RIDAA
Repositorio Institucional
Digital de Acceso Abierto de la
Universidad Nacional de Quilmes

Universidad
Nacional
de Quilmes

Invernizzi, Noela

Críticas recientes a la evaluación de la investigación : ¿vino nuevo en odres viejos?

Esta obra está bajo una Licencia Creative Commons Argentina.
Atribución - No Comercial - Sin Obra Derivada 2.5
https://creativecommons.org/licenses/by-nc-nd/2.5/ar/

Documento descargado de RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes de la Universidad Nacional de Quilmes

Cita recomendada:
Invernizzi, N. y Davyt, A. (2019). Críticas recientes a la evaluación de la investigación: ¿vino nuevo en odres viejos? Redes, 25(49), 233-252. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes http://ridaa.unq.edu.ar/handle/20.500.11807/3459

Puede encontrar éste y otros documentos en: https://ridaa.unq.edu.ar

Universidad Nacional de Quilmes
Roque Sáenz Peña 352 // Bernal
Buenos Aires // Argentina
t.: (+41 11) 4365 7100
f.: (+54 11) 4365 7101
info@unq.edu.ar

Noela Invernizzi / Amílcar Davyt***

Este artículo examina las críticas recientes a la evaluación de la producción científica plasmadas en siete documentos / manifiestos publicados entre 2010 y 2016. Los objetivos planteados son de dos tipos: a) examinar y valorar las críticas a la evaluación de la ciencia –académica– realizadas en los mencionados manifiestos y b) comparar esas críticas con las realizadas en décadas anteriores, indagando si hay identidades, divergencias, cuestiones viejas y emergentes. Los textos fueron seleccionados llevando en consideración su amplia divulgación y la discusión que suscitaron. El estudio fue realizado a partir de un análisis de contenido de los textos y de su contextualización mediante revisión de literatura. Concluimos que la mayor parte de las críticas a la evaluación de la producción científica realizada en los manifiestos no es nueva, aunque hay un claro desplazamiento del centro de la discusión desde la revisión por pares hacia la evaluación cuantitativa mediante indicadores bibliométricos. Es evidente también que algunas críticas, como las que señalan un deterioro en la calidad y relevancia de la actividad científica, han ido adquiriendo más fuerza en los últimos años, conforme la escala y la eficiencia de aplicación de la evaluación cuantitativa se han ido ampliando y sus efectos sobre la dinámica de producción de conocimiento y la actividad universitaria tienden a manifestarse más claramente.

PALABRAS CLAVE: EVALUACIÓN CIENTÍFICA – REVISIÓN POR PARES – BIBLIOMETRÍA – INVESTIGACIÓN ACADÉMICA

* Universidade Federal do Paraná, Curitiba. Correo electrónico: <noela@ufpr.br>.
** Universidad de la República, Montevideo. Correo electrónico: <amilcardavyt@gmail.com>.

En la última década ha recrudecido el debate sobre las políticas de investigación científica, en especial sobre la evaluación de sus productos y resultados. Muchas discusiones se direccionan a las métricas e indicadores utilizados para la evaluación de la producción científica. Hay cuestionamientos más de fondo, sobre los efectos de tales políticas científicas sobre la propia ciencia, discutiendo la forma en que tensionan su orientación, calidad, integridad y relevancia social. Se apunta el cercenamiento de la creatividad científica y los obstáculos al trabajo interdisciplinar impuestos por el "*publish or perish*" y por las formas de acceder a financiamientos para la investigación. Se sostiene que el sistema de *peer review* se ha tornado insustentable, que ha sido substituido por las métricas, y que el monopolio de las revistas científicas parece comandar la ciencia. Se presentan también argumentos sobre los efectos del llamado "productivismo" sobre la misión de las universidades públicas, sobre el trabajo docente y sobre la formación de las futuras generaciones de científicos.

Este tipo de críticas han sido recurrentes, con cierta periodicidad, en la historia de la política de investigación, casi desde sus orígenes, o al menos desde la década de 1960, en diversos medios de comunicación científica –revistas, foros, eventos– y desde distintas áreas del conocimiento. A lo largo de las décadas transcurridas, los argumentos han sido diversos.

Situados en este contexto histórico y en la actualidad, se examinan aquí siete posturas, expresadas en sendos documentos / manifiestos, publicados entre 2010 y 2016. Esta selección no es comprensiva; pueden encontrarse más documentos sobre el tema en el período. El criterio de selección de estos se debe a su citación recurrente en la literatura y en comunicaciones en eventos científicos, y a la discusión generada más allá de ellos, en diarios, revistas y blogs. Así, consideramos que, aunque estos manifiestos sean bastante heterogéneos, pues surgieron de situaciones diferentes, tienen finalidades diversas, y parten de individuos o de colectivos, el conjunto sintetiza los principales aspectos de la discusión actual sobre las políticas de evaluación de la ciencia; es decir, son representativos de los debates recientes en esta temática. Son ellos: "Slow Science Manifesto" - Berlín; "The Slow Science Academy" (2010) (en adelante, Slow); "Innovation, Sustainability, Development: a New Manifesto" - Sussex; Centro STEPS - Universidad de Sussex (2010) (en adelante, STEPS); "San Francisco Declaration on Research Assesment" (DORA) - San Francisco; American Society for Cell Biology *et al.* (2012) (en adelante, DORA); "Charte de la Désexcellence" - Bruselas; L'Atelier des Chercheurs - Universidad Libre de Bruselas (2014) (en ade-

lante, Charte); "The Leiden Manifesto" - Leiden; Hicks *et al.* (2015) (en adelante, Leiden); "The Academic Manifesto: from an occupied to a public University" - Países Bajos; Halffman y Radder (2015) (en adelante, Academic), y "Saving Science" - Estados Unidos; Sarewitz (2016) (en adelente, Saving).

El estudio fue realizado mediante análisis de contenido de los manifiestos. Para ello, fue identificado un conjunto de dimensiones a partir de las principales cuestiones en debate sobre evaluación científica entre las décadas de 1960 y 2000 tomando como referencia el artículo de Davyt y Velho (2000). Ello permitió sistematizar los argumentos que se mantienen en la actualidad, registrándolos en una matriz. La aparición de nuevas temáticas en los manifiestos fue registrada como cuestiones emergentes. Para situar y profundizar la discusión sobre este debate, recurrimos a una revisión amplia de la literatura sobre la temática, tanto actual como aquella producida en décadas anteriores, en el contexto de las antiguas discusiones.

Los objetivos planteados son de dos tipos: examinar y valorar las críticas a la evaluación de la ciencia académica realizadas en los mencionados manifiestos y comparar esas críticas con las realizadas en décadas anteriores.

El análisis se centra en las siguientes dimensiones y preguntas: ¿qué actores firman los manifiestos?; ¿cuál es el centro de la crítica?; ¿qué criterios de calidad y relevancia científica critican y se proponen?; ¿cómo se sitúa la investigación en el ámbito de las instituciones públicas?; ¿cuáles son las alternativas propuestas?; ¿qué convergencias y divergencias se identifican entre los manifiestos? Por último, pero no menos importante: ¿qué hay de nuevo en las discusiones actuales?; ¿en qué se diferencian de las consideraciones y críticas históricas?

A continuación, en la primera sección, este artículo se detiene en la consideración sintética de las antiguas críticas a la evaluación académica, tanto en términos de la propia evaluación por pares como del uso cada vez más importante de los indicadores bibliométricos como instrumento de evaluación. Es decir, los odres viejos. Luego, en la segunda sección, se describen algunas características de los recientes documentos / manifiestos estudiados, para luego sistematizar y detallar, en las tres secciones siguientes, las críticas en ellos contenidas a la evaluación cuantitativa y al sistema de pares. En la última sección, a modo de cierre, se explora una comparación entre ambos momentos: coincidencias, divergencias, cuestiones viejas y emergentes, y se discuten algunas explicaciones acerca del rumbo tomado en relación a lo esperado a fines de siglo pasado y se plantean posibles perspectivas para este siglo XXI.

Decir que las críticas a la evaluación académica son tan antiguas como la propia evaluación académica sería casi como retrotraerse a las discusiones acerca de las bondades –o no– de Isaac Newton al evaluar los textos para las *Philosophical Transactions*, en especial los de Gottfried Leibniz. Sin embargo, cierto es que la evaluación de la investigación realizada es una práctica propia del período de la institucionalización de la ciencia moderna, en el siglo XVII (Shapin, 1996). Hacia finales de ese siglo y comienzos del siguiente, las nacientes sociedades científicas europeas establecieron sus publicaciones oficiales y los métodos para decidir cuáles textos serían publicados (Zuckerman y Merton, 1973), en formatos semejantes a los actuales. Es decir, ha sido persistente históricamente la inicial forma de evaluación de la investigación, a través de la revisión por pares de los manuscritos (Harnad, 1998). A lo largo del tiempo, mediante ese procedimiento, de carácter comunal, gobernado por reglas formales e informales, los manuscritos científicos –en el caso de que tengan los méritos, contenido, métodos y estilo juzgados adecuados por los pares– se han transformado en conocimientos consensuales, en verdades científicas y, por transitiva, las calificaciones de sus autores han sido aprobadas (Chubin y Hackett, 1990). Se legitiman mutuamente, en ese proceso, investigadores y artículos (Fabbri y Latour, 1995).

No tan atrás en el tiempo vamos a buscar todos los odres viejos, pero sí tal vez hace medio siglo. Es decir: las primeras anotaciones críticas hacia algunos aspectos de la evaluación por pares se registran en la pasada década de 1960, y prosiguen durante las últimas décadas del siglo pasado. A esta literatura, relativamente antigua, recurrimos para ejemplificar las viejas críticas.

Antes de profundizar en ellas, debemos anotar que además de la primigenia evaluación por pares de manuscritos, es a mediados del siglo pasado que se articulan formas de evaluación, también por pares, de propuestas de investigación o proyectos, planteadas en la solicitud de financiamiento a las primeras agencias de fomento, creadas o consolidadas desde la década de 1940 en adelante (Chubin y Hackett, 1990). En este caso, la diversidad ha sido mayor que en la evaluación de manuscritos para publicación, desde sistemas con muy importante control por parte de los pares hasta situaciones donde solamente se avalan o legitiman las decisiones institucionales (Roy, 1984; Travis y Collins, 1991).

El contrato social entre ciencia y sociedad establecido en torno a 1950 proveía sustento conceptual para el funcionamiento de este sistema de eva-

luación, donde las decisiones recaían, al menos parcialmente, en los pares. Dentro del modelo conceptual lineal, el punto de partida del desarrollo tecnológico y socioeconómico estaba en los científicos y el conocimiento por ellos producido. Así, era lógico esperar que quienes pudieran opinar y decidir cuál era la "buena ciencia" fueran los propios pares. Desde sus orígenes, las agencias delegaron en los científicos las microdecisiones de financiamiento (Dickson, 1988).

Poco después, ya en la década de 1960, comienzan a desarrollarse, por parte de los organismos públicos, formas de medición de la actividad científica: es el inicio de la "cienciometría" (Price, 1986a), y en particular del desarrollo de la teoría y la metodología de los indicadores bibliométricos (Holbrook, 1992). Este desarrollo ocurre de la mano de la creación del *Institute for Scientific Information* (ISI) por parte de Eugene Garfield y de la publicación de su *Science Citation Index* (SCI) desde 1963 (Price, 1978). Los datos del SCI entraron en el escenario de la evaluación cuando la *National Science Foundation* los incorporó en sus *Science Indicators Reports* de 1972 (De Bellis, 2014). Desde entonces, su uso se extendió rápidamente en la política y la administración científica de muchos países (Gläser y Laudel, 2007).

A lo largo de la década de 1990, estos indicadores cuantitativos pasaron a tener un peso fundamental, por diversos factores. La enorme expansión de la actividad científica limitó cada vez más el lugar de la evaluación por pares (Gläser y Laudel, 2007). El ISI "evolucionó": fue comprado por la Thomson Corporation en 1992, que creó la Web of Science, WoS (De Bellis, 2014). La década de 1990 fue momento también de la difusión de una cultura de "nueva administración", rendición de cuentas, evaluación externa y *rankings*, con la llegada del New Public Management a las universidades y centros académicos (Alvesson y Spicer, 2016).

Las primeras críticas claramente documentadas a la evaluación por pares, se dieron en la década de 1960, comenzando con el histórico debate acerca de las prioridades de investigación en la conocida revista *Minerva*, contando con Alvin Weinberg como uno de sus exponentes (Weinberg, 1963; 1964). El centro de su crítica fue el "universo cerrado" de la evaluación, es decir, internalista y disciplinario, sin miradas de fuera del campo disciplinar o, más aún, de fuera de la comunidad científica, que permitieran evaluar mejor el impacto de la propuesta o manuscrito en el conjunto de la sociedad (Weinberg, 1963). Estos planteos llevaron a debates públicos en ambientes como el Congreso norteamericano o la revista *Science* (Mitroff y Chubin 1979), en la misma dirección, relacionando la evaluación con las discusiones sobre la autonomía de los investigadores.

Un matiz de esta crítica refería a la idea de sistema "cerrado y sesgado", es decir, que privilegia a ciertas redes de investigadores, con visiones comunes sobre la disciplina, o pertenecientes a instituciones de alto prestigio y reconocimiento, en detrimento de otros grupos o sectores ajenos a ellos; una forma del conocido efecto Mateo (Merton, 1968) combinado con posibles conflictos de interés (Chubin y Hackett, 1990; Martin e Irvine, 1983). Desde fines de la década de 1990, además, varios estudios han confirmado diversos sesgos y tendencias particularistas (Wenneras y Wold, 1997; Link, 1998); es decir, en el proceso, sobre todo en la asignación de pares evaluadores, existen criterios subjetivos, no explícitos, que llevan a evaluaciones no imparciales.

En la época se anotaba también que la evaluación por pares presenta resistencia a las ideas innovadoras, favoreciendo líneas de investigación tradicionales; es decir, se calificaba al sistema como conservador, ya que artículos o propuestas con ideas o técnicas heterodoxas, nuevas, creativas, fuera de las reglas aceptadas, tienen mayores dificultades que aquellos de líneas de investigación bien establecidas y consolidadas; en suma, otro sesgo, en este caso, anticreativo (Chubin y Connolly, 1982; Roy, 1984; Ziman, 1994).

Ha habido críticas centradas en la eficiencia del proceso, en el sentido de desperdicio de recursos y tiempo de los investigadores, así como de la administración en general, que podrían ser aplicados directamente a la investigación (Roy, 1984; Ziman, 1994). También se ha criticado la falta de transparencia habitual, el anonimato de los pares –*blind review*, a veces *double-blind*–, quienes por lo tanto no deben rendir cuentas de sus opiniones y decisiones; esta característica permitiría revisiones tendenciosas o maliciosas, inclusive –o más aún– de mala calidad (Chubin y Hackett, 1990).

Se ha mencionado, además, que el sistema de evaluación promueve la competencia y no la necesaria cooperación y colaboración entre investigadores (Roy, 1984).

Desde los comienzos del uso de indicadores bibliométricos como apoyo o complemento a la evaluación académica, surgió otro conjunto de críticas o cuestionamientos. La primera de las críticas, conceptual, es la idea de que gran parte de la actividad científica no se describe en la literatura, en particular el conocimiento tácito; en la formación de investigadores, la publicación es solamente uno de los tipos de comunicación (Gilbert, 1978; Manten, 1980).

Pero además, se planteaba que el procedimiento de construcción de las bases de datos de publicaciones –sobre las que se basan los indicadores–, fue llevando a considerar estándares únicos que no distinguen entre áreas del conocimiento, fases de la investigación y/o del desarrollo de las comu-

nidades académicas o regiones del planeta. En particular, se entendía que no se consideran las significativas diferencias en la organización, en el sistema de comunicación y en el comportamiento de los investigadores entre áreas del conocimiento: los puntos de partida de las bases se relacionaron con áreas básicas, llamadas comúnmente "duras" (Frame, 1985; Price, 1986b; Velho, 1989), poco aplicables a las dinámicas de otras áreas del conocimiento, en especial las sociales (Lindsey, 1978).

Por otra parte, los indicadores bibliométricos fueron desarrollados en situaciones de financiamiento sobre una base y una práctica académica bien establecidas; no es lo mismo aplicarlos en situaciones en las cuales los esfuerzos están destinados a crear una infraestructura para el crecimiento de la ciencia. Varios autores anotaron tempranamente el sesgo de la cobertura de las revistas y publicaciones utilizadas en las mediciones, favoreciendo a los países científicamente centrales (Rabkin e Inhaber, 1979; Frame, 1980). Es decir, se han señalado aquí sesgos entre regiones del mundo y entre fases de desarrollo científico.

Aun dentro de países que se pueden denominar "centrales científicamente", desde hace tiempo se han señalado críticas al sesgo del idioma inglés como medio de comunicación predominante y excluyente (Narin y Carpenter, 1975; Jagodzinski-Sigogneau, Courtial y Latour, 1982; Arvanitis y Chatelin, 1988). Se llegaron a construir bases de datos bibliométricos alternativas, basados en países de otras lenguas, para equilibrar las evaluaciones –el ejemplo clásico es la base de datos bibliográficos Pascal del Institute de l'Information Scientifique et Technique del Centre National de la Recherche Scientifique (INIST-CNRS).

Esta variedad de críticas, que aborda desde la idoneidad técnica de los mecanismos de evaluación hasta la geopolítica de la ciencia, se intensifican, conforme indican los autores, hacia la década de 1990, contexto en que varios trabajos teórico-conceptuales relevantes y removedores, como los de Etzkowitz (1990), Cozzens *et al.* (1990), Ziman (1994), Gibbons *et al.* (1994) y otros cercanos a la década de 2000, exponen que el quehacer científico estaría pasando por cambios substanciales, como mayores demandas por resultados de utilidad social y económica, creciente privatización, internacionalización y aumento de casos de mala conducta científica, entre otros aspectos; de hecho, pusieron en el tapete la discusión acerca de posibles nuevas formas de relacionamiento entre las actividades de investigación, la economía y la sociedad en general, en lo que algunos denominaron aparición de un nuevo contrato social entre ciencia y sociedad. En este nuevo contexto, surgieron propuestas de ampliar la evaluación de la investigación más allá de la comunidad científica, involucrando los usuarios de la ciencia,

a partir de los mencionados textos conceptuales relevantes y, en particular, de lo actuado por algunas agencias de financiamiento en el mundo, en esa misma década (Hills y Dale, 1995; Van den Beemt y Van Raan, 1995; Avalos, 1997; Davyt y Velho, 2000). En vista de esta situación, los últimos autores mencionados concluyen con optimismo por los cambios que se estaban produciendo, augurando que el nuevo contrato social entre ciencia y sociedad tendría uno de sus capítulos relevantes en la apertura de la evaluación académica a otros actores, entre otros cambios en el sistema (Davyt y Velho, 2000).

Aunque hay elementos de continuidad en la década actual con la situación emergente en la década de 1990, la escala en la que ocurren los fenómenos que se comenzaban a señalar en esos momentos se amplía considerablemente. Seguramente esta es la razón de la aparición del amplio conjunto de documentos / manifiestos en la actualidad, de los cuales, como se señaló inicialmente, seleccionamos los que a continuación se describen.

La tabla 1 pretende caracterizar brevemente cada uno de los siete documentos / manifiestos utilizados en el análisis posterior.

Como puede verse en la tabla 1, los textos seleccionados fueron promovidos principalmente por académicos, cinco de ellos por europeos y uno por un norteamericano; el séptimo fue planteado por editores y casas editoriales de revistas científicas, además de firmantes que apoyan el planteo.

En todos los textos de la selección aparecen, de distintas maneras, críticas a la evaluación académica centradas en los procesos de pares y críticas a las evaluaciones cuantitativas. Sin embargo, los planteos en torno al uso de indicadores bibliométricos ocupan mayor espacio y presentan más detalles en la mayoría de ellos. Esto podría deberse, en principio, a que desde la época de las antiguas críticas, reseñadas en la primera sección, se ha intensificado su uso, de la mano de su sofisticación; en algunas instancias y procesos de decisión académica, inclusive, el uso de indicadores cuantitativos tiende a restar protagonismo a la evaluación por pares, principalmente por la incapacidad del sistema de dar cuenta de la sustancial ampliación de la

Tabla 1. Documentos / manifiestos, sus autores y sus contenidos

2010 - STEPS "Innovation, Sustainability and Development. A new manifesto"	Heredero del histórico "Manifiesto de Sussex", publicado cuarenta años antes, se centra en el papel de la C&T para el desarrollo, la justicia social y la sustentabilidad. Presenta una crítica inicial sobre el no cumplimiento de las promesas de la ciencia y propone la reorientación de la innovación a partir de una consideración política explícita de su dirección, distribución y diversidad.	STEPS - Centro de investigación formado por profesores del Institute of Development Studies (IDS) y la Science Policy Research Unit (SPRU) de la Universidad de Sussex, Reino Unido. Mantiene un consorcio con varios países y regiones del mundo.
2010 - Slow "The Slow Science Manifesto"	Muy sucinto, demanda una desaceleración del desarrollo científico, para garantizar mayor calidad y discusión del conocimiento. Propone refundar las academias como espacio, tiempo y medios para recrear la torre de marfil perdida por los científicos.	Grupo de académicos alemanes –no identificados–. Es importante notar que hay muchas vertientes en el movimiento Slow Science (véase Gosselain, 2012).
2012 - DORA "San Francisco Declaration on Research Assessment"	Demanda mejorías en la evaluación de los resultados de la investigación científica por las agencias financiadoras, instituciones académicas y otros. Llama a disociar la calidad de la investigación del periódico que la publica –critica el uso indebido del *Journal Impact Factor.*	Grupo de editores y de casas editoras de revistas científicas reunidos en el Encuentro Anual de la American Society for Cell Biology. En noviembre de 2018 había reunido 13.200 firmas individuales y 733 institucionales.
2014 - Charte "Charte de la désexcellence"	La carta es un documento de varias páginas que abarca todas las dimensiones de la vida universitaria, que habría sido precarizada como resultado de la cultura de la excelencia impuesta por las reformas neoliberales de los acuerdos de Bolonia. En el capítulo sobre la investigación académica se critica el uso de indicadores cuantitativos y criterios de excelencia que serían indiferentes a los contenidos y enmascararían la pérdida de calidad de la ciencia.	Profesores de la Universidad Libre de Bruselas: A-E. Bourgaux, E. Bribosia, B. D'Hainaut-Zveny, J-M. Decroly, Ch. Deligne, O. Gosselain, J-J. Heirwegh, P. Lannoy, G. Lebeer, A. Livingstone Smith, E. Martinez, J. Moriau, C. Nagels, J. Pieret, V. Piette, C. Reyniers, B. Truffin, M. Van Criekingen, K. Vanhaesebrouck, E. Wolff.
2015 - Leiden "The Leiden Manifesto for Research Metrics"	Critica el mal uso de indicadores bibliométricos, y la tendencia a la substitución del *peer review* por estos para la evaluación científica y la gobernanza de la ciencia. Formula diez principios que condensan buenas prácticas para la evaluación de la investigación basada en métricas y considera que estas son auxiliares –no sustitutas– del procesos de *peer review*.	Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke, Ismael Rafols. La idea del Manifiesto surgió en un el Congreso Indicadores de Ciencia y Tecnología, realizado en la ciudad de Leiden en 2014.
2015 - Academic "The Academic Manifesto: From an occupied to a public university"	Publicado en formato de artículo académico en la revista *Minerva*, abarca varias dimensiones de la vida universitaria, transformada por la nueva lógica de administración centrada en la ideología de la excelencia, y sus desdoblamientos como medición, competición, eficiencia. La investigación académica es uno de los temas abordados, cuestionando fuertemente lo que consideran el régimen fetichista de los indicadores.	Profesores de universidades holandesas: Willem Halffman (Nijmegen University) y Hans Radder (VU University Amsterdam). Fue traducido y discutido en más de 14 países hasta 2017 (véase Halffman y Radder, 2017).
2016 - Saving "Saving Science"	Publicado en *The New Atlantis*, este artículo con tono de manifiesto llama a salvar la ciencia que se habría vuelto contestable, no confiable, sin uso, e inclusive errada. Esto sería producto de una política científica "à la Bush", de la incapacidad del sistema de *peer review* para asegurar su calidad y de las formas de evaluación de las carreras académicas.	Daniel Sarewitz, Director del Center for Science, Policy and Outcomes, Arizona State University. El texto fue replicado, comentado y criticado en numerosos blogs sobre ciencia.

actividad científica. Antes de continuar con esta reflexión, se presentan detalles de las críticas que aparecen en los textos a la "cuantitivización" de la evaluación académica.

Podría decirse que los textos más "radicales" en la crítica sobre estos aspectos son: Slow, DORA, Charte y Saving. En primer lugar, puede destacarse una crítica conceptual de algunos de los textos: los indicadores bibliométricos no evalúan adecuadamente la "calidad" de la ciencia, sino solamente el desempeño de los investigadores y sus productos –o *performance*–, según Academic, "una ilusión tácticamente bien pensada" (Halffman y Radder, 2015: 167). Los datos sobre productos han sustituido al juicio sobre las actividades, dice de alguna manera Leiden, sosteniendo que ello puede llegar a dañar la ciencia con instrumentos originalmente desarrollados para mejorarla. En esta dirección se suma DORA, también, al sostener que hay un mal uso sistemático y generalizado de los indicadores bibliométricos: en última instancia, ellos ignoran el contenido científico de los productos evaluados, midiendo, por ejemplo, el impacto del vehículo donde se publica. Saving plantea esta crítica de forma más profunda, asumiendo que estos problemas son fallas estructurales del sistema.

Otro de los cuestionamientos principales refiere a la aplicación de un estándar común, único, tanto a las diferentes áreas del conocimiento científico como a las fases de la investigación –de básico a aplicado– como a regiones del mundo con diferentes grados de desarrollo científico. Tanto DORA como Academic y Leiden sostienen que la ampliación del uso de indicadores y su sofisticación ha potenciado esta problemática. DORA ejemplifica afirmando que los patrones de citaciones son área-específicos. Leiden también entra en detalles sobre esta cuestión, atacando el uso de indicadores de performance universal, como si pudiera existir una idea de excelencia científica abstracta; sostiene que la evaluación debe adaptarse a las misiones de los grupos de investigación, a los tipos de investigación y a las formas de publicar resultados científicos de las diversas áreas del conocimiento. Enfatiza además que el uso de indicadores debe tomar en cuenta el contexto socioeconómico y cultural, lo que lleva implícito adaptarlos a los contextos de ciencia menos desarrollada.

Algunos de los textos refieren a la manipulación de los indicadores o "*gaming*", como estrategia practicada en gran escala para enfrentar las evaluaciones cuantitativas. DORA en particular relata las manipulaciones de los editores de revistas, en particular de los factores de impacto. Por su parte, tanto Charte como Academic discuten prácticas de los investigadores como la de ampliación del número de publicaciones a expensas del

contenido –división de resultados de la investigación en menores o pequeñas partes–, como forma de burlar sistemáticamente los indicadores. Es decir, las críticas a la manipulación de los números abarcan todos los niveles, revistas, instituciones e investigadores, a modo de círculos viciosos.

Se destaca también, en varios manifiestos, la crítica a la centralidad de la lengua inglesa en las bases de publicaciones, y con ella la identificación de la literatura *mainstream* de –algunos de– los países centrales que publican en esa lengua como parámetro de calidad. Esto es sostenido particularmente por Leiden, que argumenta que la calidad se iguala a la publicación en inglés, lo que es problemático para áreas de investigación sobre cuestiones regionales y locales; también por Academic, que identifica el sesgo hacia revistas de lengua inglesa afectando también a los *rankings* nacionales –en este caso, holandeses–; como era de esperar, también la Charte reivindica a las publicaciones en otras lenguas y en revistas nacionales y locales.

Algunos de los documentos hacen notar que la producción y circulación de datos de bibliometría surgen de bases de datos que son propiedad de empresas privadas, lo que lleva a preocupación por la falta de transparencia en el origen de esos datos, así como en su procesamiento y cálculo. DORA y Leiden hacen énfasis en esta cuestión, en especial en la disponibilidad y acceso por parte del público –¡en particular los propios evaluados!–, reclamando por apertura y transparencia.

Las fuertes críticas a la evaluación cuantitativa de la investigación, al uso y abuso de los indicadores bibliométricos, llevan a algunos de los documentos a una defensa o reivindicación de la evaluación por pares, en particular debido a que, a diferencia de la primera, sí evalúa contenidos, o al menos puede hacerlo.

DORA en particular contrapone la capacidad de la evaluación por pares frente al uso de métricas de impacto para valorar el contenido de la investigación. También Slow apoya este tipo de evaluación, con el agregado de la necesidad de ampliar la interacción de pares cara a cara.

Los restantes documentos asumen posiciones más críticas a la revisión por pares, mencionando en todos los casos versiones o matices de las críticas históricas. En particular se menciona el "universo cerrado" de la evaluación, como punto de partida para propuestas de ampliación de la participación de actores no científicos –a la manera del ya mencionado

Weinberg (1963; 1964). Con distintos matices, así lo afirman STEPS, Charte, Academic y Saving.

Es mencionado también el privilegio o sesgo hacia investigadores e instituciones que ya cuentan con prestigio, en diversas formas del mencionado "efecto Mateo", así como otros sesgos. A ello, Saving le agrega los crecientes conflictos de interés que minan el proceso. A partir de todo ello, en Leiden se propone generar nuevos indicadores cuantitativos, diseñados especialmente para contrarrestar esos sesgos, en el contexto de demandas por mayor transparencia en todo el proceso.

También la crítica conceptual acerca de la promoción de la competencia antes que la colaboración y cooperación es reiterada en los manifiestos actuales, especialmente en Academic, Charte y Saving, aunque en gran parte englobando todas las formas de evaluación actuales.

La crítica más radical es presentada por Saving, en el sentido de que la evaluación por pares es incapaz de orientar la ciencia al mismo tiempo que velar por su calidad. Presenta como argumento que hay cada vez más cuerpos de conocimiento científico publicados en periódicos con *peer review* que no son reproducibles, presentan baja calidad y baja confiabilidad, y en muchos casos son directamente inútiles, además de haber pasado por procesos con conflictos de interés. Vincula toda esta situación con los actuales sistemas de promoción académica, que incluyen muchas presiones para producir publicaciones aceleradamente.

Aunque la cuestión de la relevancia y distinción de actores implicados ya fue mencionada, tanto en las críticas históricas como en los manifiestos, vale la pena su destaque.

El punto de partida es la idea de que las agendas de investigación, y por lo tanto el contenido de la ciencia producida, están fuertemente determinadas por el ámbito y los actores involucrados en la evaluación. El texto que tal vez hace mayor énfasis en la necesidad de un redireccionamiento del contenido de la actividad científica es STEPS, criticando fuertemente el universo cerrado e internalista que da lugar a las agendas de investigación actuales y propone la incorporación de nuevos actores públicos y privados. Por su parte, Academic plantea la necesidad de procesos deliberativos más amplios y abiertos para orientar la ciencia hacia el bien común. DORA menciona la posibilidad de evaluar cualitativamente la calidad de la actividad

científica mediante su impacto en la práctica y en la política. Como dejamos ver sobre el final de la sección anterior, Saving es radical en la atribución de los problemas actuales de la ciencia a su universo cerrado, argumentando que no debe ser "protegida" de la sociedad sino que debe mantener una relación directa, abierta e íntima con ella.

Por su parte, Charte manifiesta sus temores por la gran y creciente influencia directa de las empresas en la orientación de la investigación académica, y plantea una posición mucho más mesurada, por llamarle de alguna manera, reivindicando la autonomía de las agendas de investigación con respecto a las demandas y necesidades sociales –más allá de "escucharlas"–. Más radical en esta dirección es Slow, reivindicando, en forma opuesta a los primeros, una reconstrucción de las academias, entendidas como "torres de marfil", situación imprescindible para producir ciencia de calidad. Estas academias implicarían una mirada cara a cara, de autores y pares, en una especie de retorno al pasado distante.

Algunas décadas después de las primeras críticas evidentes, reseñadas en la primera sección, de alguna manera ha ocurrido una cierta "legitimación ambigua" de los sistemas de evaluación académica. Es decir, las críticas se mantienen, se reiteran, se transforman, emergen nuevas, pero hay adhesión a las dinámicas y formas de evaluación construidas.

A lo largo del siglo pasado, se pasó de una absoluta hegemonía de la evaluación por pares, donde solamente los científicos podían evaluar el trabajo de sus colegas, en miradas predominantemente internas, con reglas de acceso, exclusión y calidad determinadas de maneras autónomas (Chubin y Hackett, 1990), al creciente peso de formas de evaluación impuestas externamente (Alvesson y Spicer, 2016). Al comenzar el nuevo siglo, cada vez más, evaluaciones cuantitativas, basadas en indicadores cientométricos, en particular bibliométricos, tienden a complementar, y muchas veces a sustituir los mecanismos de evaluación por pares (Donovan, 2007; Weingart, 2005).

En ese contexto, existe una especie de disciplinamiento de la comunidad científica: surge un conjunto de prácticas, en individuos y en instituciones, como estrategias de supervivencia en las carreras académicas. Los artículos científicos resultantes de estas prácticas se han vuelto "documentos de disciplina", ponderan Alvesson y Spicer (2016), en los que el cumplimiento de reglas parece ser mucho más importante que tener algo relevante para comunicar. Estos y otros autores (Martin, 2016; Wilsdon *et al.*, 2015),

destacan la frecuente manipulación de los indicadores y prácticas de "*gaming*", como "*salami slicing*", reciclaje de artículos, autoplagio, publicación del mismo contenido en varios artículos, entre otros. Estas conductas resultan del desplazamiento de objetivos incitado por los sistemas de evaluación y distribución de recursos y prestigio: la búsqueda de altos puntajes en las medidas de desempeño se ha tornado un objetivo en sí mismo, más que un medio de medir si determinado nivel de desempeño ha sido alcanzado (Butler, 2007). O sea, los investigadores construyen estrategias de publicación y, por lo tanto, de investigación, de acuerdo con los indicadores utilizados (Wouters *et al.*, 2015). Además, la literatura identifica el fenómeno de "*cynical compliance*": aun siendo críticos de las formas de evaluación, los investigadores las cumplen, ya que es la forma de conseguir promociones, obtener financiamiento, etc. (Alvesson y Spicer, 2016).

De alguna manera, el camino seguido en estas décadas no ha hecho más que confirmar los "viejos" pronósticos, los "miedos" planteados frente a la construcción de los primeros sistemas de evaluación: un vaciamiento de los contenidos, una traslación hacia miradas sobre el desempeño de individuos e instituciones y no sobre la relevancia de los resultados obtenidos.

Entre las explicaciones posibles de esa legitimación ambigua, del disciplinamiento de las "masas de científicos", en primer lugar debemos señalar la masificación de científicos e instituciones, a lo largo del último medio siglo. Si en torno a la década de 1960 –cuando se da la consolidación de la evaluación por pares en agencias, instituciones y publicaciones periódicas– podía identificarse apenas un millón de científicos en el mundo (Salomon, 1996), hacia 1990 –cuando los indicadores bibliométricos ya estaban comenzando a desplazar a los pares– se contaban en torno a cinco millones, y hoy, probablemente, en un número cercano a diez; el último número manejado en el reporte periódico de Unesco es de 7,8 millones de personas a tiempo completo en el año 2013 (Unesco, 2015).

Por otra parte, en este siglo se quebró el monopolio de la continuadora de la ya mencionada primera empresa de indicadores bibliométricos, el Institute for Scientific Information de Eugene Garfield, transformado en la Thomson Corporation hacia fines de siglo pasado. Así, además de la Web of Science de esta empresa, actualmente tiene gran peso la Scopus de Elsevier, así como el Google Scholar y otras organizaciones de la misma índole (De Bellis, 2014). Estas organizaciones fueron sofisticando sus herramientas y las posibilidades de cuantificación en general, construyendo nuevos indicadores e índices –al estilo del índice *h*, propuesto por el argentino Jorge Hirsch en 2005 y ampliamente difundido actualmente en el mundo (Hirsch, 2005).

Algunos autores han mencionado como elemento relevante el impacto que tuvo, en la década de 1990, la irrupción del denominado New Public Management, el "gerencialismo" –como posible traducción de *managerialism*– en las universidades e instituciones de investigación, con su ideología de la calidad y el desempeño, con sus evaluaciones externas y sus *rankings* universales (Alvesson y Spicer, 2016; Goldreich, 2015; Gosselain, 2012).

La literatura del campo de estudios sociales de la ciencia y la tecnología habla de "nuevo contrato" entre ciencia y sociedad desde la década de 1990, al menos, hasta la fecha, con distintas versiones y matices (Cozzens *et al.*, 1990; Ziman, 1994; Gibbons *et al.*, 1994). Probablemente con base en esta literatura es que surge una mención emblemática –por su amplitud–, a fines de siglo, de la Conferencia Mundial sobre la Ciencia de Budapest de junio de 1999, donde se aprueba una idea de nuevo contrato social para la ciencia y la tecnología que implica una orientación del conocimiento científico-tecnológico al desarrollo y a las necesidades de las poblaciones, por un lado, y una mayor participación de otros actores en la orientación de las políticas y acciones, democratizando así el conocimiento (Unesco, 2000).

En esta última década, los documentos analizados rediscuten o reorientan este planteo de cambio de contrato en dos direcciones. Por un lado, aquellos que reclaman un retorno al "internalismo", a la torre de marfil, a la ciencia básica, al accionar de los pares en la toma de decisiones sobre cuestiones concernientes a la ciencia, como surge claramente en el documento Slow. De hecho, se habla de un retorno al idílico contrato antiguo, de mediados del siglo XX: más autonomía, menos incidencia de otros actores.

Por otro lado, en sentido opuesto, algunos manifiestos reclaman una reorientación radical de las agendas de investigación para servir a fines políticos, como es el caso de STEPS y Saving. De hecho, esto implica una profundización del mencionado contrato social del Siglo XXI. En una posición intermedia, moderada, se encuentra el documento Charte, que acepta que las agendas sean discutidas con actores no científicos, pero manteniendo siempre la autonomía de las decisiones.

En realidad, luego de estas décadas de avance en el predominio de las evaluaciones cuantitativas, con miradas evaluativas desde la burocracia, cabe preguntarse si los avances tímidos de incorporación de nuevos actores en la evaluación, que efectivamente se dieron en algunos espacios, se han mantenido y se pueden profundizar.

En suma, la mayor parte de las críticas a la evaluación de la ciencia realizada en los manifiestos son vino nuevo que rellena los viejos odres, aunque hay una clara inversión del peso de lo cualitativo –revisión por pares– a lo cuantitativo –indicadores bibliométricos–. Es evidente tam-

bién que algunas críticas, especialmente aquellas que señalan un deterioro de la calidad y relevancia de la ciencia producida, han ido adquiriendo más fuerza en los últimos años, conforme la escala y la eficiencia de aplicación de la evaluación cuantitativa se han ido ampliando y sus efectos sobre la dinámica de producción de conocimiento tienden a manifestarse más claramente.

Alvesson, M. y A. Spicer (2016), "(Un)Conditional surrender? Why do professionals willingly comply with managerialism", *Journal of Organizational Change Management*, vol. 29, Nº 1, pp. 29-45.

American Society for Cell Biology *et al.* (2012), "San Francisco Declaration On Research Assessment (DORA)". Disponible en <https://sfdora.org/read/>.

Arvanitis, R. e Y. Chatelin (1988), "National strategies in tropical soil sciences", *Social Studies of Science*, vol. 18, Nº 1, pp. 113-146.

Avalos, I. (1997), "El CONICIT: casa de pares e impares (o cómo no hay ideas equivocadas, sino extemporáneas)", en Sutz, J. (ed.), *Innovación y desarrollo en América Latina*, Caracas, CLACSO / AECI / Nueva Sociedad, pp. 151-162.

Butler, L. (2007), "Assessing university research: a plea for a balanced approach", *Science and Public Policy*, vol. 34, Nº 8, pp. 565-574.

Chubin, D. y T. Connolly (1982), "Research trails and science policies: local and extra-local negotiation of scientific work", en Elias, N., H. Martins y R. Whitley (orgs.), *Scientific establishments and hierarchies: sociology of the sciences*, vol. VI, Albany, Reidel Publishing Company.

Chubin, D. y E. Hackett (1990), *Peerless science: peer review and US science policy*, Albany, State University of New York Press.

Cozzens, S. *et al.* (orgs.) (1990), *The research system in transition*, Dordrecht, Kluwer Academic Publishers.

Davyt, A. y L. Velho (2000), "A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro?", *História, Ciências, Saúde-Manguinhos*, vol. 7, Nº 1, pp. 93-116.

De Bellis, N. (2014), "History and evolution of (biblio)metrics", en Cronin, B. y C. Sugimoto (eds.), *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*, Cambridge, The MIT Press, pp. 23-44.

Dickson, D. (1988), *The new politics of science*, Chicago, University of Chicago Press.

Donovan, C. (2007), "Introduction: Future pathways for science policy and research assessment: metrics vs peer review, quality vs impact", *Science and Public Policy*, vol. 34, Nº 8, pp. 538-542.

Etzkowitz, H. (1990), "The second academic revolution: the role of the research university in economic development", en Cozzens, S. *et al.* (orgs.), *The research system in transition*, Dordrecht, Kluwer Academic Publishers, pp. 109-124.

Fabbri, P. y B. Latour (1995), "La retórica de la ciencia: poder y deber en un artículo de ciencia exacta", en Fabbri, P., *Tácticas de los signos*, Barcelona, Gedisa, pp. 265-290.

Frame, J. D. (1980), "Measuring scientific activity in lesser developed countries". *Scientometrics*, vol. 2, Nº 2, pp. 133-145.

—— (1985), "Problems in the use of literature-based S&T indicators in developing countries", en Morita-Lou, H. (org.), *Science and technology indicators for development*, Boulder, Westview Press Inc., pp. 117-121.

Gibbons, M. *et al.* (1994), *The new production of knowledge: the dynamics of science and research in contemporary societies,* Londres, Sage [en castellano: Gibbons, M. *et al.* (1997), *La nueva producción del conocimiento: la dinámica de la ciencia y la investigación en las sociedades contemporáneas*, Barcelona, Pomares-Corredor].

Gilbert, N. (1978), "Measuring the growth of science: a review of indicators of scientific growth", *Scientometrics*, vol. 1, Nº 1, pp. 9-34.

Gläser, J. y G. Laudel (2007), "The social construction of bibliometric evaluation", en Whitley, R. y J. Gläser (eds.), *The changing governance of the sciences. The Advent of Research Evaluation Systems*, Dordrecht, Springer, pp. 101-123.

Goldreich, O. (2015), "Content-Oblivious Quality Measures and the Control of Academia", Department of Computer Science, Weizmann Institute of Science. Disponible en <http://www.wisdom.weizmann.ac.il/~oded/F/measures-en.pdf>.

Gosselain, O. (2012), "Slow Science et Désexcellence: Quelques Poches de Résistance En Belgique, Politique Des Sciences", Seminario Politique des Sciences, París, EHESS. Disponible en <https://pds.hypotheses.org/1968>.

Halffman, W. y H. Radder (2015), "The Academic Manifesto: From an Occupied to a Public University", *Minerva*, vol. 53, Nº 2, pp. 165-187.

Harnad, S. (1998), "Web matters: the invisible hand of peer review", *Nature - Web matters*, 5 de noviembre. Disponible en <https://www.nature.com/articles/nature28029>.

Hicks, D. *et al.* (2015), "The Leiden Manifesto for research metrics", *Nature*, vol. 520, Nº 7.548, pp. 429-431.

Hills, P. y A. Dale (1995), "Research and technology evaluation in the United Kingdom", *Research Evaluation*, vol. 5, Nº 1, pp. 35-44.

Hirsch, J. (2005), "An index to quantify an individual's scientific research output", *Proceedings of the National Academy of Sciences*, vol. 102, Nº 46, pp. 16569-16572. Disponible en <https://www.pnas.org/content/pnas/102/46/16569.full.pdf>.

Holbrook, J. (1992), "Why measure science?", *Science and Public Policy*, vol. 19, Nº 5, pp. 262-266.

Jagodzinski-Sigogneau, M., J-P. Courtial y B. Latour (1982), "How to measure the degree of independence of a research system?", *Scientometrics*, vol. 4, Nº 2, pp. 119-133.

L'Atelier des Chercheurs - Universidad Libre de Bruselas (2010), "Charte de la désexcellence", Bruselas, Universidad Libre de Bruselas.

Lindsey, D. (1978), *The scientific publication system in social science*, San Francisco, Jossey-Bass Publishers.

Link, A. (1998), "US and non- US submissions", *Journal of the American Medical Association*, vol. 280, Nº 3, pp. 246-247.

Manten, A. (1980), "Publication of scientific information is not identical with communication", *Scientometrics*, vol. 2, Nº 4, pp. 303-308.

Martin, B. R. (2016), "Editors' JIF-boosting stratagems - Which are appropriate and which not?", *Research Policy*, vol. 45, Nº 1, pp. 1-7.

—— y J. Irvine (1983), "Assessing basic research: Some partial indicators of scientific progress in radio astronomy", *Research Policy*, vol. 12, Nº 2, pp. 61-90.

Merton, R. K. (1968), "The Matthew Effect in Science", *Science*, vol. 159, Nº 3.810, pp. 56-63 [en castellano: "El efecto Mateo en la ciencia", en Merton, R. K. (1977), *La sociología de la ciencia*, Madrid, Alianza, pp. 554-578].

—— (1973), *The sociology of science: theoretical and empirical investigations*, Chicago, University of Chicago Press, "'Recognition' and 'Excellence': instructive ambiguities", pp. 419-438 [en castellano: Merton, R. K. (1977), *La sociología de la ciencia*, Madrid, Alianza, "'Reconocimiento' y 'Excelencia'. Ambigüedades instructivas", pp. 531-553].

Mitroff, I. y D. Chubin (1979), "Peer review at the NSF: a dialectical policy analysis", *Social Studies of Science*, vol. 9, Nº 2, pp. 199-232.

Narin, F. y M. Carpenter (1975), "National publication and citation comparisons", *Journal of the American Society for Information Science*, vol. 26, Nº 2, pp. 80-93.

Price, D. J. de Solla (1986a), "Measuring the size of science", en Price, D. J. de Solla, *Little Science, Big Science ...and beyond*, Nueva York, Columbia University Press, pp. 135-154.

—— (1986b), "Citations measures of hard science, soft science, technology and nonscience", en Price, D. J. de Solla, *Little Science, Big Science ...and beyond*. Nueva York, Columbia University Press, pp. 155-179.
—— (1978), "Editorial statements", *Scientometrics*, vol. 1, Nº 1, pp. 3-8.
Rabkin, Y. M. y H. Inhaber (1979), "Science on the periphery: a citation study of three less developed countries", *Scientometrics*, vol. 1, Nº 3, pp. 261-274.
Roy, R. (1984), "Alternatives to review by peers: a contribution to the theory of scientific choice", *Minerva*, vol. 22, Nº 3-4, pp. 316-328.
Salomon, J-J. (1996), "La ciencia y la tecnología modernas", en Salomon, J-J., F. Sagasti y C. Sachs (comps.), *Una búsqueda incierta: Ciencia, tecnología y desarrollo*, México, Editora de la Universidad de las Naciones Unidas / Fondo de Cultura Económica, pp. 49-86.
Sarewitz, D. (2016), "Saving Science", *The New Atlantis*, vol. 49, pp. 5-40.
Shapin, S. (1996), *The scientific revolution,* Chicago, University of Chicago Press [en castellano: Shapin, S. (2000), *La revolución científica. Una interpretación alternativa*, Barcelona y Buenos Aires, Paidós].
STEPS Centre - University of Sussex (2010), "Innovation, Sustainability, Development: a New Manifesto", Brighton, Universidad de Sussex.
The Slow Science Academy (2010), "The Slow Science Manifesto", Berlín, The Slow Science Academy.
Travis, G. y H. Collins (1991), "New light on old boys: cognitive and institutional particularism in the peer review system", *Science, Technology and Human Values*, vol. 16, Nº 3, pp. 322-341.
Unesco (2000), *Declaración sobre la Ciencia y el Uso del Saber Científico. La ciencia para el siglo XXI: Un nuevo compromiso*, París, Unesco.
—— (2015), *Unesco Science Report: towards 2030*, París, Unesco.
Van den Beemt, F. C. H. D. y A. F. J. van Raan (1995), "Evaluating research proposals", *Nature*, vol. 375, Nº 6.529, p. 272.
Velho, L. (1989), "Avaliação acadêmica: a hora e a vez do baixo clero", *Ciência e Cultura*, vol. 41, Nº 10, pp. 957-968.
Weinberg, A. M. (1963), "Criteria for scientific choice", *Minerva*, vol. I, Nº 2, pp. 159-171.
—— (1964), "Criteria for scientific choice II: the two cultures", *Minerva*, vol. II, Nº 1, pp. 3-14.
Weingart, P. (2005), "Impact of Bibliometrics Upon the Science System: Inadvertent Consequences?", *Scientometrics*, vol. 62, Nº 1, pp. 117-131.
Wenneras, C. y A. Wold (1997), "Nepotism and sexism in peer-review", *Nature*, vol. 387, Nº 6.631, pp. 341-343.

Wilsdon, J. *et al.* (2015), *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, Stoke Gifford, HEFCE.

Wouters, P. *et al.* (2015), *The Metric Tide: Literature Review. (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management)*, Stoke Gifford, HEFCE.

Ziman, J. (1994), *Prometheus bound: science in a dynamic steady state*, Cambridge, Cambridge University Press.

Zuckerman, H. y R. K. Merton (1973), "Institutionalized Patterns of Evaluation in Science", en Merton, R. K., *The sociology of science: theoretical and empirical investigations*, Chicago, University of Chicago Press, pp. 460-496 [en castellano: Merton, R. K. (1977), *La sociología de la ciencia*, Madrid, Alianza, "Pautas institucionalizadas de evaluación en la ciencia", pp. 579-621].