

# Ejes de discusión en la evaluación de la ciencia: revisión por pares, bibliometría y pertinencia\*

**Nerina Fernanda Sarthou\*\***

Fecha de recepción: 02 de septiembre de 2015 · Fecha de aceptación: 26 de enero de 2016 · Fecha de modificación: 15 de febrero de 2016

DOI: <http://dx.doi.org/10.7440/res58.2016.06>

**RESUMEN** | Este artículo busca reunir en un sólo documento al conjunto de ejes problemáticos que rodean la definición e implementación de mecanismos y criterios referidos a la evaluación de la ciencia académica. Se revisitan y ordenan los consensos sobre sus defectos y las propuestas esgrimidas para contrarrestarlos a partir de una revisión bibliográfica que gira en torno a tres interrogantes: *quiénes* participan de la evaluación; *cómo* se lleva a cabo la evaluación y *qué* es lo que se evalúa. Por último, se exponen brevemente algunas notas referidas a la literatura específica generada desde América Latina, destacándose el reducido espacio que ocupa esta temática hasta el momento.

**PALABRAS CLAVE** | Bibliometría, pertinencia (Thesaurus); evaluación de la ciencia, revisión por pares (palabras clave de autor).

## Key Points of Discussion in Scientific Research Evaluation: Peer Review, Bibliometrics and Relevance

**ABSTRACT** | This article seeks to bring together in a single document the set of issues surrounding the definition and implementation of mechanisms and criteria relating to the evaluation of scientific research. It revisits and orders the basic agreements regarding their defects and the proposals for countering them based on a bibliographical review revolving around three questions: *who* participates in the assessment; *how* the assessment is carried out; and *what* is to be evaluated. Finally, it briefly presents some notes referring to the specific literature generated from Latin America, highlighting the limited space dedicated to this topic until now.

**KEYWORDS** | Bibliometrics (Thesaurus); research evaluation, peer review, relevance (Author's keywords).

## Eixos de discussão na avaliação da ciência: revisão por avaliadores, bibliometria e pertinência

**RESUMO** | Este artigo procura reunir num só documento o conjunto de eixos problemáticos que rodeiam a definição e implementação de mecanismos e critérios relacionados com a avaliação da ciência académica. Revisitam-se e ordenam os consensos sobre seus defeitos e as propostas apresentadas para combatê-los a partir de uma revisão bibliográfica que gira em torno de três questões: *quem* participa da avaliação; *como* se realiza a avaliação e *o que* se avalia. Por último, expõem-se brevemente algumas anotações relacionadas com a literatura específica gerada a partir da América Latina, destacando o reduzido espaço que ocupa essa temática até o momento.

**PALAVRAS-CHAVE** | Bibliometria, pertinência (Thesaurus); avaliação da ciência, revisão por avaliadores (palavras do autor).

\* Este artículo es producto de la investigación desarrollada como becaria Posdoctoral de Conicet (Argentina) y como docente-investigadora del Centro de Estudios Interdisciplinarios en Problemáticas Internacionales y Locales (Ceipil) de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA), Argentina. Contó con el financiamiento del Conicet, a través de su programa de Becas Posdoctorales.

\*\* Doctora en Ciencia Política por la Universidad Nacional de San Martín, Argentina. Investigadora asistente de Conicet, Argentina. Miembro del Ceipil y profesora de la Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina. Entre sus últimas publicaciones se encuentran: "Los instrumentos de política como enfoque de análisis de los sistemas de pago al mérito. Contribuciones analíticas a partir del caso argentino". *Perfiles Educativos* XXXVII (149): 150-168, 2015, y "El Programa de incentivos a docentes investigadores en Argentina: a dos décadas de su implementación" (en coautoría). *Ciencia, Docencia y Tecnología* 26 (50): 1-34, 2015.

✉ [nfsarthou@yahoo.com.ar](mailto:nfsarthou@yahoo.com.ar)

## Introducción

Si se me permite una metáfora (o quizá una hipérbole), la evaluación es a la ciencia como la democracia a nuestra sociedad: sabemos que es un sistema de gobierno imperfecto, que tiene problemas y que los genera en su aplicación, que es mejorable en su práctica en muchos países y organizaciones, pero lo cierto es que —y quizá es una asunción normativa— no podemos hacer ciencia sin una evaluación de ésta, porque sin ella no sería tal. (Sanz Menéndez 2014, 137)

La evaluación de la ciencia académica<sup>1</sup> es un tema que ha sido objeto de debate en distintos momentos históricos, entre una amplitud de interlocutores y con una diversidad de ejes de discusión. En términos temporales, la práctica de la evaluación ha formado parte de la ciencia académica, al menos desde mediados del siglo XVII, cuando comenzó a constituirse un verdadero sistema de comunicación de los resultados científicos. El momento “fundacional” puede situarse en 1665, cuando se creó la primera revista —o *journal*—, y la Royal Society instauró un sistema por el cual la presentación de trabajos para su publicación en *Philosophical Transactions* debía realizarse con el informe favorable de un miembro de la sociedad (Zuckerman y Merton 1971). Específicamente, dicha práctica fue luego denominada “revisión editorial por pares”<sup>2</sup> (Weller 2001) y se extendió en todo el mundo como procedimiento de evaluación en todas las áreas del conocimiento.

1 Se entiende por ciencia académica un modo de producción de conocimientos “regido por valores, normas, prácticas de producción y de uso, mecanismos y relaciones de intercambio que i) se despliegan entre colegas científicos que actúan en roles profesionales de tales, ii) se comunican y difunden a través de medios públicos (revistas y reuniones especializadas) en los que predominantemente participan colegas profesionales, iii) se emplean como objeto de valoración o intermediario social en el intercambio de recursos (prestigio, fondos de financiamiento, atracción de discípulos, autoridad política en el campo científico, etc.) en sistemas de relaciones sociales con colegas profesionales” (Vaccarezza y Zabala 2002, 40). En contraposición, estos autores mencionan la ciencia o “investigación comercial”, que se desarrolla “en un marco de relaciones sociales i) en donde predominan actores sociales no académicos y ii) tiene como resultado esperado la valoración comercial del conocimiento, postulando que este adquirirá la cualidad de mercancía, por parte del investigador como agente económico o de otro agente de mercado” (Vaccarezza y Zabala 2002, 40). Para conocer el debate que rodea el término de ciencia académica ver: Ziman (2003) y Jiménez-Buedo y Ramos (2009).

2 La revisión editorial por pares es un proceso que, en general, se inicia cuando el autor de un documento lo envía al editor de la revista; el editor o uno de los miembros del Comité Editorial accede al documento, selecciona dos o tres revisores y les envía una copia del manuscrito. A los revisores se les solicita que evalúen el documento y elaboren un dictamen en donde recomienden aceptar sin modificaciones, aceptar con modificaciones o rechazar el manuscrito para su publicación; el editor o los editores luego deciden si aceptan o no las recomendaciones. Los editores pueden enviar una grilla a los revisores para que completen algunas líneas como guía para realizar la evaluación, o simplemente pueden solicitar comentarios.

De esta manera, las primeras instancias de evaluación de la ciencia académica fueron completamente internas a los propios colectivos científicos; la evaluación que se hacía entonces era de tipo informal, dirigida a establecer el “valor de verdad” de los enunciados propuestos, según las concepciones vigentes (Kreimer 2011). La evaluación por pares dejó de estar sólo asociada a las actividades de comunicación de la ciencia para emplearse igualmente junto a la evaluación de los méritos curriculares en instancias competitivas frente a la ocupación de determinado cargo en una organización científica, e incluso para el otorgamiento de premios o distinciones de academias o asociaciones de ciencia (Langfeldt y Kyvik 2011). De este modo, el sistema de revisión por pares o *peer review* nació y se desarrolló como un mecanismo para “servir” a la ciencia permitiendo, a partir de la participación de los propios científicos, y sólo de ellos, definir y resguardar la “calidad científica”.

Ahora bien, a medida que el Estado comenzó a incidir sistemáticamente en el desarrollo del conocimiento científico y tecnológico, luego de la Segunda Guerra Mundial, la revisión por pares se convirtió en el procedimiento empleado para la distribución de fondos públicos entre instituciones e investigadores. Previamente, el Estado, e incluso fundaciones tales como Rockefeller, solían recurrir al consejo de destacados científicos para definir estrategias; con el surgimiento de una política pública para el área de la ciencia y la tecnología, aquella práctica se amplió hasta constituirse en un método empleado de manera generalizada. De acuerdo con Rip (1994), fueron los propios políticos (a veces ex-científicos) quienes introdujeron este método al advertir la necesidad de emplear un medio considerado legítimo por los científicos, para contrarrestar la oposición por parte de ellos a la injerencia estatal en el desarrollo de la ciencia. De este modo, la revisión por pares se introdujo con el objeto de ganar la buena voluntad de los científicos, antes que por sus cualidades para alcanzar los fines de la política pública en el área.

Con el transcurso de los años, la revisión por pares extendió su campo de acción: pasó a utilizarse para evaluar investigadores individuales, grupos de investigadores, instituciones enteras, proyectos de investigación, conjuntos de proyectos agrupados en programas, políticas de promoción de la investigación, o incluso todo el sistema de investigación en su conjunto (Molas-Galart 2012). De este modo, la práctica de la evaluación de la ciencia dejó de constituir la piedra angular de un “sistema de control de calidad”, al ser adoptada además como instancia clave de un sistema de asignación de recursos (Sanz Menéndez 2014). Esta transferencia produjo la emergencia de diversas tensiones y polémicas.

El objetivo de este trabajo es reunir en un sólo documento al conjunto de ejes problemáticos que rodean la definición e implementación de mecanismos y criterios

referidos a la evaluación de la ciencia académica. Se revisitan y ordenan los consensos sobre sus defectos y las propuestas esgrimidas para contrarrestarlos, a partir de una revisión bibliográfica que gira en torno a tres interrogantes: *quiénes* participan de la evaluación; *cómo* se lleva a cabo la evaluación y *qué* es lo que se evalúa.

Para ello, el artículo se organiza de la siguiente manera. En primer lugar se introducen las principales críticas que ha recibido la práctica de la evaluación referida al mecanismo de la revisión por pares, es decir, respecto a *quiénes* participan. En segundo lugar se exhiben los cuestionamientos al empleo de indicadores cuantitativos asociados a las publicaciones científicas, es decir, a *cómo* se evalúa a través de estas herramientas. En tercer lugar se presentan los debates que ha generado el *qué* se evalúa, es decir, las discusiones en torno a “calidad”, “pertinencia” y “relevancia social”. En cuarto y último lugar se exponen brevemente algunas notas referidas a la literatura específica generada desde América Latina, destacándose el reducido espacio que ocupa esta temática hasta el momento.

### **El problema es quién evalúa: los cuestionamientos a la revisión por pares**

Uno de los problemas esgrimidos en el momento de debatir sobre la evaluación de la ciencia nos remite a los participantes en las instancias de evaluación. Existe cierto consenso en la literatura sobre los “defectos” o “problemas” del sistema de revisión por pares o *peer review* (Campanario 2002; Cole, Cole y Simon 1981; Davyt y Velho 2000; Nicholas *et al.* 2015; Scott 2007). La definición de los “pares” suele asociarse con ciertos rasgos principales: su carácter de expertos en la materia que evalúan, el anonimato del juicio que les permite expresar libremente y de manera fundamentada su opinión, la adecuada distancia respecto de la persona e institución en que labora el autor de la propuesta, así como un *ethos* de responsabilidad, compromiso y deseo de avance del conocimiento (Grediaga 2011). Sin embargo, el significado de *pares* se puede ver afectado por modificaciones introducidas al procedimiento con la intención de mejorar su implementación.

En este sentido, existen instancias en las que los pares no son expertos en la materia, por ejemplo, cuando se conforman comisiones de evaluación multidisciplinarias o cuando además de pares se convoca a los potenciales *usuarios* de los resultados de las investigaciones. También el anonimato puede verse alterado cuando se implementa el sistema de un “solo ciego” —el evaluador conoce al autor de la propuesta— o cuando la modalidad es “abierta”, es decir, tanto el evaluador como el evaluado conocen la identidad del otro (Langfeldt y Kyvik 2011). Las críticas al procedimiento de revisión por pares son variadas y han originado una profusa literatura; incluso los propios científicos, en su doble rol de evaluadores

y evaluados, han manifestado sus críticas al proceso (Mulligan, Louise y Ellen 2013).

Como se mencionó, la revisión por pares ha estado asociada a la evaluación de la ciencia desde tiempos remotos; de allí la extensión de su empleo en las diversas áreas del conocimiento y en los distintos sistemas nacionales de investigación. Sin embargo, pronto comenzaron los cuestionamientos a este procedimiento desde el ámbito de los organismos públicos. Ya a finales de los años setenta, en Estados Unidos se evidenció cierta preocupación sobre la conveniencia de utilizar el sistema de revisión por pares como mecanismo de evaluación en la asignación de fondos de investigación por la National Science Foundation. Se encontraron varios problemas mediante un estudio exhaustivo del funcionamiento de la revisión por pares por medio de la realización de entrevistas en profundidad y del empleo de herramientas estadísticas con base en 1.200 propuestas presentadas (Cole, Cole y Simon 1981). Estos autores formaron parte de un Comité de Ciencia y Política Pública y fueron especialmente contratados por la organización para evaluar la efectividad del *peer review*.

En principio, cabe destacar que las mismas críticas hacia la revisión editorial por pares se han trasladado al mecanismo de revisión por pares asociado a la política pública científica y tecnológica, es decir, a las instancias de evaluación para la adjudicación de financiamiento, de becas, de posiciones de trabajo, o a otras instancias competitivas dentro del mundo académico. De allí que se hayan mencionado como problemas de este sistema su lentitud, el favoritismo casi ineludible dentro de comunidades científicas pequeñas (Bianco, Grass y Sutz 2014), la falta de transparencia y la incierta calidad (Chubin y Hackett 1990; Nicholas *et al.* 2015).

Asimismo, según Campanario (2002), existen dos defectos centrales: la escasa fiabilidad —*reliability*— y la dudosa validez —*validity*—. Por un lado, es escasa la fiabilidad del procedimiento de evaluación por pares —entendida como la consistencia de juicios emitidos por un evaluador determinado sobre un objeto de evaluación en ocasiones sucesivas, o la consistencia de los juicios emitidos por diversos evaluadores sobre un mismo objeto—, ya que se ha comprobado que los expertos pueden coincidir al aceptar o rechazar una propuesta, pero lo hacen, a veces, por razones distintas e incluso por motivos contradictorios. Bianco, Grass y Sutz (2014) sugieren que una eventual discrepancia de juicios entre evaluadores puede tener una diversidad de orígenes, “aun en la misma especialidad cognitiva, existen diferentes tradiciones académicas, opciones metodológicas preferidas por diferentes colectivos de investigadores, y contextos institucionales y geográficos diversos que moldean las prácticas de investigación de manera diferente” (Bianco Grass y Sutz 2014, 226).

Por el otro, la validez, esto es, la capacidad del sistema para seleccionar las “mejores” propuestas, es al menos dudosa. El origen de esta crítica reside en la creencia de que los expertos poseen cierta resistencia a aceptar teorías y descubrimientos nuevos, rasgo denominado “conservadurismo” (Campanario 2002). El problema de la validez se agrava cuando las propuestas poseen un carácter interdisciplinario o refieren a áreas emergentes del conocimiento en las cuales la disponibilidad de expertos es escasa, y mayor el desafío de evaluar adecuadamente la calidad (Lamont 2009).

En la misma línea, Lee *et al.* (2013) han estudiado los diferentes sesgos que puede generar la revisión por pares, como producto de afectar la imparcialidad de las evaluaciones. Estos autores definen la noción de imparcialidad como la capacidad de cualquier evaluador para interpretar y aplicar criterios de evaluación de la misma manera. Es decir, los evaluadores imparciales llegan a evaluaciones idénticas de una presentación en relación con los criterios evaluativos porque los interpretan de la misma manera. La imparcialidad asegura que las evaluaciones son independientes de los sesgos teóricos y de la tolerancia al riesgo del revisor, y de las identidades sociales del autor y del revisor.

Si bien el modelo de revisión por pares denominado “doble ciego”, es decir, anonimato del revisor y del autor, contribuye de manera significativa a la imparcialidad, existen diversas instancias que emplean el modelo de un “solo ciego”. En algunos sistemas nacionales de investigación, las presentaciones para adquirir subsidios, el ingreso o la promoción dentro de una carrera de investigación o la postulación a una beca de formación adoptan este tipo de revisión por pares, es decir, el evaluador conoce la identidad del autor de la propuesta pero éste desconoce la del evaluador; de allí la importancia de la imparcialidad para reducir los sesgos relativos a las identidades sociales del autor.

Estos autores identifican en la literatura cuatro tipos de factores que afectan la imparcialidad de una evaluación por pares (Lee *et al.* 2013): a) aquellos vinculados con un error en la evaluación de la *calidad real* de una presentación, b) aquellos relativos a las características sociales del autor, c) aquellos referidos a las características sociales del revisor y/o d) aquellos relacionados con el contenido crítico de la presentación. Resulta relevante destacar los problemas de sesgo relativo a las características sociales del autor: sesgo basado en el prestigio del autor —estudiado por Robert Merton y denominado “Efecto Mateo”—, sesgo relacionado con la nacionalidad del autor, sesgo vinculado al idioma empleado por el autor y sesgo relativo al género del autor.

Por su parte, Langfeldt (2006) ha revisado los sesgos en los procesos de revisión por pares vinculados a los conflictos de intereses. Existen amplias variaciones entre países y organismos de fomento de la inves-

tigación científica respecto a qué tipo de relaciones descalifican a un evaluador para la participación en la asignación de subvenciones de proyectos. En algunos consejos de investigación, los pares evaluadores no pueden evaluar propuestas provenientes de su propia institución; en otros, no se les permite solicitar subvenciones a quienes participan como evaluadores en la misma convocatoria. Hasta cierto punto, las variaciones en la definición de lo que constituye conflicto de interés parecen relacionarse con diferentes contextos de evaluación y distintas magnitudes de casos problemáticos que necesitan regulación, es decir que, en esquemas altamente competitivos, la regulación de los conflictos de interés suele ser más estricta.

Basados en los problemas detectados, organismos e instituciones científicas han adoptado variantes de la revisión por pares, por ejemplo, alterando el requisito del anonimato o incorporando a otros participantes. También, con el propósito de aumentar la consistencia en el juicio de los expertos, algunos organismos suelen proporcionar hojas de instrucciones más o menos detalladas o formularios de evaluación estructurados que hagan referencia a los puntos esenciales que deben determinar la decisión final de aceptar o rechazar una propuesta. También se ha aumentado la transparencia, explicando los criterios que se utilizan, evitando el conflicto de intereses y siguiendo normas reconocidas, y se ha incrementado la publicidad sobre la conformación de las comisiones evaluadoras —aunque no necesariamente del evaluador específico de cada propuesta— y las reglas de funcionamiento que deben cumplirse (Musselin 2013).

Desde otro punto de vista, algunos autores han sugerido modificar el sistema de revisión por pares ampliando o “democratizando” (Derrick y Pavone 2013) la participación hacia los “impares” (Sutz 2014) o “comunidad extracientífica” (Frederiksen, Hansson y Wenneberg 2003). De acuerdo con Sutz (2014), los *impares* son, en principio, aquellos que en diversas esferas —política, productiva, social, cultural— buscan incorporar los resultados obtenidos a su reflexión y a su accionar. La razón de estos actores para participar en la evaluación reside en su capacidad de evaluar la relevancia de la propuesta o de los resultados de acuerdo con variadas circunstancias, que incluyen la posibilidad que tengan de usarlos, la utilidad más o menos directa que perciban que tienen para responder a sus demandas y problemas, el nivel de participación que hayan tenido en la definición de los problemas por investigar y su involucramiento durante el proceso de investigación.

Frederiksen, Hansson y Wenneberg (2003) también advierten sobre la necesidad de incorporar nuevos actores y criterios en las instancias de evaluación de la investigación financiada con fondos públicos. Subrayan que la sociedad ya no confía en un enfoque de evaluación unidimensional llevado a cabo en un sistema aislado

con criterios de calidad intracientífica. Estos autores proponen específicamente la noción de *ágora*, entendida como un común denominador de personas, lugares, redes e instituciones con interés o influencia en las luchas por la autoridad de las afirmaciones científicas.

Una propuesta distinta es elaborada por Derrick y Pavone (2013), quienes llaman la atención sobre la necesidad de volver “más democrática” la revisión por pares ampliando la participación a actores por fuera de la comunidad científica. Para ello, estos autores proponen hacer uso de los indicadores bibliométricos: estos facilitarían la incorporación de una gama más amplia de actores, logrando un modelo más cercano al ideal de evaluación de “partes interesadas”. Según estos autores, la ventaja del uso de indicadores bibliométricos reside no sólo en su transparencia y facilidad de comprensión sino también en la capacidad de éstos de traducirle al público no científico la información sobre los resultados de una investigación, dando lugar, así, a una participación activa e igualitaria en la comunidad científica dentro de los procesos de evaluación. No obstante, afirman que su proposición no es reemplazar la revisión por pares por el uso de indicadores bibliométricos, sino buscar que las ventajas de cada método contrarresten los defectos del otro promoviendo mejoras científicas y sociales en los procesos de evaluación a partir de su combinación.

Aquí aparece el eje problemático que gira en torno al empleo de indicadores bibliométricos para llevar a cabo los procesos de evaluación, es decir, emerge el dilema sobre a través de qué procedimientos conducir la valoración de una propuesta de investigación, de los avances de una institución o de una trayectoria individual en el ámbito científico. Es en este punto donde aparecen otros múltiples problemas.

### **El problema es cómo se evalúa: las limitaciones del empleo de métodos cuantitativos**

Durante la segunda mitad del siglo XX, gran parte de las evaluaciones tanto institucionales como individuales comenzaron, paulatinamente, a priorizar los *papers* sobre otras formas de producción de conocimientos (Kreimer 2015). A la par de este proceso, los indicadores asociados al conteo de *papers* publicados en determinadas revistas científicas fueron instalándose como únicas herramientas capaces de hacer posible una evaluación rigurosa y válida. La creciente necesidad de las distintas entidades gubernamentales de contar con criterios uniformes y objetivos como fuente de información a la cual recurrir en el momento de tomar decisiones técnicas, administrativas y políticas ha generado un mayor interés en la utilización de indicadores bibliométricos, dado que ofrecen un método estándar para la evaluación de la calidad y eficacia de las actividades científicas (Rueda-Clausen, Villa-Roel y Rueda-Clausen 2005).

Se considera que una de las primeras aplicaciones de indicadores bibliométricos tuvo lugar en 1927, cuando Gross y Gross publicaron su estudio sobre las referencias incluidas en artículos de revistas de Química indizadas en el *Journal of the American Chemical Society*, entendiendo que aquello podía ser utilizado como indicador para evaluar y comparar la calidad de la producción científica. No obstante, fue recién a partir de los años sesenta que esta tendencia adquirió su nombre actual, cuando se dio a conocer la publicación del Science Citation Index (SCI), producida por el Institute for Scientific Information (ISI) de la compañía Thomson Scientific, localizada en Toronto (Canadá). En la evaluación de los antecedentes de cada propuesta se reemplazó el conteo del número bruto de artículos publicados por la medición del *impacto* de dicha producción, medido por el número de citas que han merecido los trabajos de un autor o un conjunto de autores (Kreimer 2015). Así, con la creación del SCI apareció por primera vez el “Factor de Impacto” como indicador de la calidad de los artículos y de las revistas científicas.

Si bien existen tres indicadores bibliométricos centrales<sup>3</sup> —el Factor de Impacto, el Índice de Inmediatez<sup>4</sup> y la Vida Media de Citación<sup>5</sup>—, el más difundido es el primero. El Factor de Impacto de una revista ha sido definido como la medida del tamaño de la curva de citas durante los dos y tres años posteriores a la aparición de la revista. Se calcula dividiendo el número de citas que los artículos de una revista reciben en dos años por el número de artículos publicados en esos mismos años<sup>6</sup> (Amin y Mabe 2003). Tiene como propósito evaluar y comparar la importancia de una revista frente a otras de su mismo campo.

3 También existen indicadores más complejos como el popularizado “índice h de Hirsch” (índice h), que consiste en ordenar las publicaciones de un autor a partir del número de citas recibidas en orden descendente, numerarlas e identificar el punto en el cual el número de orden coincide con el número de citas recibidas por una publicación (Kreimer 2015).

4 El Índice de Inmediatez da la medida de la asimetría de la curva, es decir, el grado en que el pico de la curva se encuentra cerca del origen del gráfico. Se calcula dividiendo las citas que recibe una revista en el año en curso por el número de artículos que publica en ese año, es decir, el Índice de Inmediatez 2014 es el número medio de citas en 2014 de artículos publicados en 2014. El número que resulta indica la rapidez con la cual los artículos publicados en la revista consiguen ser citados desde el momento de publicación.

5 La Vida Media de Citación es una medida de la tasa de disminución de la curva de la citación. Es la cantidad de años que el número de citas de un artículo tarda en declinar en un 50% de su valor inicial. Es una medida de cuánto tiempo un artículo continúa siendo citado después de su publicación.

6 Así, por ejemplo, el Factor de Impacto 2014 de una revista está constituido por las citas en 2014 de artículos publicados entre 2012 y 2013, divididos por el número de artículos publicados en 2012 y 2013. El número que resulta indica la cantidad de citas anuales que en promedio recibirá un artículo a partir de los dos años después de ser publicado.

Aunque se considera que el Factor de Impacto es un método práctico para la estandarización de la evaluación de la calidad de las revistas científicas, de los artículos y, de manera paralela, de los investigadores y grupos de investigación que publican en ellas, existe una corriente creciente de autores que no lo consideran adecuado, por dos razones centrales: posee limitaciones en su elaboración y sesgos en sus resultados, y no es conveniente para evaluar determinadas áreas del conocimiento.

De acuerdo con Amin y Mabe (2003), el valor del Factor de Impacto se ve afectado por factores tanto sociológicos como estadísticos. Los factores sociológicos incluyen el área temática de la revista, el tipo de artículos que publica —cortos o completos, originales o revisiones de literatura— y el número medio de autores por trabajo —cuestión relacionada con la disciplina—. Los factores estadísticos incluyen el tamaño de la revista —número de artículos que publica anualmente<sup>7</sup>— y el tamaño de la ventana de medición de la citación (número de años). Seglen (1997) adiciona otros factores que generan sesgos en el resultado del Factor de Impacto: las limitaciones de las bases de datos empleadas —basadas en la preferencia por los artículos o *papers*, por revistas de determinadas áreas de conocimiento y por el idioma inglés— y las dinámicas de citación propias de cada disciplina.

La distorsión provocada por el lenguaje ha sido ampliamente abordada. Se ha comprobado de manera empírica que los sesgos lingüísticos desempeñan un papel importante en el momento de la comparación y de la evaluación de los sistemas científicos nacionales; si el artículo está escrito o publicado en una revista editada en un idioma distinto al inglés tendrá un impacto considerablemente más bajo en términos de citación (Van Leeuwen *et al.* 2001). Este fenómeno es aún más pronunciado en el caso de las Ciencias Sociales (Ortiz 2009) y las Ingenierías (Mutzl, Bornmann y Daniel 2013).

Profundizando en el análisis de referencias, Bornmann (2011) identifica en la literatura diversos factores que afectan la cantidad de citas que puede recibir un artículo, y con ello, los sesgos del Factor de Impacto. Entre dichos factores enumera: a) el período escogido, ya que cuanto más se cita una publicación, es más probable que sea citada en el futuro; b) la disciplina, debido a que se ha demostrado que la dinámica de la citación varía entre disciplinas e incluso entre especialidades dentro de una misma disciplina; c) la revista seleccionada, ya que se ha comprobado que la cantidad de citas de un artículo puede depender de la frecuencia de aparición de la revista, así como de su accesibilidad, visibilidad, internacionalidad, impacto, calidad o prestigio; d) las características del artículo, es decir, que el número de citas puede

verse influido por el tipo de artículo —de investigación, de revisión, editoriales, reseñas, capítulos de libros o libros—; e) el idioma en que está escrito, y f) la accesibilidad de la base de datos en la cual se aloja la revista.

Por su parte, también existen fuertes críticas que advierten sobre los efectos de la generalidad del empleo de estos indicadores en las etapas de evaluación de la ciencia académica. Guédon llama la atención y afirma que Thomson Scientific impulsó, en realidad, una doble agenda: por un lado, proporciona medios para analizar las actividades científicas desde una gran variedad de puntos de vista; por el otro, al mismo tiempo, define los términos de la evaluación en sí. El plan de negocios de esta empresa —sentencia el autor— “depende en gran medida de la capacidad de clasificar todas las facetas de la ciencia en sus propios términos de apropiación” (Guédon 2011, 166).

Frente a las críticas que ha recibido la utilización de indicadores bibliométricos, hacia principios del siglo XXI surgió una corriente de estudios que propuso como herramienta de evaluación de la investigación el empleo del currículum vitae (CV). En cuanto documento único que resume todas las actividades profesionales de los investigadores y que registra la evolución actualizada de su trayectoria a lo largo del tiempo, se considera el CV como una valiosa fuente de información longitudinal de difícil acceso por otros medios (Cañibano, Otamendi y Solís 2010). A su vez, es uno de los pocos artefactos científicos casi universales en su disponibilidad, y casi estándar en su significado (Cañibano y Bozeman 2009).

Fueron los miembros del proyecto Research Value Mapping Programme, en la actualidad ubicado en la Universidad del estado de Arizona, quienes emplearon esta metodología por primera vez. Producto de la labor de este equipo de investigación, Bozeman, Dietz y Gaughan (2001) propusieron un cambio en el paradigma de la evaluación de las actividades de investigación realizadas por individuos y organizaciones. Este cambio consiste en pasar del “Paradigma de resultados” al “Paradigma de la capacidad”. Este nuevo enfoque desplaza la evaluación centrándola en el capital humano científico y tecnológico, reagrupando las capacidades de las personas, las instituciones y otros colectivos para contribuir al conocimiento científico y tecnológico. Desde este punto de vista, es la capacidad productiva en el largo plazo lo que importa, no los productos concretos del conocimiento en momentos particulares (Cañibano y Bozeman 2009). Si bien existen organismos e instancias que utilizan el CV como herramienta de evaluación de la investigación, aún quedan por resolver cuestiones como la diversidad de formatos de captación del CV, que conduce a que las bases de datos que se generan tengan escasa o ninguna integración y comunicación dentro un mismo país, duplicando la tarea de los investigadores.

La propuesta mencionada se relaciona con el tercer y último eje de discusión: qué evaluar. El empleo de indica-

7 Los autores (Amin y Mabe 2003) mencionan como referencia promedio que los “pequeños títulos” son aquellos que publican menos de 35 artículos por año, mientras que los “grandes títulos” publican hasta 150 artículos por año.

dores bibliométricos en el marco de la revisión por pares ha sido criticado por limitar la noción sobre los productos del conocimiento y, con ello, sobre el concepto de *calidad* de una propuesta. Cuando se contempla la diversidad de usos y de comunidades de interés involucradas directa o indirectamente en los resultados de una investigación emergen nuevas consideraciones.

### El problema es qué se evalúa: calidad y pertinencia de la investigación

Otro de los puntos discutidos acerca de los procesos de evaluación de la ciencia es qué se evalúa. En primer lugar, se presentan problemas cuando el qué de la evaluación es reducido a uno de los resultados posibles de una investigación o de una carrera científica: el *paper* o artículo científico.

Kreimer (2015) ha demostrado los problemas que conlleva reducir la evaluación de la investigación a un único resultado materializado en un artículo publicado. Este autor subraya las distintas cuestiones que un *paper* oculta: un *paper* muestra el éxito, pero también esconde el fracaso de aquello no publicado; un *paper* oculta el conocimiento tácito y el papel que los autores desempeñan en un campo científico de relaciones sociales; un *paper* esconde el interés —o la necesidad— del autor —o de los autores— por legitimarse. De este modo, este autor afirma que aquellas “evaluaciones realizadas en base a los *papers* están más marcadas por los imperativos burocráticos que por un objetivo sustantivo de evaluar las prácticas científicas, los procesos y el sentido de la producción de conocimiento” (Kreimer 2015, 45). Sin embargo, ni siquiera en términos burocráticos puede decirse que la evaluación basada en la publicación resulta un camino adecuado, debido a las críticas ya mencionadas sobre el empleo de indicadores bibliométricos.

Bajo esta inquietud, Mutzl, Bornmann y Daniel (2013) clasificaron las disciplinas científicas según el mayor o menor empleo de los distintos formatos de publicación de los resultados de investigación. En términos de evaluación, estos autores corroboraron que “Artes y Humanidades” realmente debe ser tratada como un área independiente y relativamente uniforme, y que, en lugar de contar sólo artículos de revistas y sus citas, es importante incluir también monografías y antologías. Asimismo, estos autores rechazan la creencia generalizada respecto a la jerarquización de las disciplinas basada en la cantidad de artículos en revistas con referato<sup>8</sup>. Al respecto, demuestran que no sólo los proyectos de “Ciencias Sociales/Artes y Humanidades” poseen más libros y artículos sin referato, sino que también hay una alta

proporción de este tipo de resultados en los proyectos provenientes del área de las “Ciencias Aplicadas”, es decir, que aquella jerarquía de disciplinas debe ser reformulada bajo nuevos parámetros.

En esta misma línea, Hicks (2004) identifica cuatro formatos para dar a conocer los resultados de una investigación en Ciencias Sociales, cada uno enfocado en un fin específico: los artículos en revistas internacionales, los libros, los artículos en revistas académicas nacionales y las publicaciones no académicas. De este modo, la autora subraya que los resultados de un sistema de evaluación que emplee indicadores bibliométricos no serán los mismos que los resultados de una evaluación que incluya los cuatro tipos de publicaciones señaladas. Más aún, la autora advierte el riesgo que representa que los académicos traten de obtener mejores evaluaciones abandonando los otros tres formatos, en favor de los artículos en revistas internacionales; la Ciencia Social resultante sería diferente de las Ciencias Sociales de los cuatro formatos de publicación.

En línea con estas afirmaciones, y como resultado de su trabajo, Bianco, Gras y Sutz afirman que “no es posible utilizar un mismo sistema de juicios para ponderar y valorar la diversidad de objetivos de investigación, marcos conceptuales de referencia, estrategias de investigación, tipos de resultados y sus medios de difusión” (2014, 220). En este sentido, enfatizan la necesidad de que los sistemas de evaluación contemplen las diversas modalidades en las que se expresa la identidad académica de distintas orientaciones disciplinarias cuando el propósito mayor es estimular la calidad en todas ellas.

En segundo lugar, al continuar con el interrogante referido a qué evaluar aparece cierto consenso respecto a la necesidad de evaluar la *calidad* y la *pertinencia* de la ciencia; no obstante, la definición de ambos conceptos varía significativamente entre audiencias y usuarios diversos del conocimiento producido. Ya sea en el marco de procesos de evaluación para la toma de decisiones sobre la distribución de fondos públicos para el desarrollo de una investigación o para la formación académica, o en el caso de ascensos en categorías de investigación o puestos de trabajo en organismos dedicados a la producción científica y tecnológica, la cuestión de qué evaluar ha sido objeto recurrente de debate.

Las dificultades giran en torno a la diferenciación de dos elementos: por un lado, la *calidad* de la propuesta, y por otro, su *relevancia* en relación con las prioridades de investigación establecidas. Cuando el propósito es financiar investigación que sea relevante para la sociedad, hay una necesidad de tener en cuenta un conjunto más amplio de criterios. En este sentido, ha tenido lugar en los diversos sistemas de evaluación nacionales un proceso mediante el cual al concepto de *calidad* se le ha adicionado la noción de *pertinencia* o *relevancia*.

8 Se denominan revistas con referato a aquellas que poseen un sistema de evaluación de los artículos con pedido de publicación. En general, la noción de referato es utilizada como sinónimo de revisión editorial por pares (ver la nota 2).

De acuerdo con Naidorf, Giordana y Horn (2007), en un primer momento, la noción de *pertinencia* asociada a la evaluación fue introducida por los documentos especializados de la Unesco. Desde esta perspectiva, la pertinencia debe contrapesar dos aspectos: “por una parte, la tendencia al auto-encierro de las instituciones universitarias y, por otra, la tendencia de una agenda académica definida desde parámetros puramente disciplinares y, por ende, desconectados —se supone— de las realidades y los contextos sociales” (Naidorf, Giordana y Horn 2007, 24).

Más tarde, en particular durante la primera década del siglo XXI, Naidorf y Perrotta (2015) sostienen que comenzó a asociarse la pertinencia con el establecimiento de prioridades y de temas prioritarios, es decir que comenzaron a promoverse áreas del conocimiento, temas o problemas en un sentido determinado. En este marco, las autoras consideran que diversos organismos de gestión pública dedicados a la ciencia académica han establecido como prioritarias áreas consideradas estratégicas para la solución de problemas sociales identificados como acuciantes, con más o menos éxito, y con más o menos oposición, por parte de las comunidades científicas. De modo que la incorporación de la noción de *pertinencia* como criterio de evaluación de la ciencia académica se convierte así en una vía para vincular la producción de conocimiento con las necesidades sociales.

En esta línea, Braun (1998) señala que los organismos de gestión pública que buscan financiar la ciencia académica se encuentran con la tarea de reconciliar las demandas y los intereses de actores políticos y científicos. Para ello definen criterios de evaluación propios e instan a los comités de pares a aplicarlos en el momento de seleccionar las propuestas financiables. Específicamente, este autor identifica una estructura de decisión de los organismos de financiamiento dividida en *arenas*: una arena de selección, una arena política y una arena de control. Éstas se presentan de manera sucesiva durante el proceso decisorio que conduce al otorgamiento de una beca o un subsidio. Cabe subrayar que la propia noción de *arena* intenta dar cuenta de la permeabilidad de sus bordes, es decir, que si bien cada *arena* tiene una misión central, sus participantes no quedan impedidos de realizar otras observaciones durante el proceso de evaluación.

En primera instancia, en la arena de selección intervienen los científicos como evaluadores anónimos, ya sea actuando de manera individual o integrando un comité evaluador; ellos son los encargados de decidir cuáles proyectos poseen calidad científica, aunque en el marco de lo establecido por el organismo por el cual han sido convocados. En una segunda instancia, en la arena política actúan comisiones científicas junto a comisiones adicionales (*ad hoc*). La tarea de las comisiones adicionales es no perder de vista los objetivos de política científica del ministerio, agencia o ente financiador. En

contraste con la arena de selección, la decisión sobre la financiación cumple un papel significativo en la fijación de prioridades. Por último, en la arena de control actúa una comisión política. Por medio de esta participación, los actores políticos tienen la oportunidad de dirigir directamente sus demandas a los encargados de gestionar la investigación y a los científicos.

Aquí emerge la cuestión de los usos de la evaluación: de acuerdo con el uso que tendrán los resultados de la evaluación, serán quiénes participan y qué se evalúa (Molas-Gallart 2012). Cuando se involucran otros actores en la evaluación de la investigación y se tiene en cuenta la aplicación de los resultados científicos, el rol potencial de la evaluación se expande incluyendo otros propósitos como la legitimación de políticas e iniciativas pasadas, la provisión de un mecanismo de transparencia para la legislatura y el público, el apoyo a la formulación y el diseño de políticas, la distribución de recursos, el suministro de resultados para implementar mejoras en el funcionamiento, entre otros.

Como puede advertirse, en las instancias de evaluación de la ciencia académica, la búsqueda de pertinencia se vincula a procesos de revisión por pares más abiertos, teniendo en cuenta un conjunto más amplio de puntos de vista de los responsables políticos y profesionales que puedan participar sustantivamente (Scott 2007), así como otros actores y usos de los resultados que arroje el proceso de evaluación.

Desde otro punto de vista, de acuerdo con Sutz (2014), la pertinencia o relevancia de la investigación se expresa en la medida en que haya demanda. Esta puede ser inmediata, cuando se hace explícita y se formula *ex ante*, o diferida, cuando es difusa y se expresa *ex post*. La clave del asunto reside en que “esta cuestión temporal hace que la apreciación acerca de qué tipo de investigación es relevante para un país se constituya en un asunto delicado que puede dar lugar a gruesos errores de apreciación” (Sutz 2014, 67). De este modo, la *relevancia* es un concepto asociado a una construcción social. En este sentido, la autora subraya que cuando se habla de demanda de conocimiento se está haciendo referencia a una forma particular de demanda que tiene el poder de otorgar relevancia: la demanda solvente, ya sea de empresas, ya sea del Estado. Lo que sucede es que la demanda solvente no es la única posible; está también la demanda de grupos aun minoritarios y con escaso peso político, demanda de conocimientos que permitan acercar soluciones a problemas complejos.

Como puede advertirse, la noción de *pertinencia* o *relevancia* de la investigación requiere una compleja elaboración al estar en relación con los objetivos definidos por los organismos de ciencia y tecnología, pero en el marco de determinadas condiciones políticas, económicas, sociales y culturales de una nación.



## Breves notas sobre la evaluación de la ciencia académica en América Latina

Si bien ninguno de los problemas previamente mencionados ha sido ajeno a la reflexión de los distintos organismos públicos de fomento de la producción de conocimientos científicos y tecnológicos en la región, ni a la reflexión de la comunidad científica latinoamericana, cabe destacar una marcada relegación de la temática desde los Estudios de Política Científica.

La cuestión de quiénes participan en las instancias de evaluación ha sido explorada en trabajos referidos a la implementación de la revisión por pares en la educación superior (Marquina 2006), en el marco de las revistas científicas (Davyt y Velho 2000; Grediaga 2011) y en el caso de políticas específicas de promoción de la investigación en las universidades (Bianco, Gras y Sutz 2014; Sutz 2014).

Respecto a la situación de las revistas latinoamericanas existe una copiosa literatura proveniente de la Bibliotecología (Krauskopf y Vera 1995; Miguel 2011; Ríos Gómez y Herrero Solana 2005; Romanos 2008), complementada por algunos trabajos que vinculan el uso de los indicadores bibliométricos y el diseño de la política científica (De la Vega 2003; Gómez 2004; Mugnaini 2006). En el último lustro han cobrado fuerza dos líneas de interpretación con visiones sociopolíticas concretas vinculadas a una “inserción subordinada” (Kreimer 2011 y 2015) o “periférica” dentro del “Sistema Académico Mundial” (Beigel 2015; Guédon 2011; Vessuri, Guédon y Cetto 2014) del conocimiento producido en los países de América Latina.

Respecto a la discusión sobre qué evaluar, ésta se ha organizado en torno a dos ejes. Por un lado se encuentran los trabajos que enfatizan en el *paper* (Kreimer 1998), y de allí analizan las limitaciones de una evaluación restringida a este único resultado científico, en especial en el caso de países latinoamericanos, a los que se considera situados en la “periferia de la ciencia” (Beigel 2013; Kreimer 2011 y 2015). Por el otro se hallan aquellos que exploran las nociones de *calidad* y *pertinencia* incorporadas a procesos de evaluación de la investigación pública (Lattuada 2014; Naidorf y Perrotta 2015; Naidorf, Giordana y Horn 2007; Sutz 2014).

Hasta aquí se ha explorado y organizado la literatura sobre la evaluación de la ciencia académica en torno a los acuerdos sobre sus fallas y las consiguientes propuestas para revertirlas, con base en tres ejes: *quiénes* participan en la evaluación, *cómo* ésta es llevada a cabo y *qué* es lo que se evalúa. Luego se señaló de manera sintética la bibliografía generada por académicos latinoamericanos referida a aquellos mismos ejes. A continuación se mencionan algunos breves comentarios.

## Conclusiones

Luego del recorrido bibliográfico construido surge, en principio, un panorama depurado respecto a cuáles son los dilemas que giran en torno a la evaluación de la ciencia académica, al menos desde la perspectiva de los propios científicos que han volcado ríos de tinta al estudiar el tema. Dicha perspectiva no es la de cualquier actor, sino la de un actor fundamental para la práctica de la evaluación. Es el propio científico quien actúa en sus múltiples roles de autor de una propuesta, de par evaluador, de director o presidente de un organismo científico, de editor de una revista, que tiene, así, la capacidad de alterar o mantener las implicancias que posee cada proceso evaluativo.

Emerge también una mayor claridad respecto a cuáles dificultades se quieren abordar. Al contar con una visión más nítida de las problemáticas que rodean la evaluación de la investigación puede generarse una jerarquía entre ellas con base en los objetivos definidos en un momento dado, y a partir de allí tomar decisiones. Así, los defectos de los procedimientos o métodos de evaluación pueden enfrentarse de a uno por vez pero sabiendo que aún restan o pueden emerger otros.

Se desprende, asimismo, la necesidad de reforzar el rol político de la evaluación de la ciencia académica, en el sentido de vincular los propósitos de una institución con el lugar ocupado por la evaluación en el proceso de alcanzar dichos objetivos. Debe entenderse la evaluación, y también los criterios, los procedimientos y los participantes, como elementos centrales de la política científica y tecnológica, y no como instancias auxiliares. En este sentido, los ejes de la discusión en torno a la evaluación pasan a ser quiénes, cómo y qué *debería* evaluarse según determinados objetivos de política definidos por cierto actor.

Retomando el epígrafe de este artículo, si bien la evaluación de la ciencia es un sistema imperfecto, que tiene y que genera problemas en su aplicación y es mejorable en su práctica en muchos países y organizaciones, no podemos hacer ciencia sin la evaluación de ésta, aunque no sólo porque la ciencia sin evaluación no sería tal, *sino por necesidad de la política*.

El avance del Estado sobre la ciencia hizo de la evaluación un mecanismo de diálogo entre políticos y científicos. Algunos científicos, aquellos con mayor capacidad de entendimiento con los políticos, pasaron a formar parte de un grupo selecto y reducido —los evaluadores— que se encargaría de interpretar las demandas en lenguaje político y traducirlas y comunicarlas en lenguaje científico. Con el transcurso del tiempo, la evaluación de la ciencia adquirió tal magnitud que aquel mecanismo de diálogo *rudimentario* reveló imperfecciones. No

existía un único lenguaje científico sino que había varios *dialectos* —ciencia dura/blanda/aplicada/básica—, lo cual condujo a poner en duda la validez de los interlocutores; tampoco eran sólo dos las partes involucradas —políticos y científicos— sino que había otras partes *interesadas*; ni existía un claro mensaje por transmitir. Así, las instancias de evaluación de la ciencia sufrieron diversas modificaciones que buscaban que el diálogo entre políticos y científicos se renovara.

## Referencias

- Amin, Mayur y Michael Mabe. 2003. "Impact Factors: Use and Abuse". *Medicina* 63 (4): 347-354.
- Beigel, Fernanda. 2013. "Centros y periferias en la circulación internacional del conocimiento". *Nueva Sociedad* 245: 110-123.
- Beigel, Fernanda. 2015. "Circuitos segmentados de consagración académica: las revistas de Ciencias Sociales y Humanas en la Argentina". *Información, Cultura y Sociedad* 32: 11-35.
- Bianco, Mariela, Natalia Gras y Judith Sutz. 2014. "Reflexiones sobre la práctica de la evaluación académica". En *Veinte años de políticas de investigación en la Universidad de la República: aciertos, dudas y aprendizajes*, coordinado por Mariela Bianco y Judith Sutz, 209-235. Montevideo: Ediciones Trilce.
- Bornmann, Lutz. 2011. "Peer Review and Bibliometric: Potentials and Problems". En *University Rankings*, editado por Jung Cheol Shin, Robert K. Toutkoushian y Ulrich Teichler, 145-164. Berlin: Springer.
- Bozeman, Barry, James S. Dietz y Monica Gaughan. 2001. "Scientific and Technical Human Capital: An Alternative Model for Research Evaluation". *International Journal of Technology Management* 22 (7/8): 716-740.
- Braun, Dietmar. 1998. "The Role of Funding Agencies in the Cognitive Development of Science". *Research Policy* 27 (8): 807-821. [http://dx.doi.org/10.1016/S0048-7333\(98\)00092-4](http://dx.doi.org/10.1016/S0048-7333(98)00092-4)
- Campanario, Juan Miguel. 2002. "El sistema de revisión por expertos (peer review): muchos problemas y pocas soluciones". *Revista Española de Documentación Científica* 25 (3): 166-184. <http://dx.doi.org/10.3989/redc.2002.v25.i3.107>
- Cañibano, Carolina y Bari Bozeman. 2009. "Curriculum Vitae Method in Science Policy and Research Evaluation: The State-of-the-Art". *Research Evaluation* 18 (2): 86-94. <http://dx.doi.org/10.3152/095820209X441754>
- Cañibano, Carolina, Javier Otamendi y Francisco Solís. 2010. "Investigación y movilidad internacional: análisis de las estancias en centros extranjeros de los investigadores andaluces". *Revista Española de Documentación Científica* 3: 428-457. <http://dx.doi.org/10.3989/redc.2010.3.736>
- Chubin, Daryl y Edward Hackett. 1990. *Peerless Science: Peer Review and U.S. Science Policy*. Nueva York: State University of New York Press.
- Cole, Stephen, Jonathan Cole y Gary A. Simon. 1981. "Chance and Consensus in Peer Review". *Science* 214 (4523): 881-886. <http://dx.doi.org/10.1126/science.7302566>
- Davyt, Amílcar y Lea Velho. 2000. "A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro?". *História, Ciências, Saúde* VII (1): 93-116. <http://dx.doi.org/10.1590/S0104-59702000000200005>
- De la Vega, Iván. 2003. "Cienciometría y política científica en la periferia: el caso de Venezuela". *Revista Espacios* 24 (1): 5-17.
- Derrick, Gemma y Vincenzo Pavone. 2013. "Democratising Research Evaluation: Achieving Greater Public Engagement with Bibliometrics-Informed Peer Review". *Science and Public Policy* 40 (5): 563-575. <http://dx.doi.org/10.1093/scipol/sct007>
- Frederiksen, Lars Frode, Finn Hansson y Søren Barlebo Wenneberg. 2003. "The Agora and the Role of Research Evaluation". *Evaluation* 9: 149-172. <http://dx.doi.org/10.1177/1356389003009002003>
- Gómez, Yuri J. 2004. "Política científica colombiana y bibliometría: usos". *Nómadas* 22: 241-254.
- Grediaga Kuri, Rocío. 2011. "El papel de pares y publicaciones periódicas en la validación del conocimiento científico". *Revista Mexicana de Investigación Educativa* 16 (48): 7-14.
- Gross, P.L.K. y E.M. Gross. 1927. "College Libraries and Chemical Education". *Science* 66 (1713): 385-389.
- Guédon, Jean-Claude. 2011. "El acceso abierto y la división entre ciencia 'principal' y 'periférica'". *Crítica y Emancipación* 6: 135-180.
- Hicks, Diana. 2004. "The Four Literatures of Social Sciences". En *Handbook of Quantitative Science and Technology Research*, editado por Henk F. Moed, Wolfgang Glänzel y Ulrich Schmoch, 476-496. Dordrecht: Springer.
- Jiménez-Buedo, María e Irene Ramos Vielba. 2009. "¿Más allá de la ciencia académica?: modo 2, ciencia posnormal y ciencia posacadémica". *Arbor* CLXXXV (738): 721-737.
- Krauskopf, Manuel y María Inés Vera. 1995. "Las revistas latinoamericanas de corriente principal: indicadores y estrategias para su consolidación". *Revista Interciencia* 20 (3): 144-148.
- Kreimer, Pablo. 1998. "Publicar y castigar. El paper como problema y la dinámica de los campos científicos". *Redes* 5(12): 51-73.
- Kreimer, Pablo. 2011. "La evaluación de la actividad científica: desde la indagación sociológica a la burocratización. Dilemas actuales". *Propuesta Educativa* 2: 59-77.
- Kreimer, Pablo. 2015. "Los mitos de la ciencia: desventuras de la investigación, estudios sobre ciencia y políticas científicas". *Nómadas* 42: 33-51.
- Lamont, Michele. 2009. *How Professors Think: Inside the Curious World of Academic Judgement*. Cambridge - Londres: Harvard University Press.
- Langfeldt, Liv. 2006. "The Policy Challenges of Peer Review: Managing Bias, Conflict of Interests and Inter-

- disciplinary Assessments". *Research Evaluation* 15 (1): 31-41. <http://dx.doi.org/10.3152/147154406781776039>
29. Langfeldt, Liv y Svein Kyvik. 2011. "Researchers as Evaluators: Tasks, Tensions and Politics". *Higher Education* 62 (2): 199-212. <http://dx.doi.org/10.1007/s10734-010-9382-y>
  30. Lattuada, Mario. 2014. "La evaluación de la investigación en las universidades argentinas. Contextos, culturas y limitaciones". *Revista Iberoamericana de Ciencia, Tecnología y Sociedad* 9 (27): 157-164.
  31. Lee, Carole, Cassidy Sugimoto, Guo Zhang y Blaise Cronin. 2013. "Bias in Peer Review". *Journal of the American Society for Information Science and Technology* 64 (1): 2-17. <http://dx.doi.org/10.1002/asi.22784>
  32. Marquina, Mónica. 2006. *La evaluación por pares en el escenario actual de aseguramiento de la calidad de la educación superior. Un estudio comparativo de seis casos nacionales*. Buenos Aires: Coneau.
  33. Miguel, Sandra. 2011. "Revistas y producción científica de América Latina y el Caribe: su visibilidad en SciELO, RedALyC y SCOPUS". *Revista Interamericana de Bibliotecología* 34 (2): 187-199.
  34. Molas-Gallart, Jordi. 2012. "Research Governance and the Role of Evaluation A Comparative Study". *American Journal of Evaluation* 33 (4): 583-598.
  35. Mugnaini, Rogério. 2006. "Caminhos para adequação da avaliação da produção científica brasileira: impacto nacional versus internacional". Disertación doctoral, Universidade de São Paulo.
  36. Mulligan, Adrian, Hall Louise y Raphael Ellen. 2013. "Peer Review in a Changing World: An International Study Measuring the Attitudes of Researchers". *Journal of the American Society for Information Science and Technology* 64 (1): 132-161. <http://dx.doi.org/10.1002/asi.22798>
  37. Musselin, Christine. 2013. "How Peer Review Empowers the Academic Profession and University Managers: Changes in Relationships Between the State, Universities and the Professoriate". *Research Policy* 42 (5): 1165-1173. <http://dx.doi.org/10.1016/j.respol.2013.02.002>
  38. Mutzl, Rüdiger, Lutz Bornmann y Hans-Dieter Daniel. 2013. "Types of Research Output Profiles: A Multilevel Latent Class Analysis of the Austrian Science Fund's Final Project Report Data". *Research Evaluation* 22: 118-133. <http://dx.doi.org/10.1093/reseval/rvs038>
  39. Naidorf, Judith y Daniela Perrotta. 2015. "La ciencia social politizada y móvil de una nueva agenda latinoamericana orientada a prioridades". *Revista de la Educación Superior* 44 (174): 19-46. <http://dx.doi.org/10.1016/j.resu.2015.05.001>
  40. Naidorf, Judith, Patricia Giordana y Mauricio Horn. 2007. "La pertinencia social de la Universidad como categoría equívoca". *Nómadas* 27: 22-33.
  41. Nicholas, David, Anthony Watkinson, Hamid Jamali, Edi Herman, Carol Tenopir, Rachel Volentine y Kenneth Levine. 2015. "Peer Review: Still King in the Digital Age". *Learned Publishing* 28 (1): 15-21. <http://dx.doi.org/10.1087/20150104>
  42. Ortiz, Renato. 2009. *La supremacía del inglés en las ciencias sociales*. Buenos Aires: Siglo Veintiuno Editores.
  43. Ríos Gómez, Claudia y Víctor Herrero Solana. 2005. "La producción científica latinoamericana y la ciencia mundial: una revisión bibliográfica (1989-2003)". *Revista Interamericana de Bibliotecología* 28 (1): 43-61.
  44. Rip, Arie. 1994. "The Republic of Science in the 1990s". *Higher Education* 28 (1): 3-23.
  45. Romanos de Tiratel, Susana. 2008. *Revistas argentinas de Ciencias Sociales y Humanidades: visibilidad en bases de datos internacionales*. Buenos Aires: Instituto de Investigaciones Bibliotecológicas - Universidad de Buenos Aires - Facultad de Filosofía y Letras.
  46. Rueda-Clausen Gómez, Christian, Cristina Villa-Roel Gutiérrez y Christian E. Rueda-Clausen Pinzón. 2005. "Indicadores bibliométricos: origen, aplicación, contradicción y nuevas propuestas". *MedUNAB* 8 (1): 29-36.
  47. Sanz Menéndez, Luis. 2014. "La evaluación de la ciencia y la investigación". *RES. Revista Española de Sociología* (21): 137-148.
  48. Scott, Alister. 2007. "Peer Review and the Relevance of Science". *Futures* 39: 827-845. <http://dx.doi.org/10.1016/j.futures.2006.12.009>
  49. Seglen, Per. O. 1997. "Why the Impact Factor of Journals Should Not Be Used for Evaluating Research". *British Medical Journal* 314: 498-502.
  50. Sutz, Judith. 2014. "Calidad y relevancia en la investigación universitaria: apuntes para avanzar hacia su convergencia". *Revista Iberoamericana de Ciencia, Tecnología y Sociedad* 27 (9): 63-83.
  51. Vaccarezza, Leonardo y Juan Pablo Zabala. 2002. *La construcción de la utilidad social de la ciencia: Investigadores en biotecnología frente al mercado*. Quilmes: Universidad Nacional de Quilmes.
  52. Van Leeuwen, Ted, Henk Moed, Robert Tijssen, Martijn Visser y Anthony Van Raan. 2001. "Language Biases in the Coverage of the Science Citation Index and Its Consequences for International Comparisons of National Research Performance". *Scientometrics* 51 (1): 335-346.
  53. Vessuri, Hebe, Jean-Claude Guédon y Ana María Cetto. 2014. "Excellence or Quality? Impact of the Current Competition Regime on Science and Scientific Publishing in Latin America and Its Implications for Development". *Current Sociology* 62 (5): 647-665. <http://dx.doi.org/10.1177/0011392113512839>
  54. Weller, Ann. 2001. *Editorial Peer Review: Its Strengths and Weaknesses*. Nueva Jersey: American Society for Information Science and Technology.
  55. Ziman, John. 2003. *¿Qué es la ciencia?* Madrid: Cambridge University Press.
  56. Zuckerman, Harriet y Robert Merton. 1971. "Patterns of Evaluation in Science. Institutionalisation, Structure and Functions of Referee System". *Minerva* 9 (1): 66-100. <http://dx.doi.org/10.1007/BF01553188>