas a artículos clave en el umbral del núcleo del *h-index* para aumentar su valor artificialmente (Van Raan, 2006). Todo ello provoca que una medida que podría potencialmente aportar mucha información de interés termina adulterándose y, por tanto, alejándose de aquello que pretende medir, la reputación científica (Koltum y Hafner, 2021).
 - Existen problemas técnicos, entre los que destaca desambiguación de nombres (Hirsch, 2005), aspecto que con la generalización del ORCID podrían minimizarse poco a poco.
 - Finalmente, el *h-index* depende de la fuente de datos utilizada para su cómputo (Bar-Illan, 2008). Dadas las diferencias de cobertura entre las fuentes, el *h-index* de un autor puede ser completamente diferente si lo estamos midiendo en bases de datos muy elitistas (*Science Citation Index* y *Social Science Citation Index*), elitistas (*Scopus*) o inclusivas (*Google Scholar*, *ResearchGate*).

Aun cuando el *h-index* es un parámetro fácil de entender y rápido de capturar en las principales bases de datos bibliográficas, su uso discriminatorio para evaluar a una persona está altamente desaconsejado, de forma prácticamente unánime por la comunidad bibliométrica. Este valor debe normalizarse, contextualizarse o complementarse con otras medidas.

i10-Index

El *i10-index* es un indicador popularizado por *Google Scholar* que lo incorpora en los perfiles de autor de esta base de datos (*Google Scholar Profiles*). El indicador contabiliza la cantidad de publicaciones que han logrado un mínimo de 10 citas. Este cálculo resuelve algunas limitaciones del *h-index* y premia la capacidad de la persona de publicar muchos artículos con un mínimo de impacto (Teixeira Da Silva, 2021). El principal problema de este indicador es

la dependencia de las áreas disciplinares (precisa de una normalización) y de la edad académica.

4.3. Indicadores bibliométricos a nivel de publicación

En este apartado se incluyen las comúnmente llamadas “*article-level metrics*”. No obstante, con el fin de no ser restrictivo a la hora de relacionar una publicación con un artículo de revista, se ha preferido describir estas métricas usando la nomenclatura “a nivel de publicación”.

Tabla 5.

Listado de indicadores a nivel de publicación (artículo)

Métricas a nivel de publicación	Características, ventajas e inconvenientes
<i>Highly-cited papers (HCP)</i>	<p>Este indicador es calculado por <i>Clarivate Analytics</i> a partir de los datos de <i>Essencial Science Indicators (ESI)</i>, y se basa en la identificación de artículos altamente citados en su disciplina, considerando el año de publicación, para eliminar sesgos por antigüedad.</p> <p>Para ello, en cada año y disciplina, se calculan las citas recibidas por todas las publicaciones correspondientes y, con esos datos, se calcula el percentil 1% de las citas. Este valor genera un número umbral. Si la publicación de un individuo alcanza este valor umbral, pasa a ser considerado como un documento altamente citado.</p> <p>Un documento altamente citado puede dejar de serlo si al año siguiente no logra alcanzar el umbral de publicación del año correspondiente (conforme pasan los años, el umbral es más exigente, pues los artículos van recopilando citas conforme pasa el tiempo).</p> <p>Este indicador es utilizado asimismo para medir el impacto de los autores. Cuando un autor o autora logra un número determinado de publicaciones altamente citadas, pasa a ser un autor altamente citado (<i>Highly Cited Researcher</i>). Este valor, a su vez, es usado para medir el impacto científico de las universidades (cuántos autores altamente citados trabajan en una universidad) en el <i>Academic Ranking of</i></p>

Field-Weighted Citation Impact (FWCI)

World Universities (ARWU), conocido como el ranking de Shanghai.

Con independencia de otros usos secundarios, este indicador a nivel de publicación permite conocer la capacidad de un individuo a la hora de publicar un trabajo de impacto en su área, controlando el año de publicación; es un elemento elitista y diferenciador en un currículum académico.

Se trata de un indicador normalizado a nivel de publicación, elaborado a partir de la base de datos *Scopus* (Colledge, 2014).

La normalización controla la categoría científica, el tipo documental y el año de publicación.

El indicador se basa en la relación (ratio) entre las citas del documento analizado y el número promedio de citas recibidas por todos los documentos similares durante un período de tres años. Cada disciplina hace una contribución igual a la métrica, lo que elimina las diferencias en el comportamiento de las citas de los investigadores.

El parámetro permite por tanto conocer si el impacto en citas de una publicación está por debajo o por encima del promedio de citas que el total de documentos del mismo tipo (por ejemplo, artículos, capítulos de libro, *conference papers*) ha recibido en el mundo.

Este indicador debe usarse cautelosamente porque:

- Su valor fluctúa durante los tres primeros años desde que el artículo se publicó (cada año su valor puede variar). A partir del cuarto año, su valor se estabiliza..
 - Este indicador se puede inflar artificialmente con publicaciones “hipercolaborativas”.
 - El indicador puede verse afectado por la influencia de artículos atípicos, especialmente si es un *dataset* pequeño, y puede ser muy elevado para tipologías
-

documentales minoritarias (con pocas citas pueden recibir un FWCI muy alto).

A pesar de estas limitaciones, es un indicador a tener en cuenta a la hora de evaluar el impacto de una publicación en particular.

Scival (Elsevier) calcula adicionalmente el FWCI a nivel de autor, a partir de un promedio del valor logrado por las publicaciones de forma individual. Este parámetro es de interés, pero no debe usarse cuando la cantidad de publicaciones es muy reducida, pues unos pocos artículos pueden distorsionar el valor final.

Category Normalized Citation Impact (CNCI)

El CNCI es un indicador equivalente al FWCI, pero calculado sobre la base de datos de *Web of Science Core Collection*. Por tanto, es una métrica a nivel de publicación normalizada (por campo, tipo documental y año de publicación)⁸. En este sentido, presenta las mismas ventajas e inconvenientes del FWCI, con la añadidura de los sesgos introducidos por la cobertura de *Web of Science*.

Por tanto, este indicador es difícilmente aplicable a áreas sociales, humanas y técnicas, donde pocos trabajos van a lograr tener valores CNCI significativos como para poder discriminar y comparar impacto.

Field Citation Ratio (FCR)

El índice de citas de campo/dominio (FCR) indica el rendimiento relativo de las citas de una publicación en comparación con artículos de similar antigüedad en su área o dominio temático. Un valor de más de 1.0-1.5 indica una cita superior al promedio.

El FCR se calcula para todas las publicaciones en *Dimensions* que tienen al menos dos años de antigüedad y se publicaron a partir del 2000.

8

<https://incites.help.clarivate.com/Content/Indicators-Handbook/ih-normalized-indicators.htm?Highlight=cnci#>

Por tanto, se trata nuevamente de un indicador normalizado, dependiente de la base de datos (en este caso *Dimensions*), y con la novedad de la exigencia de dos años de antigüedad, por lo que no puede ser usado para publicaciones antiguas. Este indicador resulta interesante nuevamente como complemento, para valorar el impacto relativo de una publicación considerando en este caso una plataforma con una cobertura no elitista, aunque con otros sesgos disciplinares (Martín-Martín, et al., 2021).

4.4. Nuevas métricas y métricas alternativas

Las métricas alternativas constituyen una de las líneas más prolíficas de trabajo en el área de la Bibliometría durante la última década, desde su surgimiento con la publicación del *Altmetrics Manifesto* (Priem et al., 2011). Su aparición viene derivada por una parte del abuso de las métricas tradicionales de revista (especialmente el JIF) a la hora de evaluar los trabajos y personas y, a la vez, como consecuencia de la aparición de numerosas plataformas *online* que eran capaces por primera vez de recoger, recopilar y computar una nueva generación de métricas orientadas a medir la interacción persona-documento en un contexto en línea a nivel masivo.

Esto suponía poder descender a nivel de documento (dejando el agregado “revista”) a la hora de evaluar el impacto o interés suscitado por un trabajo. Un nivel de granularidad que no era posible anteriormente. La interacción de individuos específicos con documentos específicos (no necesariamente artículos de revista, sino cualquier tipología documental) abría las puertas una nueva variedad de métricas, en un orden de magnitud planetario e instantáneo.

Aunque las *altmetrics* se popularizaron como métricas a nivel de artículo (*article-level metrics*), este término es restrictivo pues las métricas alternativas pueden referirse tanto a objetos digitales (cualquier tipo documental), a personas (Martín-Martín, Orduña-Malea y Delgado López-Cózar, 2018) u otros agregados, como revistas o universidades. Por tanto, cuando se habla de *Altmetrics* se corre el riesgo de estar generalizando y hablando de métricas completamente diferentes, aunque todas estén agrupadas por este término paraguas, que define su naturaleza rompedora con las métricas tradicionales bibliométricas.

El uso de las *Altmetrics* en evaluación científica ha generado una amplia y dilatada discusión en la literatura científica, tanto en su interpretación (Haustein, Bowman y Costas, 2016) como en su uso en evaluación (Sud y Thelwall, 2014; Haustein, 2016; Sugimoto et al., 2017; Thelwall, 2018c; 2020). La tabla 6 recopila las principales ventajas e inconvenientes de las *Altmetrics* en general, pero obviamente se precisa descender a nivel de métrica y proveedor de métrica para poder evaluar la idoneidad de un indicador concreto.

Tabla 6.

Altmetrics: ventajas e inconvenientes

Métricas alternativas	Características, ventajas e inconvenientes
Ventajas	<p>Las principales ventajas de las métricas alternativas son las siguientes:</p> <ul style="list-style-type: none"> ● Proporcionan evidencias alternativas o complementarias al impacto medido en citas, lo que puede suponer la identificación de un impacto o relevancia en entornos no científicos (por ejemplo, profesionales o gubernamentales). ● Son rápidas. Las métricas se generan a alta velocidad (Fang y Costas, 2020), por lo que en algunos casos puede mostrar evidencias de impacto con cierta antelación, dado que el proceso de acumulación de citas es más lento, especialmente en algunas disciplinas y fuentes. ● El volumen de datos de algunas de estas métricas es manifiestamente superior al de otras métricas bibliográficas, en algunos casos incluso en algún orden de magnitud, por lo que los análisis estadísticos pueden proporcionar mayor granularidad. ● Son métricas que pueden aportar un alto nivel de especificidad, al poder medir documentos e individuos concretos, centrar la medición en intervalos temporales muy cortos o determinar filtros específicos con relativa facilidad. ● Son métricas que aportan potencialmente una mayor diversidad, al no centrarse en artículos de revistas y en las citas recibidas por estos artículos procedentes de otros artículos. Otras tipologías (incluyendo material complementario, datos o software) y otras evidencias de impacto (comunicación y difusión de resultados fuera de la comunidad científica) pueden aflorar, ayudando a comprender mejor, en algunos casos, el impacto de un trabajo.
Inconvenientes	<p>Los principales inconvenientes de las métricas alternativas son los siguientes:</p>

-
- Existe una gran dificultad a la hora de traducir la evidencia en impacto, es decir, en entender qué significa o aporta la evidencia localizada.
 - Son métricas con falta de robustez y potencial de alterar su valor (*European Commission, 2017*).
 - Existe una captación limitada de las redes sociales en varias disciplinas y países (*European Commission, 2017*). Es decir, los datos vienen fuertemente delimitados por el uso de las plataformas que proporcionan los datos. Si en un país o comunidad no se usa una plataforma concreta, sus datos nunca podrán ser relevantes. Por ejemplo, en países como China no tiene sentido analizar *Twitter*, puesto que no es utilizada (está vetada) y tienen sus propias redes (por ej. WeChat).
 - Falta de libre acceso a los datos subyacentes (*European Commission, 2017*).
 - Existe una alta evanescencia de los datos.
 - La prevalencia de los datos es muy baja. Es decir, para algunas métricas alternativas, solamente unas pocas publicaciones alcanzarán valores significativos, mientras que un amplio porcentaje de documentos no recibirá evidencias de impacto.
 - Existe una fuerte dependencia de proveedores de servicios (*PlumX, Altmetric.com, Crossref Event Data*). Las características de estos proveedores y su estabilidad son esenciales (Ortega, 2020).
 - Son métricas altamente manipulables.
 - Son métricas que precisan ser normalizadas y “limpiadas”. Generalmente no se pueden usar sin un proceso de revisión y filtrado de datos.
-

Entre todas las métricas alternativas disponibles, la literatura científica ha mostrado las buenas prestaciones únicamente de las siguientes métricas:

Readers (Mendeley): este indicador refleja la cantidad de personas que han incluido una referencia bibliográfica en *Mendeley*, un gestor de referencias bibliográficas. Este indicador ha mostrado una correlación con el número de citas muy elevada (Thelwall, 2018b). La razón es que los autores, en el momento de redactar un artículo y buscar referencias bibliográficas, seleccionan trabajos que, con una alta probabilidad, van a acabar siendo citados en sus artículos. Por ello, el número de *Readers* logrado por una publicación es un indicador temprano de citas muy efectivo, aunque con diferencias entre áreas.

Policy citations / Clinical citations: estos indicadores, aunque bajo el paraguas de las *Altmetrics*, cuantifican el número de citas que recibe un trabajo, con la salvedad de que amplían la variedad de documentos citantes. En este caso, al contemplar las citas procedentes de informes de políticas gubernamentales o guías clínicas (en el área de medicina y salud) se puede llegar a calibrar mejor el impacto de un trabajo en entornos más allá de los circuitos científicos de publicación de trabajos. La cita de un trabajo en un informe realizado por una organización, tanto gubernamental como no gubernamental, puede indicar la utilidad de los resultados de un estudio a la hora de elaborar políticas públicas, por ejemplo. Estos datos se consideran de gran interés a la hora de evaluar un trabajo, especialmente en algunos campos (por ejemplo, salud, psicología, ciencias de los alimentos, agricultura, etc.).

Media citations: de forma similar al caso anterior, esta métrica cuantifica la mención de trabajos científicos en distintos medios de comunicación. Aun cuando esto podría suponer una debilidad de cara a la facilidad para poder manipular los datos (menciones falsas), los proveedores de este tipo de métricas operan con un listado curado de medios (tanto generales como especializados), que ayudan a filtrar y depurar los datos. La parte negativa es que existe un sesgo hacia ciertos tipos de trabajos (polémicos, temas de interés general, mediáticos) en contra de trabajos de disciplinas más pequeñas o técnicas, con menor capacidad de captar el interés del público general. Por ese motivo, son métricas complementarias para usar en determinados casos (por ejemplo, en disciplinas con bajas tasas de citas pero alta demanda profesional), y nunca como un valor determinante en ningún proceso evaluativo.

El resto de métricas alternativas, como por ejemplo el número de *tweets* acumulados, el *engagement* (e.g. *likes*, *comments*, *replies*, *shares*), el número total de visualizaciones,

descargas, etc., son manifiestamente inapropiadas a la hora de ser utilizadas en procesos evaluativos, y deberían descartarse en todas las áreas disciplinares. En algunos casos puntuales podrían ser utilizadas como evidencia complementaria, si y sólo si existe un proceso de curación de estas métricas, una comprobación de que no han sido alteradas o manipuladas, y un trabajo de contextualización y comprensión de la evidencia aportada.

5. Efectos adversos del uso de indicadores en procesos evaluativos

Como en cualquier sistema de incentivos al personal (en este caso para mejorar el desempeño docente e investigador del PDI en las universidades), la selección de indicadores y su seguimiento es crucial para analizar el impacto de la política pública. Sin duda, el sistema de evaluación a nivel individual puesto en marcha en el ecosistema universitario ha tenido efectos positivos sobre la docencia, investigación y transferencia del PDI ya que los méritos establecidos para las distintas figuras de profesorado han supuesto un estímulo para una gran mayoría de profesores e investigadores, especialmente en las etapas más tempranas de la carrera académica. Pero, como todos los sistemas de incentivos que perduran en el tiempo, empieza a manifestar signos de agotamiento en todos los sistemas universitarios a nivel mundial y, por ende, a nivel europeo y español (véase Delgado López-Cózar y Martín-Martín, 2022).

Estos síntomas de agotamiento no son exclusivos de los sistemas de evaluación al profesorado universitario; responden a cuestiones más profundas relacionadas con las respuestas de los individuos evaluados ante sistemas de incentivos y de monitorización del rendimiento. En este sentido, es ampliamente conocida la llamada *Ley de Campbell* (1976), que viene a indicar que cuanto más se usa un indicador social de naturaleza cuantitativa con el fin de tomar decisiones sociales, los procesos sociales que se pretenden monitorizar con este indicador estarán sujetos a más presiones y distorsiones. Es decir, cuando un sujeto es evaluado y conoce el indicador utilizado para ser evaluado, esto generará un cambio en su comportamiento que puede llevar a comportamientos estratégicos, contrarios a la naturaleza de la calidad investigadora, y conductas no éticas.

Este principio social tiene muchas derivadas y variaciones, como la *Ley de Goodhart* (“Cuando una medida se convierte en un objetivo, la medida deja de ser una buena medida”) o la *Crítica de Lucas* (1976), quien considera ingenuo tratar de “predecir los efectos de un cambio en política pública a partir de las relaciones observadas en datos históricos”. Es decir, que los modelos macroeconómicos no funcionan pues los agentes económicos adaptan su comportamiento para verse favorecidos por los efectos de los cambios

(Puerta y Galeano, 2005). Estos efectos llegan a su máxima con el conocido como *Efecto Cobra*, por el cual una política orientada a resolver un determinado problema termina logrando completamente lo contrario de lo que pretendía. El nombre a este efecto proviene de una política gubernamental llevada a cabo en la India, en la que se buscaba disminuir la población de cobras venenosas. Para ello, se definió una política destinada a ofrecer una recompensa económica a las personas que mataran. No obstante, las personas, viendo la recompensa económica, comenzaron a criar en cautividad cobras con el fin de matarlas y recibir el dinero. Cuando el gobierno tuvo conocimiento de este hecho, las recompensas terminaron. Sin embargo, todas las personas que estaban criando cobras de forma clandestina decidieron liberarlas, pues ya no podían obtener dinero por ellas. La consecuencia final fue un incremento en la población de cobras (Varpio et al., 2017). Efectos parecidos han ocurrido en otros lugares, de manera que este fenómeno también es conocido como el efecto de consecuencias no intencionadas (*unintended consequences*).

Una de las mayores problemáticas del uso de indicadores en procesos evaluativos es que puede incentivar comportamientos adversos e inadecuados. A medida que estos indicadores han sido aceptados e incluidos dentro de los sistemas de evaluación de la investigación de cualquier país, y juegan además un papel fundamental y determinante, los investigadores adaptan su comportamiento para ajustarse a esta medición. Esto puede provocar en algunos casos malos comportamientos, donde los resultados generados son preferentes al propio avance de la Ciencia. En ocasiones parece que el objetivo de una publicación sea obtener citas y no transmitir nuevo conocimiento, resolver un problema o discutir ideas. A modo de ejemplo, China ofrece incentivos monetarios a los autores en función de sus publicaciones en revistas con un alto Factor de Impacto (*Ministerio de Educación de la República Popular China, 2020*), generando un objetivo financiero (para los investigadores) en lugar de un objetivo científico. Estos requisitos generan un mayor número de envíos de artículos a ciertas revistas, intentando maximizar la productividad científica.

El objetivo de este apartado es describir las malas prácticas de publicación científica que se han denunciado en la literatura internacional. Entre estas prácticas inapropiadas se deben distinguir dos casos diferentes:

- Prácticas orientadas a mejorar la reputación científica de la persona autora *per se*. En estos casos puede no existir un proceso evaluativo necesariamente, pues la persona busca la reputación y el prestigio de la comunidad científica.
- Prácticas orientadas expresamente a superar un proceso evaluativo específico, como puede ser un sexenio de investigación o una acreditación. En estos casos, el prestigio científico no es buscado necesariamente.

Por otro lado, se deben distinguir las prácticas de tipo más individual o reducido, y las prácticas que provienen de un sistema organizado para ello. Este es el concepto que ampliamente se conoce como *paper mills* (Teixeira da Silva, 2021; Day, 2022).

A continuación, se describen algunas de las prácticas inapropiadas más reconocidas y detectadas a nivel internacional.

Autocitación excesiva

El proceso de autocitación no es una mala conducta en sí misma. Resulta comprensible, lógico y justo que una persona cite trabajos suyos previos. Esto sucede especialmente en áreas y campos en las que pocas personas trabajan en una línea de investigación específica. En algunos casos, el propio autor/a es la única persona con publicaciones en un tema. Por tanto, no citarse es incluso una mala práctica, pues dejas fuera del estudio trabajos pertinentes. Además, los autores pueden de esta forma construir su propio discurso científico.

El problema se plantea cuando los autores se autocitan de forma excesiva e injustificada (Szomszor, Pendlebury y Adams, 2020). En este caso, incluso se modifica y altera la redacción para incluir citas no pertinentes. En ocasiones, esta práctica se realiza de forma concreta para aumentar las citas a algunos trabajos específicos, que estratégicamente son los que pueden hacer aumentar el *h-index* de un académico.

Autoría excluida (ghost authorship)

En este caso, una persona que ha colaborado activamente en el trabajo es excluida de la autoría final del trabajo. En algunos casos la persona no es consciente pues su nombre se ha eliminado justo en el momento de envío del trabajo. En otros casos la persona es

consciente, y ha participado en el trabajo probablemente de forma forzosa o ha aceptado a cambio de otros favores.

Autoría falsa (gift/honorary authorship)

Una de las principales trampas y prácticas no éticas en la publicación científica es la falsa autoría, llamada de formas ligeramente diferentes en la literatura (Teixeira da Silva y Dobránszki, 2016).

Una forma de autoría falsa es “la autoría invitada”. Una persona que no ha participado en el estudio pero que acaba apareciendo como coautora del trabajo. El motivo de la aparición puede ser diverso. Nos podemos encontrar con casos directos de chantaje (autoría a cambio de favorecer la promoción) o favores mutuos (dos individuos acuerdan escribir un trabajo cada uno, y finalmente los dos trabajos son firmados por los dos). En otros casos se hace por ayudar a determinadas personas en sus procesos de acreditación o solicitud de sexenios y que, por algún motivo, no pueden publicar. En otros casos son políticas directas de la persona investigadora principal. En muchos grupos, centros e institutos, la persona responsable aparece como coautora de todos los trabajos publicados en el centro o en el marco de un proyecto de investigación, aunque no haya participado en el trabajo o ni siquiera lo haya leído. No obstante, con el paso del tiempo este tipo de prácticas se han ido limitando en algunos centros e institutos a través del diseño y circulación de políticas éticas de publicación, como en el caso del CSIC (*Comité de Ética del CSIC*, 2021).

En algunos de estos casos la participación ha sido económica, es decir, la persona ha aportado financiación, ha pagado los costes de publicación, ha prestado un laboratorio o software, etc. En estos casos, la persona ha facilitado la investigación. En estos casos se aboga por mencionar a estas personas en los agradecimientos. Sin embargo, aparecer en agradecimientos no es valorado en procesos evaluativos, de ahí que su aparición sea finalmente como coautores. Esta es una circunstancia ampliamente debatida: el concepto de autoría. Probablemente, la existencia de autoría técnica (personal técnico) y una autoría logística (recursos humanos, económicos, tecnológicos) debieran contar en procesos evaluativos, de forma que no adulterasen el concepto de autoría intelectual.

En otros casos existe directamente un mercado de la coautoría, donde empresas se dedican a ofrecer “puestos” de coautoría en artículos ya realizados y aceptados⁹, parte de las actividades de lo que se conoce como *paper mills*. Por ese motivo, muchas revistas exigen un informe justificado a los autores cuando existe algún cambio en la autoría de los trabajos aceptados (renuncia o adición de autores).

Autoría no solicitada

Este supone un caso especialmente interesante. Los autores de un trabajo incluyen como coautora a otra persona (generalmente muy reconocida en su campo) sin que ésta última sea consciente de ello.

El motivo es que esta persona puede ser una facilitadora a la hora de que el trabajo sea aceptado. El equipo editorial puede ser más laxo al ver a una persona reconocida como coautora, confiar en la calidad del trabajo o simplemente pensar que ese artículo le va a reportar un mayor número de citas. Para evitar este problema, muchas revistas envían un email a todos los coautores para que confirmen la autoría del trabajo recién enviado. De esta forma, si el autor percibe que alguien está usando su nombre, puede ponerse en contacto con la revista y denunciar el hecho. Por lo tanto, en la actualidad, se está frenando este tipo de prácticas inapropiadas desde las propias revistas. Con todo, en muchas ocasiones la persona autora no se va a dar cuenta de lo ocurrido (el email puede llegar a su carpeta de *spam*) o se puede aprovechar de la circunstancia.

Cártel de citas (citation cartel)

A nivel informal, los autores comienzan a crear sus redes de contactos para intercambiar citas. De esta forma, cuando un autor A escribe un nuevo trabajo, trata de citar al autor B, y viceversa. Si las citas son puntuales y pertinentes no pareciera haber un excesivo problema. No obstante, cuando las citas son masivas y no necesariamente pertinentes, nos encontramos ante un caso de engaño, conocido como “cártel de citas” (Fister Jr.; Fister y Perc, 2016).

Esta acción no solamente es llevada a cabo por los individuos, sino que se extiende a las propias revistas, que tratan de ese modo de inflar artificialmente su Factor de Impacto a al

⁹ <https://francis.naukas.com/2019/07/25/el-mercado-negro-de-la-coautoría-de-artículos-científicos/>

recibir citas externas de otras revistas. Aunque las bases de datos elitistas (JCR, *Scopus*) detectan semiautomáticamente el comportamiento anómalo entre revistas, y puede expulsarlas del sistema (cada año se expulsan revistas por autocitación excesiva o redes de citas artificiales), si los artículos particulares no han sido retractados, las citas quedan para sus autores.

Citas coercitivas (coercitive citation)

Se trata de una práctica de chantaje que consiste en que la persona editora/revisora de una revista científica solicita a un autor que agregue citas de forma artificial a un artículo (de la revista o de la propia persona que actúa de revisora) antes de que la revista acepte su publicación. Los autores se ven presionados a incorporar la cita para evitar que su trabajo sea rechazado (Lynch, 2012).

En el caso del personal editorial de una revista, esta circunstancia se realiza con el propósito de aumentar el Factor de Impacto de la revista de forma artificial, por lo que las citas sugeridas son a otros trabajos de la propia revista. En algunas ocasiones, estas citas pueden haberse incrustado sin que la propia persona autora sea consciente. Una vez los trabajos ya han sido aceptados, y durante el proceso de maquetación y publicación, se pueden introducir citas por lo que la persona autora solamente puede darse cuenta en el momento en que el trabajo ya está publicado, aunque es probable que ni siquiera se dé cuenta, pues los autores, por norma general, no revisan su sección de referencias una vez se ha publicado el trabajo, al existir una confianza en la revista.

En ocasiones se argumenta que las recomendaciones de la persona editora pueden estar justificadas en cierta medida. Por ejemplo, si un artículo se envía a una revista muy específica y el artículo no cita a ningún trabajo previo de esa revista, puede surgir la siguiente pregunta: ¿por qué se envía el trabajo a esta revista?, ¿es la más apropiada?, y si la revista sí ha tratado previamente el tema, ¿por qué no se citan esos trabajos previos? Como se puede apreciar, es un tema delicado y con muchas aristas y matices.

En el caso de los revisores ciegos, pueden sugerir, bajo el anonimato, la cita a trabajos suyos. Estas sugerencias pueden ser pertinentes o no pertinentes. Si la cita realmente es importante y ha sido omitida, es deber de la persona que realiza la revisión indicar no sólo su inclusión sino su discusión, si procede (como con cualquier otro trabajo pertinente

publicado por otra persona). No obstante, en ocasiones, la sugerencia puede ser poco pertinente o incluso nada pertinente.

Plagio (plagiarism)

Esta práctica supone la apropiación indebida de contenidos o ideas de otras personas sin el debido reconocimiento de autoría, socavando el sistema de recompensas en la Ciencia (Resnik, 1998). Por tanto, supone la usurpación de la propiedad intelectual, y es una actividad punible.

Estas prácticas se realizan con el propósito de mostrar ideas o resultados como propios para llevarse el mérito. Aunque los resultados plagiados puedan estar ya publicados (en trabajos antiguos o incluso en otro idioma), a ojos de los nuevos lectores las ideas aparecerán como originales y podrían recibir sus citas correspondientes.

Esta práctica no debe confundirse con el reciclado de contenido (llamado erróneamente autoplagio). Una persona puede reutilizar contenido suyo previamente utilizado en otros trabajos. Esto suele suceder en la descripción de una metodología o procedimiento seguido para obtener unos datos, que ha podido ser el mismo para dos estudios, y el texto es el mismo o muy parecido. En este caso no se puede hablar de mala *praxis*.

Muchas personas lo que realizan es un artículo metodológico (*protocol paper, research protocol, methods papers*), en el que se describen los procedimientos o protocolos a seguir ante un determinado análisis. De esta forma, cuando ese método se aplica en trabajos específicos, no sólo se aligera la sección metodológica, sino que se genera una cita al trabajo metodológico. Esta es una buena práctica (de hecho, el trabajo metodológico podrá ser muy útil a otras personas trabajando en el mismo área o interesadas en aprender ciertos protocolos), siempre y cuando estos trabajos sean debidamente revisados como cualquier otro trabajo de investigación.

Publicación comprada

En esta ocasión, la persona autora envía su trabajo a una revista o fuente en la que no va a existir un proceso de revisión por pares real. La publicación puede tener costes de publicación (en el caso de revistas depredadoras) o bien carecer de los mismos (en el caso de amiguismo o intercambio de favores).

Publicación duplicada (duplicate publication)

La publicación duplicada sucede cuando una persona envía un mismo trabajo a distintas revistas o fuentes para su publicación (Norman y Griffiths, 2008). Esta mala práctica puede realizarse de forma simultánea (la persona envía el trabajo a una segunda revista antes de que la primera haya enviado el dictamen de revisión) o en diferido (cuando el primer trabajo ya está aceptado o publicado y se envía a una segunda fuente). En ocasiones los autores envían exactamente el mismo trabajo, con alguna pequeña variación, o lo traducen a otro idioma (sin informar de ello a la revista). Por tanto, esta práctica atenta contra la ética científica.

Publicación inventada (data fabrication and falsification)

En este caso no se trata de un plagio, sino de la invención de datos. Dentro de esta categoría entra la falsificación y la invención de datos (Resnik, 2014). En estos casos, los autores simplemente pueden inventarse datos para que un análisis estadístico cuadre con las hipótesis, manipular imágenes, inventarse referencias bibliográficas que apoyan sus estudios o mentir sobre la procedencia de los datos, entre muchas otras casuísticas.

Publicación Salami (salami slicing)

Esta práctica consiste en dividir una investigación en muchas partes con entidad propia para ser publicadas (concepto de Unidad Mínima Publicable, UMP), con el fin de tener un mayor número de artículos publicados y de citas (Jackson et al., 2014, Tolsgaard et al., 2019). En muchas ocasiones, el trabajo final queda completamente desfigurado y el grado de solapamiento de los distintos trabajos puede ser muy elevado (alto nivel de reciclaje de contenidos).

Esta práctica viene agudizada por los criterios de extensión que las revistas publican según la tipología de artículo. Estos límites eran comprensibles en la era papel, puesto que las revistas no podían sobrepasar un número total de páginas, pues no sólo encarecía la revista, sino que la hacía más difícilmente manejable.

Hoy día, las revistas persisten en estipular extensiones a los trabajos, aunque éstos se publiquen solamente en PDF o HTML. El motivo viene derivado de la creencia de que los trabajos breves se leen y se citan más que los trabajos excesivamente largos, debido a la

economía de la atención en un entorno digital. Sin embargo, aun cuando esto es cierto para la lectura de textos en blogs y plataformas de redes sociales, no parece que ocurra lo mismo con las publicaciones científicas, donde la extensión de los trabajos puede no estar relacionada con la citación, o incluso existir una relación positiva en algunas áreas (Haustain, Costas y Larivière, 2015; Fox, Paine y Sauterey, 2016).

Este sistema beneficia a todas las partes: la persona autora publica más trabajos; los distintos trabajos van a incluir prácticamente las mismas citas, por lo tanto, los trabajos citados reciben más citas; el autor aprovecha y cita su primer trabajo en el segundo trabajo; y la editorial publica dos trabajos en lugar de uno (cobrando dos APCs en lugar de uno).

Revisión falsa (fake reviewer)

Algunas revistas suelen solicitar a las personas que envían un nuevo manuscrito a revisión la recomendación de potenciales revisores, especialmente si la revista tiene una plantilla de revisores escasa o el tema del trabajo enviado es de interés, pero no es un tema vehicular en la revista. Ante esta situación, la persona responsable de la correspondencia crea un nombre y cuenta de email falsas, y posteriormente recomienda a esa persona inventada como *reviewer* de un trabajo en una revista. Cuando las revistas no verifican la existencia e idoneidad de estos revisores, se pueden encontrar casos en los que una persona evalúa su propio trabajo.

Revisión indulgente (peer-review ring)

En estos casos, un conjunto de personas acuerda ser indulgente entre ellos cuando reciban solicitudes de revisión. Aunque el proceso de revisión sea supuestamente ciego, los autores hablan entre ellos y conocen las autorías de cada trabajo. De esta forma se consigue lograr la aceptación de todos los trabajos, especialmente cuando desde la editorial (*editor-in-chief*) no realiza una revisión final del trabajo, o incluso forma parte del *peer-review ring*.

En muchas ocasiones, los anillos de *peer-review* se detectan porque estas revisiones se realizan prácticamente de forma automatizada a través de plantillas de revisión. Es decir, el informe de revisión es un mensaje genérico que se envía como respuesta a todas las solicitudes de revisión que la persona revisora falsa recibe.

Revisor falso

Este fraude es similar al de la revisión falsa. En este caso, se envía un manuscrito en el que un autor o coautor es falso. De esta forma, si no se comprueba su identidad, la revista podrá incorporar a esta persona como un potencial revisor en el futuro. De esta manera se consigue introducir un caballo de Troya en la revista. Cuantos más caballos de Troya existan (todos dirigidos a una misma persona o grupo de personas), más posibilidades existen de controlar los procesos de revisión en esa revista. En muchas ocasiones, el *fake author* es creado utilizando cuentas de ORCID falsas para aparentar veracidad (Teixeira da Silva, 2021).

No obstante, se han enumerado y descrito algunos de los efectos adversos que se generan como consecuencia de los usos de indicadores en procesos evaluativos de personas. Sin embargo, estos efectos perniciosos que se describen no son los predominantes entre las personas y científicos, y gracias a ello la generación y transmisión del conocimiento y la ciencia a través de las publicaciones es una realidad palpable como demuestra la bibliometría y la Evaluación de la Ciencia.

6. Conclusiones

6.1. Posibilidades y límites de la evaluación

Estamos en el momento oportuno de reinventar la evaluación de la investigación en el *Espacio Europeo de Educación Superior*, en el sistema universitario español y en el sistema universitario canario. No obstante, como hemos visto a lo largo del informe, no existe una solución única y sencilla. Parece que se deduce del repaso de la literatura internacional y de las iniciativas puestas en marcha a nivel supranacional y nacional que existe una corriente hacia una evaluación multidimensional y con criterios e indicadores más cualitativos e inclusivos, que reconozcan una amplia variedad de resultados (no sólo publicaciones en JCR), así como tareas de investigación desarrolladas a lo largo de la carrera científica.

En definitiva, estamos inmersos en un momento de optimismo acerca de la necesidad de un cambio estimulante en la evaluación de la investigación en España, en Europa y a nivel mundial. Es necesaria una profunda reflexión sobre la estrategia evaluadora como elemento central de la calidad docente e investigadora, y es una responsabilidad de la comunidad científica y universitaria internacional. Ahora bien, existen grandes dificultades para alcanzar un consenso a nivel de país y de región. Sin duda, la iniciativa de la *Comisión Europea*, que actúa como facilitador, constituye una llamada a implementar un cambio real y rápido en todos los países europeos, y modificar la estructura está en manos de los que toman decisiones a nivel de gobiernos (del país) y también a nivel de gobierno universitario.

Parece razonable buscar un equilibrio o balance entre la gobernanza de la evaluación de la investigación frente a la acreditación de la evaluación de la investigación. Es necesaria una reflexión previa para determinar la estrategia y la misión de la investigación en los momentos actuales de cambio, así como reflexionar y repensar conceptos como la excelencia científica. Europa marcará el camino, pero no tiene la competencia en política de educación superior y, por tanto, será el gobierno español y los gobiernos autonómicos en el ejercicio de sus competencias los que dictarán reglas o aclaraciones al respecto.

Entre las posibilidades de un nuevo modelo de evaluación se destaca el reto de potenciar la diversidad en dos sentidos. Un primer sentido hace referencia a la necesidad de mayor diversidad en los criterios y métodos de evaluación (cuantitativos y cualitativos), respetando la autonomía universitaria y la responsabilidad última de los equipos de gobierno de los

países y de las universidades. Un segundo sentido es que si se quieren potenciar perfiles diferentes en las instituciones universitarias “multimisión”, la evaluación del PDI se debe basar en un modelo holístico e integral, flexible y más justo y que permita diferentes elecciones (más docencia, más investigación, más transferencia) y especialización al docente e investigador a lo largo de toda su trayectoria académica.

Como se ha visto, con el paso del tiempo, se ha ido alcanzando cierto consenso en la necesidad de corregir el uso “abusivo” del uso de indicadores para la evaluación individual de la investigación, de manera que tanto evaluadores como especialistas en Bibliometría son conscientes de que usar únicamente indicadores cuantitativos para la evaluación individual consagra ventajas logísticas, pero también desventajas cuando se trata de la evaluación de personal investigador como se ha visto. Las corrientes más actuales de pensamiento que se han estudiado en la literatura, y a la que nos adherimos como equipo del proyecto/Informe encargado por la ACCUEE, es que los indicadores como el JIF, el *h-index* o los índices normalizados, ayudan a tomar la decisión, pero no deben ser instrumentos determinantes.

Además, en la evaluación de la investigación existen importantes diferencias por áreas que se pueden asumir en una evaluación de la investigación holística siempre que existan Comités Técnicos por disciplina en cada proceso de evaluación o Comités generales por área que se complementen con uno o dos evaluadores técnicos expertos en su campo para limitar algún tipo de injusticias que puedan aflorar durante el proceso de evaluación. Esto resultará aún más relevante en un contexto de implementación de la ciencia abierta, con énfasis en la conexión de datos con la sociedad y la integridad (LERU, 2021).

A este respecto han sido muchas las voces contrarias a un uso abusivo de la Bibliometría, como se ha visto detalladamente. Si la acreditación se basa desde las etapas más tempranas principalmente en el *output* de las publicaciones, esto puede conducir a un sesgo en favor de algunos perfiles de investigadores y a una falta de reconocimiento de otras contribuciones, y lo que es más importante, puede conducir a la frustración para candidatos que tienen un perfil diferente o que han hecho elecciones distintas en sus carreras. Los jóvenes académicos pueden sentirse mal juzgados y agraviados porque sus fortalezas no consiguen el mismo peso en la evaluación que las ratios de publicaciones y el desempeño de la actividad investigadora, minusvalorándose la transferencia de conocimiento, en general. Por ejemplo, es importante reconocer la divulgación científica que, en muchas ocasiones, consume tiempo de redactar publicaciones o propuestas de proyectos para

conseguir financiación, así como compartir datos, métodos o código durante todo el ciclo de la investigación. Igualmente, las actividades de tutorización, compartir materiales de docencia innovadores, establecer redes y dinamizarlas, aportar consejo a los *policy makers* son contribuciones muy relevantes que deben tenerse en cuenta.

Aunque algunas de las recomendaciones realizadas desde la literatura y experiencia internacional de evaluación de jóvenes investigadores no son aplicables directamente a todos los contextos regionales (incluyendo el SUC), no es descabellado pensar que sentar las bases de máximos de un buen proceso de acreditación (de cara a las nuevas figuras que puedan aparecer en la LOSU) puede ayudar a establecer buenos hábitos y asegurar el éxito en acreditaciones futuras a nuevas figuras, además de ampliar las expectativas que tienen los *stakeholders* (los propios equipos de gobierno de las universidades, equipos decanales, departamentos, grupos y centros de investigación) respecto a las acreditaciones de la ACCUEE.

Por tanto, la propuesta que se formule tendrá que ser posibilista e integral, teniendo en cuenta no solo el *output* investigador sino todas las contribuciones de la vida académica (formación, docencia, investigación y transferencia de conocimiento). Como señala la *LERU* (2022), si las contribuciones no son debidamente reflejadas en los criterios de producción, no serán reconocidas y, por tanto, serán infravaloradas. Algunas exclusiones pueden conducir a la creencia de que ciertas actividades son normales y, por tanto, se espera que el investigador las realice (como, por ejemplo, la contribución a la evaluación por pares de los investigadores). En la literatura, a este fenómeno se le denomina trabajo “invisible”, y en estudios recientes se examina poniendo el énfasis en los temas de género, raza o clase social (Braun, 2017; Cooper, 2021).

6.2. Observaciones finales

A continuación se exponen las principales conclusiones obtenidas por este informe:

- La evaluación de la calidad y el impacto de los investigadores tiene consecuencias cruciales para el reclutamiento, promoción y reconocimiento a lo largo de su carrera académica. Sin embargo, el sistema actual de “*publish or perish*” ampliamente aceptado en la mayoría de los países supone una barrera en la evaluación de la excelencia en la cultura investigadora (así como en el progreso de la ciencia abierta). Como resultado, múltiples iniciativas (tanto *bottom-up* a nivel de investigadores y *top-down* a través de organismos como la *Comisión Europea*) han surgido pidiendo

una reforma, en la que se quiere avanzar hacia nuevas formas de evaluación de la excelencia más cualitativas, inclusivas, y que consideren otras tareas académicas básicas, incluida la docencia y la transferencia social.

- A través del análisis de las diferentes iniciativas y manifiestos a lo largo de los últimos diez años se denota el creciente interés y la necesidad urgente de este cambio. Sin embargo, la mayoría de estas declaraciones proponen recomendaciones (por ejemplo, de indicadores) o principios a nivel macro, siendo un reto muy complejo la incorporación a nivel micro (por ejemplo, en un sistema de evaluación específico). En este sentido, faltan ejemplos de acciones prácticas y concretas (por ejemplo, un piloto en una agencia de evaluación) que se puedan reproducir en otros contextos posteriormente.
- A través de la revisión de la literatura científica, se observa el creciente interés sobre este tema, así como diferentes líneas de debate que este tema ha suscitado. En este sentido, este interés versa sobre los siguientes temas: manifiestos, declaraciones y artículos que abogan por un necesario cambio en el sistema de evaluación; análisis de los diferentes indicadores utilizados en los sistema de evaluación (por ejemplo, críticas al JIF) así como desarrollo de nuevos indicadores para vencer sus debilidades; las diferencias y el desempeño en los diferentes sistemas de evaluación (por ejemplo, en los sistemas de Reino Unido y Italia); crítica a las bases de datos utilizadas para la evaluación (principalmente, *Web of Science* y *Scopus*) y sus efectos en algunas disciplinas científicas (por ejemplo, ciencias sociales y humanidades); discusión del uso de métricas alternativas; comportamientos oportunistas (por ejemplo, de investigadores adaptándose a los diferentes sistemas de evaluación); la discusión entre la medición cuantitativa (bibliométrica) y cualitativa (*peer-review*), así como una conjugación de ambas; aspectos transversales (como la ciencia abierta, interdisciplinariedad o género). La revisión denota la amplia variedad de debate que se ha generado a lo largo del tiempo por parte de la comunidad científica.
- A través de la revisión de los indicadores bibliométricos en la evaluación de la investigación a diferentes niveles (a nivel de fuente, persona y nivel de publicación) así como el uso de métricas alternativas, se describen las diferentes ventajas e inconvenientes para ser usados en la evaluación. Esta revisión demuestra que todos los indicadores bibliométricos deben analizarse y usarse con cautela, ya que

presentan una serie de limitaciones que, como se ha visto, deben ser consideradas para ser usados en prácticas de evaluación.

- El actual sistema de evaluación ha comportado una serie de malas praxis y comportamientos adversos por parte del personal de investigación, que afectan tanto a nivel de producción (por ejemplo, *salami slicing* o unidad mínima publicable de la investigación-UMP) como de impacto (autocitas), que pueden tener un recorrido puntual e individual, pero también organizado y de alcance masivo, que socava el proceso de comunicación científica en su raíz.
- El gran reto del cambio de la evaluación no sólo repercutirá en una reformulación de los indicadores, sino que también requerirá recursos, inversiones y compromiso para ser llevado adelante y adoptado por los sistemas de evaluación (Agencias de Evaluación, profesionales de la evaluación, etc.). Esto implica una reformulación de lo que es la excelencia de investigación y lo que conlleva el impacto de la investigación tanto el valor tangible como el intangible para nuestra sociedad actual.

6.3. Aspectos a considerar

Una vez revisadas las limitaciones de algunos indicadores y los efectos adversos de la evaluación de la investigación, y en vista de los resultados obtenidos y reflejados a lo largo del informe, a la hora de elaborar una propuesta de modelo de evaluación se deberán tener en cuenta los siguientes aspectos, tal y como se abordará en el segundo informe (Orduña-Malea, Bautista-Puig y Perez-Esparrells, 2022):

Aspectos relacionados con los baremos cerrados

Para evitar este problema, los criterios deben ser discutidos para cada candidato con el fin de contextualizar los indicadores (Rafols, 2019) y hacer más flexible una comparación de los diferentes perfiles. Para ello, se puede hacer una distinción, por un lado, entre los requisitos mínimos y los extras que pueden recibir reconocimiento sobre el desarrollo y desempeño de sus actividades y, por otro, su contexto, así como los valores derivados de su uso.

Aspectos relacionados con el trabajo compartido de los asesores y de los investigadores evaluados

La apertura de la perspectiva de evaluación supone un reto tanto para los evaluadores como para las personas evaluadas. Para evitar la carga de trabajo adicional que supone la evaluación integral del candidato se puede proponer la comunicación entre evaluadores y evaluados especialmente en la evaluación cualitativa y la trayectoria (parte narrativa).

Aspectos relacionados con el estrés del investigador que sigue con filosofía del checklist

Existe el riesgo de que se lance el mensaje de tener que ser excelente en todos los aspectos del trabajo académico, lo que puede generar estrés. Por ello, y para evitar este problema, resulta de gran relevancia transmitir a las personas evaluadas que no deben cumplir con todos los ítems y que solo necesitan cubrir unos requisitos mínimos y unos umbrales.

Aspectos relacionados con las alegaciones

Otra de las limitaciones es que como los pesos relativos de estos criterios proliferantes no están claros, lo cual es necesario si se quiere conseguir una mayor diversidad y reconocimiento de las contribuciones, puede surgir la impresión de que no hay consistencia en el juicio, lo que provoca acusaciones de injusticia y favoritismo. En las Universidades de la *LERU* por investigaciones y experiencias se conoce que estas incertidumbres ofrecen condiciones favorables por parcialidad en los juicios (*LERU, 2022*). Para evitar este problema lo más importante es fijar unos criterios de acuerdo con la comunidad universitaria canaria.

Aspectos relacionados con el coste en recursos humanos

Otra de las limitaciones es que, si se quiere conseguir una evaluación integral basada en instrumentos de evaluación heterogéneos y contextualizados, dicha evaluación tiene un gran coste en términos de recursos humanos. Por ello, la propuesta formulada intentará un equilibrio entre la evaluación cuantitativa y cualitativa (su combinación y que ninguna sea determinante hace que se minimicen los problemas de cada una) y que muchos de los requisitos de la parte de investigación estén automatizados, además de procesos de formación de los evaluadores.

En el caso de ACCUEE, un aspecto clave de la evaluación es que las actividades y la información recopilada estén alineadas con los objetivos del tipo de acreditación al que se accede actualmente (Ayudante Doctor, Contratado Doctor y Contratado Doctor Investigador) y que dicha evaluación esté en línea con los criterios y normas que exigen las universidades canarias para que se alineen las estrategias de evaluación de la actividad

docente e investigadora de los jóvenes docentes e investigadores que se quieren incorporar al SUC. El ideal es llegar a diseñar enfoques de evaluación más adecuados (de máximos) con una dirección común por parte de todas las partes implicadas para poder alcanzarlo (ACCUEE, *Dirección General de Universidades*, universidades canarias).

Existen buenas iniciativas que ya están en marcha a nivel de país y a nivel de universidades europeas. En los Países Bajos y, en concreto, en la *Universidad de Leiden*, el programa nacional de Calificación Docente Superior (*Senior Teaching Qualification*) y la vía de promoción del profesor asociado (profesor titular) demuestran un mayor reconocimiento de la docencia, al igual que varias iniciativas a nivel de centros o facultades dentro de la *Universidad de Leiden*, donde ponen mucho énfasis en que deben tenerse en cuenta las diferencias entre disciplinas e institutos. Además, en la *Universidad de Leiden* se hace hincapié en incardinar todo el proceso de evaluación de la investigación con el plan estratégico de la propia universidad.

Por otra parte, la evaluación y sus evaluadores deben estar preparados para virar sus acciones desde los indicadores bibliométricos más al uso hasta prestar más atención a aquellos logros conseguidos en el ecosistema de investigación del joven investigador. Resulta que existe un trabajo invisible o intangible que contribuye a los logros individuales y colectivos del investigador senior que debe tenerse en cuenta cuando se evalúa al investigador junior. Dicha contribución al trabajo de equipo (*team-research*) está perdida o diluida en el proceso de evaluación individual de los jóvenes investigadores (LERU, 2022). Estaríamos hablando, en este caso, de una evaluación formativa tan importante o más que la evaluación sumativa que es la que se tiene en cuenta en la actualidad.

También podrá ser un hito muy importante y distintivo del SUC prestar atención no sólo a los temas relativos a género, responsabilidades familiares, o problemas de salud, sino también a los profesores maduros (o con experiencia como profesores asociados), a los profesores inmigrantes, exiliados y buscadores de asilo y a los profesores con discapacidad. Siguiendo la experiencia internacional de algunas universidades se pueden nombrar en los Comités observadores que atiendan toda esta casuística como los “guardianes de género” en la *Universidad de Lund* (Suecia) o como los “delegados de igualdad” en la *Universidad de Génova* (Italia), que tengan en cuenta el impacto de este tipo de factores en los procesos de selección. Dichos observadores pueden formar parte de los paneles a la hora de tomar las decisiones o simplemente estar de observadores o comisarios para que se eviten todo este tipo de sesgos en los procesos de evaluación. Debe

tenerse en cuenta de alguna manera los sesgos que existen en las publicaciones científicas de las mujeres en cuanto a la hora de aparecer como primeras firmantes o autoras de correspondencia, participar en colaboraciones internacionales o con empresas.

Para terminar, este informe ofrece algunas buenas ideas, recomendaciones y sugerencias al hilo de las evidencias mostradas sobre que está ocurriendo a nivel internacional en la práctica de la evaluación de la investigación y cómo efectuar un cambio cultural muy necesario en la evaluación de los investigadores que está puesto en marcha en las mejores universidades europeas (LERU, YERUN, etc.) así como en los ecosistemas de investigación más innovadores de Europa.

En suma, la necesidad de una reforma del sistema de evaluación de la investigación a nivel individual que ha sido el objeto de este informe debe ser lo suficientemente flexible como para adaptarse a la multitud de disciplinas, culturas de investigación, niveles de madurez de la investigación, misiones específicas de las instituciones y carreras académicas.

Partiendo de la evidencia científica y la identificación de los problemas y las posibles soluciones llevadas a cabo en otros entornos, el siguiente paso debería consistir en sentar las bases de un modelo de máximos, abierto e integral que pueda servir de guía para la evaluación del profesorado en el ecosistema investigador canario (SUC). Obviamente, no es algo que simplemente se formule, negocie e implemente con celeridad en una Comunidad Autónoma ultraperiférica como Canarias, sino que comienza con la toma de conciencia por parte de la ACCUEE de cómo se pueden sentar las bases para hacer la evaluación pública de la investigación de manera diferente en un futuro que es el objeto de este informe y del siguiente informe (Orduña-Malea, Bautista-Puig y Perez-Esparrells, 2022). También será muy interesante aplicar el modelo experimentalmente a un grupo reducido de PDI.

Finalmente, habrá un coste económico asociado a implantar un nuevo modelo, pero esa inversión debe venir del Estado en una gran medida. Si realmente se quiere potenciar la ciencia en España, no se puede pretender evaluar bien, rápido y barato por las Agencias de Evaluación con competencias en esta materia que existen en nuestro país. Parece condición necesaria financiar la infraestructura técnica que apoye al modelo de evaluación integral, pero no suficiente si no se considera un elemento estratégico y clave de la política pública de ciencia y educación superior de nuestro país.

7. Referencias

- Abramo, G., & D'Angelo, C. A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics*, 87(3), 499-514. <https://doi.org/10.1007/s11192-011-0352-7>
- Abramo, G., & D'Angelo, C. A. (2013). Rethinking research evaluation indicators and methods from an economic perspective: the FSS indicator as a proxy of productivity. In *14th International Society of Scientometrics and Informetrics Conference—(ISSI–2013)*.
- Abramo, G., Cicero, T., & D'Angelo, C. A. (2012). Revisiting the scaling of citations for research assessment. *Journal of Informetrics*, 6(4), 470-479. <https://doi.org/10.1016/j.joi.2012.03.005>
- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2019). When research assessment exercises leave room for opportunistic behavior by the subjects under evaluation. *Journal of Informetrics*, 13(3), 830-840. <https://doi.org/10.1016/j.joi.2019.07.006>
- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: the effects of changing the rules of the game during the game. *Scientometrics*, 88(1), 229-238. <https://doi.org/10.1007/s11192-011-0373-2>
- Abramo, G., D'Angelo, C. A., & Grilli, L. (2021). The effects of citation-based research evaluation schemes on self-citation behavior. *Journal of Informetrics*, 15(4), 101204. <https://doi.org/10.1016/j.joi.2021.101204>
- Akbaritabar, A., Bravo, G., & Squazzoni, F. (2021). The impact of a national research assessment on the publications of sociologists in Italy. *Science and Public Policy*, 48(5), 662-678. <https://doi.org/10.1093/scipol/scab013>
- Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus and Web of Science. *Journal of Data and Information Science*, 4(1), 1-21. <https://doi.org/10.2478/jdis-2019-0001>
- Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, 38(6), 895-905. <https://doi.org/10.1016/j.respol.2009.02.001>
- Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of informetrics*, 3(4), 273-289. <https://doi.org/10.1016/j.joi.2009.04.001>
- Alonso, S., Cabrerizo, F., Herrera-Viedma, E., & Herrera, F. (2010). hg-index: A new index to characterize the scientific output of researchers based on the h-and g-indices. *Scientometrics*, 82(2), 391-400. <https://doi.org/10.1007/s11192-009-0047-5>
- Amara, N., & Landry, R. (2012). Counting citations in the field of business and management: Why use Google Scholar rather than the Web of Science. *Scientometrics*, 93(3), 553-581. <https://doi.org/10.1007/s11192-012-0729-2>
- Anthony, D. (2005). The nursing research assessment exercise 2001: An analysis. *Clinical Effectiveness in Nursing*, 9(1-2), 4-12. <https://doi.org/10.1016/j.cein.2005.09.001>

- Baccini, A., & De Nicolao, G. (2022). Just an artifact? The concordance between peer review and bibliometrics in economics and statistics in the Italian research assessment exercise. *Quantitative Science Studies*, 3(1), 194-207. https://doi.org/10.1162/qss_a_00172
- Ball, D. F., & Butler, J. (2000). Research assessment exercises in United Kingdom universities—learning to innovate. *International Journal of Services Technology and Management*, 1(4), 341-357. <https://doi.org/10.1504/IJSTM.2000.001578>
- Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271. <https://doi.org/10.1007/s11192-008-0216-y>
- Baneyx, A. (2008). “Publish or Perish” as citation metrics used to analyze scientific output in the humanities: International case studies in economics, geography, social sciences, philosophy, and history. *Archivum immunologiae et therapiae experimentalis*, 56(6), 363-371. <https://doi.org/10.1007/s00005-008-0043-0>
- Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., & Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy*, 44(2), 451-466. <https://doi.org/10.1016/j.respol.2014.08.004>
- Bonaccorsi, A., & Cicero, T. (2016). Distributed or concentrated research excellence? Evidence from a large-scale research assessment exercise. *Journal of the Association for Information Science and Technology*, 67(12), 2976-2992. <https://doi.org/10.1002/asi.23539>
- Bordons, M., Fernández, M., & Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics*, 53(2), 195-206. <https://doi.org/10.1023/a:1014800407876>
- Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, 103(3), 1123-1144. <https://doi.org/10.1007/s11192-015-1565-y>
- Bornmann, L., & Haunschild, R. (2018). Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data. *PloS one*, 13(5), e0197133. <https://doi.org/10.1371/journal.pone.0197133>
- Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics*, 5(3), 346-359.
- Bornmann, L., Thor, A., Marx, W., & Schier, H. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. *Journal of the Association for Information Science and Technology*, 67(11), 2778-2789. <https://doi.org/10.1002/asi.23627>
- Braun, Y., e.a. (2017). The Burden of Invisible Work in Academia: Social Inequalities and Time Use in Five University Departments, *Humboldt Journal of Social Relations*, 39, 228-245. <https://www.jstor.org/stable/90007882>

- Butler, N., & Spoelstra, S. (2014). The regime of excellence and the erosion of ethos in critical management studies. *British journal of management*, 25(3), 538-550. <https://doi.org/10.1111/1467-8551.12053>
- Camí, J. (1997). Impactolatría: diagnóstico y tratamiento. *Medicina Clínica*, 109(13), 515-524. <https://www.jcamí.eu/documents/Impactolatria.PDF>
- Campbell, Donald. (1976). Assessing the impact of planned social change. Occasional Paper Series, # 8.
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and program planning*, 2(1), 67-90.
- Cecchi, D., Mazzotta, I., Momigliano, S., & Olivanti, F. (2020). Convergence or polarisation? The impact of research assessment exercises in the Italian case. *Scientometrics*, 124(2), 1439-1455. <https://doi.org/10.1007/s11192-020-03517-2>
- Clarivate Analytics (2021). *Introducing the Journal Citation Indicator A new approach to measure the citation impact of journals in the Web of Science Core Collection*. Disponible en: https://clarivate.com/wp-content/uploads/dlm_uploads/2021/05/Journal-Citation-Indicator-discussion-paper-2.pdf
- Colledge, L. (2014). *Snowball metrics recipe book*. Elsevier https://www.snowballmetrics.com/wp-content/uploads/snowball-recipe-book_HR.pdf
- Cooper, M. (2021). Research: Women Leaders Took on Even More Invisible Work During the Pandemic, *Harvard Business Review*. Disponible en: <https://hbr.org/2021/10/research-women-took-on-even-more-invisible-work-during-the-pandemic>
- Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of informetrics*, 1(3), 193-203. <https://doi.org/10.1016/j.joi.2007.02.001>
- Crespo, N., & Simões, N. (2021). The problem of credit in research evaluation—the case of Economics. *Annals of Library and Information Studies*, 68(3), 225-229. https://repositorio.iscte-iul.pt/bitstream/10071/23121/1/article_83057.pdf
- Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices?. *Journal of the American Society for Information Science and Technology*, 52(7), 558-569. <https://doi.org/10.1002/asi.1097>
- Comité de Ética del Consejo Superior de Investigaciones Científicas (CSIC) (2021). *Código de Buenas Prácticas Científicas del CSIC*. Disponible en: <https://www.icmm.csic.es/img/Codigo-de-Buenas-Practicas.pdf>
- Dawson, J., Findlay, A., & Sparks, L. (2004). The UK research assessment exercise (RAE) 2001 and retail research output. *The International Review of Retail, Distribution and Consumer Research*, 14(4), 479-491. <https://doi.org/10.1080/0959396042000260906>
- Day, A. (2022). Exploratory analysis of text duplication in peer-review reveals peer-review fraud and paper mills. *Scientometrics* 127, 5965–5987. <https://doi.org/10.1007/s11192-022-04504-5>

- De Winter, J. C., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics*, 98(2), 1547-1565. <https://doi.org/10.1007/s11192-013-1089-2>
- Delgado López-Cózar, E. (2018). *De la ruta de oro a la ruta verde de la comunicación científica: negocio editorial y bibliométrico, publicación libre, acceso abierto, evaluación total e individualizada*. Disponible en: <https://digibug.ugr.es/handle/10481/53696>
- Delgado López-Cózar, E., & Martín-Martín, A. (2022). *Detectando patrones anómalos de publicación científica en España: Más sobre el impacto del sistema de evaluación científica*. Disponible en: https://www.researchgate.net/publication/363535388_Detectando_patrones_anomalos_de_publicacion_cientifica_en_Espana_Mas_sobre_el_impacto_del_sistema_de_evaluacion_cientifica
- Delgado López-Cózar, E., Orduña-Malea, E., & Martín-Martín, A. (2019). Google Scholar as a data source for research assessment. In Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M. (Eds.) *Springer handbook of science and technology indicators* (pp. 95-127). Springer, Cham. https://doi.org/10.1007/978-3-030-02511-3_4
- Delgado-López-Cózar, E., Ràfols, I., & Abadal, E. (2021). A call for a radical change in research evaluation in Spain. *Profesional de la información*, 30(3). <https://doi.org/10.3145/epi.2021.may.09>
- Demetrescu, C., Ribichini, A., & Schaerf, M. (2020). Are Italian research assessment exercises size-biased?. *Scientometrics*, 125(1), 533-549. <https://doi.org/10.1007/s11192-020-03643-x>
- Doyle, J. R., & Arthurs, A. J. (1998). Grade inflation in the UK's 1996 research assessment exercise?. *Omega*, 26(4), 461-465. [https://doi.org/10.1016/S0305-0483\(97\)00077-7](https://doi.org/10.1016/S0305-0483(97)00077-7)
- DORA (2012). *San Francisco Declaration on Research Assessment*. Disponible en: <https://sfdora.org/>
- Dunne, S., Harney, S., & Parker, M. (2008). Speaking out: The responsibilities of management intellectuals: A survey. *Organization*, 15(2), 271-282. <https://doi.org/10.1177/1350508407087871>
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131-152. <https://doi.org/10.1007/s11192-006-0144-7>
- Elton, L. (2000). The UK research assessment exercise: unintended consequences. *Higher Education Quarterly*, 54(3), 274-283. <https://doi.org/10.1111/1468-2273.00160>
- European Commission, Science Europe y European University Association et al. (2022). *Coalition for Advancing Research Assessment (COARA)*. Disponible en: https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/commission-signs-agreement-reforming-research-assessment-and-endorses-san-francisco-declaration-2022-11-08_en
- European Commission, Directorate-General for Research and Innovation (2017). *Evaluation of research careers fully acknowledging Open Science practices: rewards, incentives and/or recognition for researchers practicing Open Science*. Disponible en: <https://data.europa.eu/doi/10.2777/75255>

- European Commission, Directorate-General for Research and Innovation (2017). Next-generation metrics: responsible metrics and evaluation for open science, Publications Office. Disponible en: <https://data.europa.eu/doi/10.2777/337729>
- European Commission, Directorate-General for Research and Innovation (2020), Whittle, M., Rampton, J., *Towards a 2030 vision on the future of universities in Europe*, Publications Office. Disponible en: <https://data.europa.eu/doi/10.2777/510530>
- European Commission (2021). *The EU's Open Science Policy*. Disponible en: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en#:~:text=The%20EU's%20open%20science%20policy,-Open%20science%20is&text=That%20is%20why%20the%20Commission,of%20citizens%20and%20end%20users
- Expert Group on Altmetrics (2017). *Next Generation Metrics*. Disponible en: <https://data.europa.eu/doi/10.2777/337729>
- Fang, Z., & Costas, R. (2020). Studying the accumulation velocity of altmetric data tracked by Altmetric.com. *Scientometrics*, 123(2), 1077-1101. <https://doi.org/10.1007/s11192-020-03405-9>
- Feenstra, R. A., Delgado López-Cózar, E., & Pallarés-Domínguez, D. (2021). Research misconduct in the fields of ethics and philosophy: researchers' perceptions in Spain. *Science and engineering ethics*, 27(1), 1-21. <https://doi.org/10.1007/s11948-021-00278-w>
- Fernández Carro, R. (2020). What is a scientific article? A principal-agent explanation. *Social Studies of Science*, 51(2), 298-309. <https://doi.org/10.1177/0306312720951860>
- Fister Jr, I., Fister, I., & Perc, M. (2016). Toward the discovery of citation cartels in citation networks. *Frontiers in Physics*, 49. <https://doi.org/10.3389/fphy.2016.00049>
- Fox, C. W., Paine, C. T., & Sauterey, B. (2016). Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*, 6(21), 7717-7726. <https://doi.org/10.1002/ece3.2505>
- French Open Science Committee (2022). *Paris Call on Research Assessment*. Disponible en: <https://osec2022.eu/paris-call>
- Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, 41(4), 277-304. <https://doi.org/10.1023/B:MINE.0000005155.70870.bd>
- Geuna, A., & Piolatto, M. (2016). Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). *Research Policy*, 45(1), 260-271. <https://doi.org/10.1016/j.respol.2015.09.004>
- Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of informetrics*, 6(4), 674-688. <https://doi.org/10.1016/j.joi.2012.07.001>
- Harzing, A. W., & Alakangas, S. (2017). Microsoft Academic: is the phoenix getting wings?. *Scientometrics*, 110(1), 371-383. <https://doi.org/10.1007/s11192-016-2185-x>

- Haustein, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*, 108(1), 413-423. <https://doi.org/10.1007/s11192-016-1910-9>
- Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting "altmetrics": viewing acts on social media through the lens of citation and social theories. In C. Sugimoto (Ed.). *Theories of informetrics and scholarly communication* (pp. 372-406). Alemania: De Gruyter. <https://doi.org/10.1515/9783110308464-022>
- Haustein S, Costas R, & Larivière V (2015). Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLOS ONE* 10(5): e0127830. <https://doi.org/10.1371/journal.pone.0120495>
- Haustein, S., & Larivière, V. (2015). The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects. En; Isabell M. Welp, Jutta Wollersheim, Stefanie Ringelhan, Margit Osterloh (Eds.). *Incentives and performance* (pp. 121-139). Springer, Cham. https://doi.org/10.1007/978-3-319-09785-5_8
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, 1(1), 387-395. https://doi.org/10.1162/qss_a_00020
- Hug, S. E., & Brändle, M. P. (2017). The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics*, 113(3), 1551-1571. <https://doi.org/10.1007/s11192-017-2535-3>
- Jappelli, T., Nappi, C. A., & Torrini, R. (2017). Gender effects in research evaluation. *Research Policy*, 46(5), 911-924. <https://doi.org/10.1016/j.respol.2017.03.002>
- Järvelin, K., & Persson, O. (2008). The DCI index: Discounted cumulated impact-based research evaluation. *Journal of the American Society for Information Science and Technology*, 59(9), 1433-1440. <https://doi.org/10.1002/asi.20847>
- Karlsson, S. (2017). Evaluation as a travelling idea: Assessing the consequences of Research Assessment Exercises. *Research Evaluation*, 26(2), 55-65. <https://doi.org/10.1093/reseval/rvx001>
- Kim, J. et al. (2014). Diverse Heterogeneous Information Source-Based Researcher Evaluation Model for Research Performance Measurement. En: *Mobile, Ubiquitous, and Intelligent Computing* (pp. 261-266). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40675-1_40
- King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of information science*, 13(5), 261-276. <https://doi.org/10.1177/016555158701300501>
- Knights, D., & Richards, W. (2003). Sex discrimination in UK academia. *Gender, Work & Organization*, 10(2), 213-238. <https://doi.org/10.1111/1468-0432.t01-1-00012>
- Koltun, V., & Hafner, D. (2021). The h-index is no longer an effective correlate of scientific reputation. *PLoS ONE*, 16(4), 1-16. <https://doi.org/http://dx.doi.org/10.17632/wsrd8m2h6.1>

- Kostoff, R. (1997). Use and misuse of metrics in research evaluation. *Science and Engineering Ethics*, 3(2), 109-120. <https://doi.org/10.1007/s11948-997-0002-x>
- Kostoff, R. (1998). The use and misuse of citation analysis in research evaluation. *Scientometrics*, 43(1), 27-43. <https://doi.org/10.1007/bf02458392>
- Kousha, K., & Thelwall, M. (2009). Google book search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, 60(8), 1537-1549. <https://doi.org/10.1002/asi.21085>
- Kousha, K., Thelwall, M., & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for information science and technology*, 62(11), 2147-2164. <https://doi.org/10.1002/asi.21608>
- Larivière, V. et al. (2016). A simple proposal for the publication of journal citation distributions. *BioRxiv*, 062109. <https://doi.org/10.1101/062109>
- Lariviere, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. En Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M. (Eds.) *Springer handbook of science and technology indicators* (pp. 3-24). Springer, Cham. https://doi.org/10.1007/978-3-030-02511-3_1
- League of European Research Universities (LERU) (2012). *LERU Report on Research Universities and Research Assessment*. Disponible en: <https://www.leru.org/files/Research-Universities-and-Research-Assessment-Full-paper.pdf>
- Leiden Manifesto (2015). Leiden Manifesto for Research Metrics. *Nature*, 520, 429-431. <http://www.leidenmanifesto.org>; <https://doi.org/10.1038/520429a>
- LERU (2022). *LERU Framework for the Assessment of Researchers: Disponible en: https://www.leru.org/files/Publications/LERU_PositionPaper_Framework-for-the-Assessment-of-Researchers.pdf*
- Ley Orgánica 11/1983, de 25 de agosto, de Reforma Universitaria. (BOE núm.209, de 1 de septiembre de 1983). <https://www.boe.es/eli/es/lo/1983/08/25/11>
- Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa. (BOE núm.295, de 10 de diciembre de 2013). <https://www.boe.es/eli/es/lo/2013/12/09/8/conLucas>, R. E. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46.
- Lynch, John G. (2012). Business Journals Combat Coercive Citation. *Science*, 335(6073), 1169. <https://doi.org/10.1126/science.335.6073.1169-a>
- Ma, L. (2022). Information, platformized. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24713>
- Martin, B. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36(3), 343-362. <https://doi.org/10.1007/bf02129599>

- Martín-Martín, A., Orduña-Malea, E., & Delgado López-Cózar, E. (2018). Author-level metrics in the new academic profile platforms: The online behaviour of the Bibliometrics community. *Journal of informetrics*, 12(2), 494-509. <https://doi.org/10.1016/j.joi.2018.04.001>
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of informetrics*, 12(4), 1160-1177. <https://doi.org/10.1016/j.joi.2018.09.002>
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871-906. <https://doi.org/10.1007/s11192-020-03690-4>
- Marston, C., & Ayub, A. (2000). Relationship between publications in selected journals and research assessment exercise rankings in 1996 for UK accountancy departments. *Accounting Education*, 9(1), 93-102. <https://doi.org/10.1080/096392800413672>
- Mcculloch, S. (2017). Hobson's choice: the effects of research evaluation on academics' writing practices in England. *Aslib Journal of Information Management*, 69(5), 503-515. <https://doi.org/10.1108/AJIM-12-2016-0216>
- Mingers, J. (2014). Problems with SNIP. *Journal of Informetrics*, 8(4), 890-894. <https://doi.org/10.1016/j.joi.2014.09.004>
- Mingers, J., & Meyer, M. (2017). Normalizing Google Scholar data for use in research evaluation. *Scientometrics*, 112(2), 1111-1121. <https://doi.org/10.1007/s11192-017-2415-x>
- Mingers, J., & Yang, L. (2017). Evaluating journal quality: A review of journal citation indicators and ranking in business and management. *European journal of operational research*, 257(1), 323-337. <https://doi.org/10.1016/j.ejor.2016.07.058>
- Ministerio de Educación de la República Popular China (2020). Varios Dictámenes sobre la Regulación del Uso de Indicadores Relacionados con Documentos SCI en Instituciones de Educación Superior y el Establecimiento de una Correcta Orientación a la Evaluación. Disponible en: http://www.moe.gov.cn/srcsite/A16/moe_784/202002/t20200223_423334.html
- Ministerio de Universidades (2022). *Proyecto de Real Decreto por el que se modifica el Real Decreto 1393/2007, de 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias oficiales y el Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de doctorado.* Disponible en: <https://www.educacionyfp.gob.es/dam/jcr:b48a6ae4-ce24-464a-a0c4-01e379a52e35/ordenacion-ensenanzas-universitarias.pdf>
- Ministerio de Universidades (2022). Anteproyecto de Ley Orgánica del sistema universitario. Disponible en: <https://www.universidades.gob.es/portal/site/universidades/menuitem.21ef60083f296675105f2c10026041a0/?vgnnextoid=660607559eaab710VgnVCM1000001d04140aRCRD>

- Moed, H. F. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity?. *Scientometrics*, 74(1), 153-161. <https://doi.org/10.1007/s11192-008-0108-1>
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of informetrics*, 4(3), 265-277. <https://doi.org/10.1016/j.joi.2010.01.002>
- Moed, H. F., Van Leeuwen, T. N., & Reedijk, J. (1998). A new classification system to describe the ageing of scientific journals and their impact factors. *Journal of Documentation*, 54(4), 387-419. <https://doi.org/10.1108/EUM0000000007175>
- Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, 65(8), 1627-1638. <https://doi.org/10.1002/asi.23071>
- Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. (2020) The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS biology* 18(7), e3000737. <https://doi.org/10.1371/journal.pbio.3000737>
- Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS biology*, 16(3), e2004089. <https://doi.org/10.1371/journal.pbio.2004089>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228. <https://doi.org/10.1007/s11192-015-1765-5>
- Mongeon, P., Waltman, L., & Rijcke, S. (2016). What do we know about journal citation cartels? A call for information. *Leiden Madtrics* [blog]. Disponible en: <https://www.cwts.nl/blog?article=n-q2w2b4>.
- Norman, I., & Griffiths, P. (2008). Duplicate publication and 'salami slicing': ethical issues and practical solutions. *International Journal of Nursing Studies*, 45(9), 1257-1260. <https://doi.org/10.1016/j.ijnurstu.2008.07.003>
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of informetrics*, 1(2), 161-169.
- Northcott, D., & Linacre, S. (2010). Producing spaces for academic discourse: The impact of research assessment exercises and journal quality rankings. *Australian Accounting Review*, 20(1), 38-54. <https://doi.org/10.1016/j.joi.2006.12.001>
- O'Connell, B. T., De Lange, P., Stoner, G., & Sangster, A. (2020). Impact of research assessment exercises on research approaches and foci of accounting disciplines in Australia. *Accounting, Auditing & Accountability Journal*, 33(6), 1277-1302. <https://doi.org/10.1108/AAAJ-12-2019-4293>
- O'Connor, S. (2022). The San Francisco Declaration on Research Assessment (DORA) in Nursing Science. *Nursing Science Quarterly*, 35(2), 275-276. <https://doi.org/10.1177/08943184211070602>
- Open Scholar C.I.C. (2014). *Independent Peer Review Manifesto*. Disponible en: <https://www.openscholar.org.uk/independent-peer-review-manifesto>

- Oppenheim, C. (1995). The correlation between citation counts and the 1992 Research Assessment Exercise Ratings for British library and information science university departments. *Journal of Documentation*, 51(1), 18-27. <https://doi.org/10.1108/eb026940>
- Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of documentation*, 53(5), 477-487. <https://doi.org/10.1108/EUM0000000007207>
- Orduña-Malea, E. (2021). Carta: La insoportable levedad de la evaluación de proyectos de investigación: ¿necesitamos regularizar la actividad profesional de la evaluación científica?. *Profesional de la información*. 30(5), e300413. <https://doi.org/10.3145/epi.2021.sep.13>
- Orduña-Malea, E.; Bautista-Puig, N., & Perez-Esparrells, C. (2022). *Propuesta de un modelo de evaluación del profesorado universitario*. Canarias: Agencia Canaria de Calidad Universitaria y Evaluación Educativa (ACCUEE).
- 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. <https://www.scielo.org/pdf/resp/2016.v90/e20006/en>
- Ortega, J. L. (2020). Altmetrics data providers: A metaanalysis review of the coverage of metrics and publication. *El profesional de la información*, 29(1). <https://doi.org/10.3145/epi.2020.ene.07>
- Oviedo-García, M.A.; Casillas Bueno, J.C., & González Rodríguez, M.R. (2021). *Análisis bibliométrico e impacto de las editoriales open-access en España*. Madrid: Aneca. http://www.aneca.es/content/download/16664/202231/file/210930_Openaccess.pdf
- Overlaet, B. (2022). A Pathway Towards Multidimensional Academic Careers. A LERU Framework for the Assessment of Researchers. *LERU Position Paper*. Disponible en: https://www.leru.org/files/Publications/LERU_PositionPaper_Framework-for-the-Assessment-of-Researchers.pdf
- Paisey, C., & Paisey, N. J. (2005). The research assessment exercise 2001—insights and implications for accounting education research in the UK. *Accounting Education: an international journal*, 14(4), 411-426. <https://doi.org/10.1080/06939280500347134>
- Parra, C., Imran, M., Mirylenka, D., Daniel, F., Casati, F., & Marchese, M. (2011). A Scientific Resource Space for Advanced Research Evaluation Scenarios. En *SEBD* (pp. 203-214). https://mimran.me/papers/para_imran_daniil_sebd2011.pdf
- Prathap, G. (2010). Is there a place for a mock h-index?. *Scientometrics*, 84(1), 153-165. <https://doi.org/10.1007/s11192-009-0066-2>
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2011). Altmetrics: A manifesto. Disponible en: <http://altmetrics.org/manifesto/>
- Puerta, M. S., & Galeano, M. M. O. (2005). Reflexiones sobre la importancia de la Crítica de Lucas. *Ecos de Economía*, 9(20), 1-12.

- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, 105(45), 17268-17272. <https://doi.org/10.1073/pnas.0806977105>
- Rafols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation*, 28 (1), 7-22. <https://doi.org/10.1093/reseval/rvy030>
- Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research policy*, 41(7), 1262-1282. <https://doi.org/10.1016/j.respol.2012.03.015>
- Ràfols, I., & Molas-Gallart, J. (2022). How to reform research evaluation in Spain. *Institutional accreditation as a response to the European Agreement on research assessment. Letter. Profesional De La información*, 31(6). <https://doi.org/10.3145/epi.2022.nov.01>
- Real Decreto 1086/1989, de 28 de agosto, sobre retribuciones del profesorado universitario. (BOE núm.216, de 9 de septiembre de 1989). <https://www.boe.es/eli/es/rd/1989/08/28/1086>
- Rebora, G., & Turri, M. (2013). The UK and Italian research assessment exercises face to face. *Research policy*, 42(9), 1657-1666. <https://doi.org/10.1016/j.respol.2013.06.009>
- Resolución de 28 de noviembre de 2018, de la Secretaría de Estado de Universidades, Investigación, Desarrollo e Innovación, por la que se fija el procedimiento y plazo de presentación de solicitudes de evaluación de la actividad investigadora a la Comisión Nacional Evaluadora de la Actividad Investigadora (BOE núm. 289, de 30 de noviembre de 2018). https://www.boe.es/diario_boe/txt.php?id=BOE-A-2018-16379
- Repiso, R., Aguaded, I., Castillo-Esparcia, A., & Montero, J. (2020). Sexenio de transferencia: luces y sombras de un proyecto necesario. Análisis de una muestra de denegaciones en Ciencias de la Información y estudio de la edición de revistas y comisiones de agencia como mérito. *Anuario ThinkEPI*, 14. <https://doi.org/10.3145/thinkepi.2020.e14a02>
- Resnik, D. (1998). *The Ethics of Science* New York: Routledge
- Resnik, D. (2014). Data fabrication and falsification and empiricist philosophy of science. *Science and engineering ethics*, 20(2), 423-431. <https://doi.org/10.1007/s11948-013-9466-z>
- Rodríguez-Navarro, A., & Brito, R. (2019). Probability and expected frequency of breakthroughs: Basis and use of a robust method of research assessment. *Scientometrics*, 119(1), 213-235.
- Ronald, R., & Fred, Y. Y. (2013). A multi-metric approach for research evaluation. *Chinese Science Bulletin*, 58(26), 3288-3290. <https://doi.org/10.1007/s11434-013-5939-3>
- Rowlands, J., & Wright, S. (2020). The role of bibliometric research assessment in a global order of epistemic injustice: A case study of humanities research in Denmark. *Critical Studies in Education*, 1-17. <https://doi.org/10.1080/17508487.2020.1792523>
- Rushforth, A., & de Rijcke, S. (2015). Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands. *Minerva*, 53(2), 117-139. <https://doi.org/10.1007/s11024-015-9274-5>

- Schleicher, A. (2020). The Impact of COVID-19 on Education: Insights from Education at a Glance 2020. *OECD Publishing*. <https://www.oecd.org/education/the-impact-of-covid-19-on-education-insights-education-at-a-glance-2020.pdf>
- Sheikh, A. (2000). Publication ethics and the research assessment exercise: reflections on the troubled question of authorship. *Journal of Medical Ethics*, 26(6), 422-426. <http://dx.doi.org/10.1136/jme.26.6.422>
- Sidorkin, A. M. (2016). Campbell's Law and the Ethics of Immensurability. *Studies in Philosophy and Education*, 35(4), 321-332. doi: 10.1007/s11217-015-9482-3.
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126(6), 5113-5142. <https://doi.org/10.1007/s11192-021-03948-5>
- Sotudeh, H., Ravaei, M., & Mirzabegi, M. (2018). Comparing the opportunities provided by altmetrics and citation analysis for research evaluation. *Iranian Journal of Information Processing and Management*, 34(1), 113-138.
- Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131-1143. <https://doi.org/10.1007/s11192-013-1117-2>
- Šupak Smolčić, V. (2013). Salami publication: Definitions and examples. *Biochemia Medica*, 23(3), 237-241. <https://doi.org/10.11613/BM.2013.030>
- Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the association for information science and technology*, 68(9), 2037-2062. <https://doi.org/10.1002/asi.23833>
- Szomszor, M., Pendlebury, D. A., & Adams, J. (2020). How much is too much? The difference between research influence and self-citation excess. *Scientometrics*, 123(2), 1119-1147. <https://doi.org/10.1007/s11192-020-03417-5>
- Taylor, J. (2011). The assessment of research quality in UK universities: Peer review or metrics?. *British Journal of Management*, 22(2), 202-217. <https://doi.org/10.1111/j.1467-8551.2010.00722.x>
- Teixeira da Silva, J. A. (2021). The i100-index, i1000-index and i10, 000-index: expansion and fortification of the Google Scholar h-index for finer-scale citation descriptions and researcher classification. *Scientometrics*, 126(4), 3667-3672. <https://doi.org/10.1007/s11192-020-03831-9>
- Teixeira da Silva, J.A. (2021). Abuse of ORCID's weaknesses by authors who use paper mills. *Scientometrics* 126, 6119–6125. <https://doi.org/10.1007/s11192-021-03996-x>
- Teixeira da Silva, J. A., & Dobránszki, J. (2016). Multiple authorship in scientific manuscripts: ethical challenges, ghost and guest/gift authorship, and the cultural/disciplinary perspective. *Science and engineering ethics*, 22(5), 1457-1472. <https://doi.org/10.1007/s11948-015-9716-3>
- Thelwall, M. (2018a). Dimensions: A competitor to Scopus and the Web of Science?. *Journal of informetrics*, 12(2), 430-435. <https://doi.org/10.1016/j.joi.2018.03.006>

- Thelwall, M. (2018b). Early Mendeley readers correlate with later citation counts. *Scientometrics*, 115(3), 1231-1240. <https://doi.org/10.1007/s11192-018-2715-9>
- Thelwall, M. (2018c). Using altmetrics to support research evaluation. In *International Workshop on Altmetrics for Research Outputs Measurements and Scholarly Information Management* (pp. 11-28). Springer, Singapore. http://doi.org/10.1007/978-981-13-1053-9_2
- Thelwall, M. (2020). The Pros and Cons of the Use of Altmetrics in Research Assessment. *Scholarly Assessment Reports*, 2(1), p.2. <http://doi.org/10.29024/sar.10>
- Torres-Salinas, D., Valderrama-Baca, P., & Arroyo-Machado, W. (2022). Is there a need for a new journal metric? Correlations between JCR Impact Factor metrics and the Journal Citation Indicator—JCI. *Journal of Informetrics*, 16(3), 101315. <https://doi.org/10.1016/j.joi.2022.101315>
- Traynor, M., & Rafferty, A. M. (1999). Nursing and the research assessment exercise: past, present and future. *Journal of advanced nursing*, 30(1), 186-192. <https://doi.org/10.1046/j.1365-2648.1999.01063.x>
- UNESCO (2021a). *Thinking Higher and Beyond Perspectives on the Futures of Higher Education to 2050*. ISBN 978-980-7175-57-9. Disponible en: <https://www.iesalc.unesco.org/en/2021/05/26/report-on-the-futures-of-higher-education-envisions-collective-and-holistic-responses-to-global-challenges/>
- UNESCO (2021b). *UNESCO Recommendation on Open Science*. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa
- Van den Akker, W. J. (2016). Yes We Should: Research Assessment in the Humanities. En M. Ochsner, S.E. Hug, H.D. Daniel (Eds.). *Research assessment in the humanities: towards criteria and procedures* (pp. 23-29). Springer Open.
- Van Raan, A. F. (2005). Measurement of central aspects of scientific research: Performance, interdisciplinarity, structure. *Measurement: Interdisciplinary Research and Perspectives*, 3(1), 1-19. https://doi.org/10.1207/s15366359mea0301_1
- Varpio, L., Ajjawi, R., Monrouxe, L. V., O'Brien, B. C., & Rees, C. E. (2017). Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Medical education*, 51(1), 40-50.
- Vasen, F. The 'Ivory Tower' as Safe Bet: Science Policies and Research Evaluation in Mexico. *Education Policy Analysis Archives*, 26, 1-27.
- Vaughan, L., & Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*, 74(2), 317-330. <https://doi.org/10.1007/s11192-008-0220-2>
- Vera-Baceta, M. A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, 121(3), 1803-1813. <https://doi.org/10.1007/s11192-019-03264-z>
- Vidovich, L. (2008). Research Assessment in Singaporean Higher Education: Changing Educational Accountabilities in a Context of Globalisation. *International Education Journal: Comparative Perspectives*, 9(1), 41-52. <http://files.eric.ed.gov/fulltext/EJ894341.pdf>

- Waltman, L., van Eck, N. J., Van Leeuwen, T. N., & Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of informetrics*, 7(2), 272-285. <https://doi.org/10.1016/j.joi.2012.11.011>
- Warner, J. (2000). A critical review of the application of citation studies to the Research Assessment Exercises. *Journal of Information Science*, 26(6), 453-459. <https://doi.org/10.1177/016555150002600607>
- Wildgaard, L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar. *Scientometrics*, 104(3), 873-906. <https://doi.org/10.1007/s11192-015-1608-4>
- Wilsdon, J. (2015). *The metric tide: Independent review of the role of metrics in research assessment and management*. London: Sage. <https://dx.doi.org/10.4135/9781473978782>
- Williams, T., & Morrone, J. J. (2018). Science is strengthened by Mexico's researcher evaluation system: Factual errors and misleading claims by Neff. *Science and Public Policy*, 45(5), 742-745. <https://doi.org/10.1093/scipol/scy004>
- YERUN (2021). Yerun Position Paper on Research Assessment. Disponible en: https://yerun.eu/wp-content/uploads/2021/11/YERUN-Position-Paper_Research-assessment.pdf