

Medidas insuficientes para un cambio en la evaluación de la investigación en España: glosando las nuevas directrices de la ANECA

EMILIO DELGADO LÓPEZ-CÓZAR¹ (Universidad de Granada), ISMAEL RÀFOLS (Leiden University) y ERNEST ABADAL (Universitat de Barcelona)

Artículo recibido: 16 de noviembre de 2021
Artículo aceptado: 18 de noviembre de 2021

Delgado López-Cozar, Emilio, Ràfols, Ismael y Abadal, Ernest (2021). Medidas insuficientes para un cambio en la evaluación de la investigación en España: glosando las nuevas directrices de la ANECA. *Recerca. Revista de Pensament i Anàlisi*, 27(2), pp. 1-10. doi: <http://dx.doi.org/10.6035/recerca.6308>

INTRODUCCIÓN

En una carta publicada hace seis meses (Delgado López-Cózar, Ràfols y Abadal, 2021), pocas semanas después de que la Agencia Española de Evaluación se adhiriera a la declaración DORA (DORA, 2012), hacíamos un llamamiento a las autoridades científicas españolas para que abandonaran las políticas de evaluación de la investigación basadas en un uso excesivo e indiscriminado de los indicadores bibliométricos —especialmente el *Journal Impact Factor* (JIF)—, para valorar el desempeño individual de los académicos. En concreto, se les animaba a «que suscriban y cumplan con la DORA y sigan las recomendaciones en el uso de indicadores bibliométricos señalados en el *Manifiesto Leiden*» (Hicks et al., 2015). Una mesa redonda organizada en el mes de julio con participación de distintos agentes del sistema español de I+D+I coincidía en la necesidad de acometer ese cambio de forma urgente.

En este contexto, el 1 de octubre pasado la ANECA publicó unos «Principios y directrices para la actualización de criterios de evaluación de la investigación de ANECA 2021» (ANECA, 2021), presentándolos públicamente a través

¹ **Autor de correspondencia:** Emilio Delgado López-Cózar, edelgado@ugr.es

de su canal YouTube (ANECA TV) a fin de ampliar información y resolver las dudas planteadas.

En un ejercicio de transparencia, insólito en esta agencia en sus veinte años de existencia, pero que debiera ser la guía que oriente la rendición de cuentas de todo organismo público, la ANECA anunció, entre otras cosas:

1. Una serie de principios que inspirarán su actuación, entre los que cabría destacar la calidad, la transparencia, la integridad académica, la estabilidad en sus políticas, igualdad...
2. Una revisión simultánea de los criterios de evaluación del componente investigador en los tres programas de evaluación y acreditación del profesorado que coordina la agencia (PEP, ACADEMIA y Sexenios).
3. La adhesión al *Manifiesto de Leiden* de buenas prácticas en el uso de los indicadores bibliométricos con fines evaluativos.
4. Mantener el JIF como principal criterio de evaluación de las publicaciones, complementado con indicadores cuantitativos a nivel de artículo.
5. Introducir criterios cualitativos en la evaluación de la producción científica primando la evaluación de la calidad sobre la cantidad.
6. Apoyar al acceso abierto a las publicaciones y los datos de la investigación.
7. Estudiar cómo potenciar la multidisciplinariedad y no penalizarla.
8. Atajar las desviaciones detectadas en los comportamientos de publicación.

Una vez enfriados los ánimos tras la estridente repercusión mediática generada por el desatinado informe adjunto, «Análisis bibliométrico e impacto de las editoriales open-access en España» (Oviedo-García, Casillas-Bueno y González-Rodríguez, 2021), y en concreto por sus «listados de revistas con un comportamiento no estándar» (anexos I y II), quisiéramos desde la serenidad y la distancia compartir nuestra opinión sobre lo que consideramos más relevante, que es glosar las directrices anunciadas por la ANECA para revisar sus criterios de evaluación. Nos resistimos a que algunos aberrantes resultados del informe adjunto, producto de la impericia de sus autores y de su incapacidad de contextualizar los datos e interpretar correctamente los resultados obtenidos, opaquen lo que seguramente será más determinante para la investigación científica española de los próximos años.

1. SOBRE INCENTIVOS Y SUS EFECTOS: ACTUANDO SOBRE LAS CAUSAS PARA EVITAR SUS CONSECUENCIAS

Es una máxima no solo en la ciencia sino en la vida que, si no se efectúa un correcto diagnóstico de un problema, se errará en la terapia para resolverlo. Asimismo, si se confunden las causas con las consecuencias, será difícil enfocar cualquier asunto.

Pues bien, aquí no solo hay que abordar los negativos efectos de la evaluación (las malas prácticas de publicación y las desviaciones en la recta conducta de la investigación) sino la causa de los mismos (la presión por publicar en revistas de impacto). Obviamente, las faltas a la integridad académica deben ser penalizadas y atajadas de forma inmediata. Pero la mejor forma de hacerlo es actuando sobre la raíz del problema. Es bien conocido, como ya advertimos en nuestra carta, que un uso indiscriminado de un indicador inadecuado, como es el factor de impacto de las revistas para medir el rendimiento científico, crea incentivos perversos que conducen a prácticas aviesas y aberrantes (Weingart, 2005). Es cambiando los principios y las directrices que gobiernan la evaluación de la investigación como zanjaremos el problema, no poniendo parches a los efectos que estos generan.

2. ¿UNA EVALUACIÓN QUE INTEGRA LAS DISTINTAS CONTRIBUCIONES DE LOS ACADÉMICOS?

Haciendo gala del deseo de armonizar sus regulaciones, la ANECA anuncia una revisión de los criterios de evaluación de la investigación en paralelo en todos sus programas. Puede parecer un pequeño paso en la línea de evitar contradicciones, pero pensamos que es peor, ineficaz y vano para reorientar la actividad investigadora, en particular, y la académica, en general, hacia derroteros más deseables, es decir, hacia una ciencia *mejor*: más creativa y que contribuya más al bienestar material y cultural de los ciudadanos.

Y esto, básicamente, porque, como ya advertíamos en nuestra carta, uno de los efectos tóxicos del sistema de evaluación actual es el desequilibrio y el empobrecimiento de la actividad académica. Otorgar el máximo peso al componente investigador y, dentro de él, a la publicación científica ha conducido, si no al abandono, sí al descuido de las tareas académicas *no productivas*.

Para evitar este efecto indeseado, hay que abordar la revisión del componente investigador conjuntamente con el resto de las actividades que confor-

man el quehacer de un académico: docencia, contribución social y transferencia y gestión. Se precisa una revisión integral e integrada de los criterios que evalúen toda la actividad académica de manera más equilibrada, permitiendo diversas rutas que ofrezcan una progresión diversificada en la carrera académica. Con ello preservaremos la *biodiversidad* académica tan necesaria en el *ecosistema* científico. Además, así, se amparará la libertad individual en la elección de la ruta académica más adecuada.

3. INVESTIGAR ES ALGO MÁS QUE PUBLICAR

El protagonismo de la publicación en el documento de directrices de la ANECA es absoluto. Sigue siendo el santo y seña de la evaluación y no solo por el peso que se le otorga, sino porque pareciera que no existe investigación evaluable fuera de la publicación científica. No debe olvidar la ANECA que el objetivo último de cualquier sistema de evaluación de la investigación es mejorar la investigación misma para que la ciencia pública consiga sus objetivos hacia la sociedad: generar nuevos conocimientos y compartirlos con los actores sociales y los ciudadanos (como enfatiza la ciencia abierta). Para ello, la evaluación tendría que reconocer a aquellos investigadores que generan conocimiento original y lo comparten hacia objetivos de bienes públicos.

Justamente uno de los males detectados y de los efectos indeseados del sistema de evaluación español es que la publicación se ha convertido en una obsesión para el investigador, que solo piensa en la mejor forma de empaquetar un trabajo y no en resolver los problemas científicos, técnicos y sociales que este puede elucidar o resolver. Publicar lo que sea y como sea se ha convertido en el lema de muchos investigadores en España y en otros países con sistemas de evaluación centrados en la publicación. La relevancia, novedad y originalidad tienen un peso mucho menor en la consecución de acreditaciones y méritos.

¿Está segura la ANECA que el sistema de evaluación de la investigación vigente está mejorando la capacidad de España para producir descubrimientos? ¿No se estará produciendo una inflación de publicaciones que se justifican solo por la necesidad de responder a los requerimientos del sistema evaluador? ¿No están proliferando las publicaciones vacuas o estériles pero publicables en revistas que las acogen para hacer florecer su negocio editorial? ¿No es hora de fomentar la investigación realmente *importante* anteponiendo la evaluación de lo que se aporta por encima del medio donde se publique?

Hay que recomponer el espacio académico desactivando la presión por publicar mediante la valoración de otros productos de la investigación que permitan en igualdad de oportunidades otras vías para el progreso en la carrera. La publicación no puede ser el elemento que todo lo dé o todo lo quite.

4. ADHESIÓN A BUENAS PRÁCTICAS DE EVALUACIÓN CIENTÍFICA: ¿LEIDEN, SÍ? ¿DORA, NO?

La ANECA realizó una enfática declaración de *firma* y adhesión pública al *Manifiesto de Leiden* leyendo cada uno de sus diez principios. En cambio, se soslayó la *Declaración de San Francisco* (DORA). Ya advertíamos en nuestra carta que una cosa es lo que se dice y otra lo que se hace. En este caso, el problema es que hay contradicciones entre lo que supuestamente se dice que se quiere defender (principios del *Manifiesto de Leiden*) y lo que a continuación se declara que se quiere hacer o se hace (los criterios de evaluación publicados).

Aunque entendemos y valoramos el bienintencionado ánimo de la ANECA de seguir buenas prácticas en el uso de indicadores bibliométricos para la evaluación, debemos manifestar que Leiden y DORA son dos caras de una misma moneda.

DORA nació en el 2012 de la apremiante necesidad de mejorar la forma en que las agencias de financiación, las instituciones académicas y otros grupos evalúan la investigación científica, rechazando explícitamente las inconsistencias del JIF y su problemática aplicación a la evaluación individual de los científicos. De modo similar, el *Manifiesto de Leiden* surgió en el 2015 desde la comunidad bibliométrica por la constatación de los malos usos generalizados de los indicadores (*además* del JIF, *también* el índice H, la falta de normalización por campos, la falta de rigor estadístico, etc.) en la evaluación del desempeño científico y de la necesidad de codificar unos principios básicos para su buen uso. Así, el *Manifiesto de Leiden* apoya DORA en su crítica al JIF, critica otros usos de indicadores y propone algunas alternativas.

El mantenimiento del JIF como referente en la evaluación por parte de la ANECA con su jerarquización en forma de cuartiles y terciles, así como de otros indicadores de revista (Citescore, SJR, Article influence, JIF sin autocitas) de similar filosofía, aunque de distinto detalle técnico, va en contra de varios de los principios enunciados en el *Manifiesto de Leiden* y, por supuesto, de DORA. Sustituir el SJR por Citescore, por mucho que uno sea más transparente que otro, no cambia el fondo del problema.

El JIF contradice abiertamente varios de los principios del *Manifiesto de Leiden*:

- El principio siete, donde se recomienda que «la evaluación individual de investigadores debe basarse en la valoración *cualitativa* de su portafolio de investigación» (la cursiva es nuestra).

- El principio ocho, donde se aconseja «evitar la concreción impropia y la falsa precisión», poniendo como ejemplo justamente lo que ocurre con el JIF: «el factor de impacto de revistas se publica con tres decimales para evitar empates. Sin embargo, dada la ambigüedad conceptual y la variabilidad aleatoria de las citas, no tiene sentido distinguir entre revistas por pequeñas diferencias en el factor de impacto. Se debe evitar la falsa precisión».

- El principio tres, donde se apela a proteger la excelencia en investigación de relevancia local, poniendo justamente como ejemplo de mala práctica la política evaluativa española basada en el JIF, que incentiva la indagación en temas básicos y de interés general, penalizando los de orientación local, especialmente en las áreas de humanidades y ciencias sociales (López-Piñero y Hicks, 2015).

Ya advertimos en nuestra carta que no basta con adherirse o firmar manifiestos, sino que hay que tener voluntad de cumplirlos; pero, para conseguirlo, lo primero es interpretar correctamente lo que señalan. Saludando la voluntad de cambio expresada por la ANECA, debemos criticar su lectura e interpretación y su incorrecta aplicación a las directrices para la revisión de sus criterios.

Entendemos que la ANECA perciba la firma de DORA como una impugnación global a sus prácticas evaluativas. Abandonar repentinamente el JIF substituyendo de la noche a la mañana unas prácticas evaluativas que han guiado la actuación de la ANECA en sus veinte años de historia es difícil de asumir e implementar. Quizás es sensato producir un cambio gradual, pero el comienzo de esta etapa debe partir de una manifestación explícita del abandono de estos indicadores en un futuro próximo. Animamos a adoptar esta posición realista que asegura la estabilidad y la seguridad de los evaluados y una adaptación no traumática de muchos evaluadores.

La ANECA debe convencerse de que el enfoque de evaluación de la calidad de una publicación basado en la revista es erróneo, porque, por principio, no se puede evaluar el contenido (artículo) por su continente (revista), al igual que es absurdo valorar un mensaje por su medio de difusión. Sabemos, además, que el indicador del continente (JIF o sucedáneos) no es representativo de los artículos que contiene ni es un predictor seguro de su impacto futuro.

Pero hay otras razones que aconsejan abandonar el indicador: es perverso y crea efectos dañinos para la ciencia. Hoy el uso del JIF y sus cuartiles y terciles promueve la injusticia, premia la mediocridad y no la calidad e incentiva la comodidad y no la creatividad y el riesgo. Las revistas del primer y segundo cuartil han incrementado notablemente el número de artículos que publican como respuesta directa a la creciente demanda incitada por sistemas de evaluación como el propiciado por la ANECA. Según datos de Moya (2020), el número de documentos publicados en revistas del primer cuartil en Scopus cuadruplica al de los publicados en el cuarto cuartil. En el 2019, en las revistas del primer y segundo cuartil se publicaron el 70 % de los documentos publicados en las revistas Scopus de ese año (Moya, 2020).

Dado que la distribución de citas es asimétrica dentro de una revista, esto es, que solo unos pocos artículos concentran la mayoría de las citas, lo que estamos haciendo no es primar y valorar la calidad de los trabajos. Muchos artículos de nula repercusión se benefician de la posición ocupada por la revista donde se publican, con lo cual, lo que realmente se está recompensando, es la mediocridad. Premiar a muchos con el rendimiento de unos pocos no solo es falaz sino injusto, creando espejismos que llevan a un estado de autocomplacencia paralizante. En la medida en que este *statu quo* favorece cada vez a más autores —el número de firmantes crece sin cesar—, genera un ambiente de comodidad que explica el rechazo a cualquier cambio que pueda empeorar las condiciones de evaluación. Los científicos prefieren no adentrarse en problemas científicos que puedan tener riesgo de no producir resultados publicables y se limitan a producir artículos rutinarios que puedan ser aceptados sin excesivos problemas.

Muchas de las revistas bien posicionadas en los JIF se aprovechan de su situación de dominio para captar cuantos más artículos mejor, eliminando o suavizando la relevancia, originalidad o novedad como filtros de calidad en sus procesos de revisión, con lo que se pueden estar publicando trabajos metodológicamente irreprochables pero irrelevantes por sus resultados. Gracias a los cargos por procesamiento de artículos (APC en inglés), que los autores pagan con gusto por su rentabilidad académica, las editoriales obtienen un pingüe beneficio.

5. ¿MENOS CANTIDAD, MÁS CALIDAD?: COMPLEMENTANDO LOS INDICADORES DE IMPACTO REVISTA CON INDICADORES CUANTITATIVOS Y CUALITATIVOS A NIVEL DE ARTÍCULO

Evaluar la calidad y no la cantidad, resaltando la necesidad de originalidad en la generación de conocimiento de las aportaciones alegadas, es uno de los principios declarados que inspirarán la revisión de criterios. Para conseguirlo, la ANECA propone la introducción de indicadores bibliométricos a nivel de artículo y otros de carácter cualitativo.

No obstante, esta declaración general esconde no pocas inconcreciones y contradicciones. Ciertamente, admitir por primera vez indicadores que miden el valor individual de los artículos —*Category Normalized Citation Impact* (CNCI) a partir de la Web of Science y el *Field Weighted Citation Impact* (FWCI) de Scopus— es un paso más ajustado a lo que pretenden los criterios enunciados por la ANECA, que es averiguar la repercusión de los trabajos de investigación sometidos a evaluación.

Siendo esta una buena noticia en el sentido de que estos indicadores apuntan a captar el valor intrínseco de la investigación a través del impacto científico de cada artículo, se introducen tímidamente como complementos al indicador basado en el impacto del canal donde se publican. Así, las directrices de la ANECA establecen que:

Por regla general el nivel adjudicado a un artículo podrá seguir siendo el derivado del factor de impacto de la revista en la que se haya publicado en las bases de datos indexadas aceptadas en cada campo o las medidas como cuartiles o terciles usualmente usados en cada campo.

Por otra parte, se sugiere a los comités que han de juzgar los CV a «complementar el índice de impacto de las revistas con otras medidas e indicadores de calidad de las mismas como el índice de impacto sin autocitas o [...] el *Article Influence*», así como sustituir el SJR por el Citescore. Es evidente que la noción de calidad se sigue asociando a los índices de impacto de las revistas y a los indicadores cuantitativos, siguiendo la tradición española de las últimas décadas (Delgado López-Cózar y Martín-Martín, 2019; Feenstra y Delgado López-Cózar, 2021).

Al mismo tiempo que se opta por esta vía cuantitativa tan claramente concretada, la ANECA defiende la introducción de valoraciones cualitativas. No obstante, una mera declaración no sirve para nada si no se estipulan las

medidas que permitirán llevarlas a efecto. De hecho, se deja en manos de los comités su implementación: «Cuando proceda realizar una evaluación cualitativa detallada de la aportación (sexenio, cuatro aportaciones relevantes de ACADEMIA) los comités y comisiones incluirán orientaciones sobre las cuestiones a valorar».

Lo que resulta realmente más sorprendente es que, dentro de los denominados por la ANECA «Estándares e indicadores para la evaluación cualitativa de la producción científica», se incluya el depósito en «green open access» de las aportaciones sometidas a evaluación. No debiera confundirse el depósito en abierto de una publicación o de los datos que la soportan con un indicador de calidad. Si acaso, debería ser un requisito previo y obligatorio para todos los trabajos financiados con fondos públicos. En esta línea estimamos como muy positiva la posibilidad de exigencia futura.

6. CONCLUSIÓN

En resumen, de acuerdo con el análisis que hemos presentado, las nuevas medidas de la ANECA no suponen un cambio sustancial. Aunque parece que haya buenas intenciones para pluralizar la evaluación, España va a continuar con un sistema centralizado, burocrático e individual de evaluación con un indicador de la revista de publicación (JIF) como principal criterio. Las nuevas medidas permiten flexibilizar la aplicación de este criterio, pero, como no cambian ni la filosofía y ni los mecanismos de la evaluación, mucho nos tememos que los efectos negativos van a continuar.

Como decíamos en la carta, hay que acometer una reforma que cambie de raíz (radical) el sistema de evaluación. Ajustes menores no valen en un sistema de evaluación anclado en el siglo pasado si queremos que la ciencia nos ayude a construir el futuro.

BIBLIOGRAFÍA

ANECA (2021). *Principios y directrices para la actualización de criterios de evaluación de la investigación de ANECA 2021*. Madrid: ANECA. Recuperado de: <http://www.aneca.es/content/download/16660/202193/file/210930>

- De Moya-Anegón, Félix (2020). Research evaluation entities cause a shift of publication to Q1 journals. *El Profesional de la Información*, 29(4), e290431. doi: 10.3145/epi.2020.jul.31
- Delgado López-Cózar, Emilio y Martín-Martín, Alberto (2019). El Factor de Impacto de las revistas científicas sigue siendo ese número que devora la ciencia española: ¿hasta cuándo? *Anuario ThinkEPI*, 13, e13e09. doi: 10.3145/thinkepi.2019.e13e09.
- Delgado López-Cózar, Emilio, Ràfols, Ismael y Abadal, Ernest (2021). Letter: A call for a radical change in research evaluation in Spain. *Profesional de la Información*, 30(3). doi: 10.3145/epi.2021.may.09.
- DORA (2012). *San Francisco Declaration on Research Assessment*. Recuperado de: <https://sfdora.org/read/>.
- Feenstra, Ramón A. y Delgado López-Cózar, Emilio (2021). The footprint of a metrics-based research evaluation system on Spanish philosophical scholarship: an analysis of researchers perceptions. *arXiv:2103.11987*.
- Hicks, Diana, Wouters, Paul, Waltman, Ludo, de Rijcke, Sarah y Ràfols, Ismael (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429-431. doi: 10.1038/520429a.
- López-Piñero, Carla y Hicks, Diana (2015). Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. *Research evaluation*, 24(1), 78-89. doi: 10.1093/reseval/rvu030.
- Oviedo-García, M. Ángeles; Casillas-Bueno, José Carlos y González-Rodríguez, M. Rosario (2021). *Análisis bibliométrico e impacto de las editoriales open-access en España*. Madrid: ANECA. Recuperado de: http://www.aneca.es/content/download/16664/202231/file/210930_Openaccess.pdf.
- Weingart, Peter (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62(1), 117-131. doi: 10.1007/s11192-005-0007-7.