

El valor de tener muchas citas

ALEJANDRO G. FARJI-BRENER

Laboratorio Ecotono, CRUB-UNCOma. INIBIOMA, CONICET. Bariloche, Argentina.

“Es gracioso imaginarse a los músicos evaluados como actualmente lo somos los científicos. Una comisión de expertos contaría el número de canciones compuestas y evaluarían su calidad de acuerdo a la estación de radio adonde las pasan (estaciones, a su vez, valoradas por cuánta audiencia poseen) y por la cantidad de veces que la canción es reproducida. Rápidamente, este sistema vería progresar a aquellos autores de música más comercial por sobre aquellos compositores de otra clase de música. Lo que no es gracioso es que en el mundo real de la ciencia, los criterios dominantes y poco confiables de evaluación (como índices de impacto y número de citas de un artículo) están dominando las mentes, deformando comportamientos y determinando carreras”.

Peter Lawrence (2007)

[Palabras clave: evaluación académica, calidad de un manuscrito, factor de impacto]

[Keywords: academic evaluation, manuscript quality, impact factor]

Los investigadores somos evaluados constantemente para acceder a puestos estables de trabajo, para promover de categoría y para ganar subsidios que nos permitan equiparnos y financiar nuestras investigaciones. Pese a que la calidad de un investigador puede valorarse de maneras diferentes, la cantidad y la calidad de sus publicaciones siguen siendo los criterios más utilizados. Mientras que la cantidad de trabajos publicados es un parámetro sencillo (aunque poco revelador) de la productividad de un científico, la calidad de su producción es mucho más compleja de valorar de forma objetiva y práctica. En consecuencia, se han propuesto diferentes formas que estiman la calidad de los manuscritos. De estas formas, la más conocida es el valor de impacto de las revistas donde se publica (más conocido como ISI, que resume el número de citas de la revista). Sin embargo, el uso de estos índices es controversial por depender del área temática (sus rangos pueden variar órdenes de magnitud entre las diversas áreas de las ciencias) y, principalmente, por estimar la “calidad” de la revista y no del artículo, entre otros problemas (Seglen 1997; Moed 2002, 2005). Por eso, en muchas instancias de evaluación, una tendencia nueva y creciente es estimar la calidad de una publicación en función del número de veces que fue citada en otros artículos. Si bien antes era una tarea casi imposible de realizar, hoy este parámetro es sencillo de calcular gracias a diversas herramientas existentes en internet, como

www.scholar.google.com o scholarometer.indiana.edu, en donde se puede conocer el número de citas de cualquier manuscrito en cuestión de segundos. Pero ¿cuál es el valor de tener muchas citas? En este artículo presentaré algunos argumentos por los cuales considero que el número de citas de un manuscrito es un estimador poco confiable de su calidad científica.

CITA A CIEGAS: CUATRO RAZONES POR LAS CUALES EL NÚMERO DE CITAS ESTIMA MAL LA CALIDAD DE UN MANUSCRITO

Uno de los supuestos principales al utilizar la cantidad de citas como parámetro para evaluar la producción científica de un investigador es que un artículo más citado es de “mejor calidad” o “posee más impacto en la comunidad científica” que un artículo menos citado. Yo discrepo. La cantidad de citas no representa necesariamente la calidad de un artículo ni su nivel de aporte a la comunidad científica por varias razones. Las principales son: a) por lo general, los artículos son más citados por su pertinencia que por su valor intrínseco, b) este indicador sobrevalora trabajos de revisión por sobre artículos que desarrollan temáticas particulares, independientemente de la calidad de ambos, c) el impacto de un artículo sobre la comunidad científica no necesariamente se ve reflejado en cuántas veces dicho trabajo fue citado, y d) una

gran cantidad de veces se citan artículos de manera inadecuada o trabajos que ni siquiera son leídos. Estos motivos, que ampliaré a continuación, ponen en duda la efectividad de utilizar el número de citas que posee un artículo como una herramienta adecuada al momento de evaluar la producción de un científico.

Primero, los investigadores no citamos artículos porque son buenos, sino porque representan ejemplos adecuados para avalar o refutar hipótesis, aclarar o ejemplificar ciertos aspectos metodológicos o porque aportan información apropiada para la introducción o la discusión de cierta problemática (Lawrence 2007). Ninguno de estos argumentos implica una valoración sobre la calidad del artículo citado, sino simplemente su pertinencia para cierto propósito didáctico dentro de la estructura de nuestro manuscrito. Una argumentación en contra de esta idea sería que, dada la gran producción de artículos científicos que se pueden enmarcar en la temática que vamos a desarrollar, el criterio de excelencia se utiliza dentro un subconjunto de manuscritos previamente seleccionados por su pertinencia. Sin embargo, este razonamiento posee, al menos, dos debilidades: a) sostiene que el primer filtro de selección es la pertinencia y no la calidad. Eso implica que, posiblemente, han quedado fuera de ser citados artículos de alta calidad porque resultaron poco pertinentes al tema sobre el cual estamos escribiendo en ese momento, y b) el supuesto de que podemos acceder a todos los artículos pertinentes, para luego hacer una selección más fina y citar a los mejores, es irreal. La cantidad apabullante de literatura sobre cualquier tópico inhibe a un investigador acceder y leer todos los trabajos de cierta temática. Imaginémos que me propongo escribir un manuscrito sobre competencia entre plantas. Si utilizo www.scholar.google.com y escribo "plant competition" como palabra clave, obtengo 1500000 resultados en 0,11 s (sin incluir patentes ni citas). En el mundo real ¿quién puede acceder o leer al menos una porción significativa de esos artículos para luego seleccionar los mejores? Si considero una temática más específica (e.g., escribo "ant-plant interactions" como palabra clave) encuentro sólo 3680 resultados en 0,18 s, una cifra mucho menor pero también imposible de evaluar en su conjunto para seleccionar a los mejores manuscritos. Lo que sucede es que, en el mundo real, el azar, el conocimiento

de ciertos autores, el acceso a ciertas revistas por sobre otras y otros elementos ajenos a la calidad de un artículo limitan o sesgan nuestro acceso a la información. En consecuencia, la gran mayoría de las veces los trabajos se citan más por su pertinencia que por su calidad, inclusive si nos limitamos a considerar exclusivamente el sub-conjunto de los trabajos más "pertinentes".

El segundo motivo por el cual considero a las citas como una herramienta inadecuada para evaluar la calidad de un manuscrito es que este indicador sobrevalora los trabajos de revisión por sobre aquellos de temáticas más específicas. Debido a la limitación de espacio o restricciones en el número de referencias que imponen las revistas, es común que para ilustrar cierto tópico se prefiera citar un trabajo de revisión o meta-análisis en vez de muchos trabajos que describan situaciones, organismos o sistemas específicos (Lawrence 2007). Por ejemplo, el artículo "The nature and consequences of indirect effects in ecological communities" (Wootton 1994) ha sido citado hasta la fecha unas 561 veces. Sin embargo, el artículo "Odour learning and decision-making during food collection in the leaf-cutting ant *Acromyrmex lundii*" (Roces 1994), sólo 52 veces (www.scholar.google.com). El primer artículo ¿es 10 veces mejor que el segundo? El primero detalla y resume de forma integral la importancia de los efectos indirectos en las comunidades ecológicas, sin duda un tema de relevancia. Pero el segundo, aunque está enfocado en un organismo específico, es un artículo excelente que presenta un diseño experimental elegante y aporta descubrimientos novedosos con implicancias generales. Quien haya leído ambos trabajos encontrará muy difícil elegir sólo uno de ellos sobre la base de sus virtudes académicas. Sin embargo, al emplear el número de citas para evaluar a sus autores podríamos llegar a la conclusión errónea que J.T. Wootton es un "mejor" científico que F. Roces, o que el primero merece más un ascenso, un subsidio o un mejor empleo que el segundo. Esta diferencia en el número de citas de un orden de magnitud corresponde, simplemente, a que un artículo de revisión puede servir de ejemplo o apoyo a muchos más trabajos que uno que trata un tema más específico. Sin embargo, los artículos de revisión (como los artículos de cualquier otra temática) pueden ser buenos, regulares o malos. El hecho de que un artículo sea de revisión no

lo convierte "ipso facto" en un buen artículo, ni automáticamente en un aporte más valioso que un buen trabajo que pone a prueba alguna hipótesis particular, o que describe algún sistema, problemática u organismo específico. Más paradójico aún, al premiar en exceso a los artículos de revisión por ser más citados se desestimularía la investigación y la escritura sobre problemáticas particulares, cuando éstos últimos trabajos son imprescindibles para la existencia de los primeros. Es verdad que los trabajos de revisión pueden ofrecer una síntesis del estado del arte de un área específica y constituir un aporte relevante al conocimiento. Mis argumentos no pretenden desestimar la importancia de los artículos de revisión. Sencillamente, pretenden ejemplificar cómo ciertos manuscritos corren con ventaja a la hora de considerar a las citas como indicadores de calidad, independientemente de su valor intrínseco.

Tercero, los artículos más citados no necesariamente son los que más impacto poseen sobre la comunidad científica. Si definimos impacto como la capacidad de un artículo de modificar de alguna manera las ideas o el desarrollo de las actividades académicas de sus lectores (Bornman et al. 2008), veremos que artículos relativamente poco citados pueden tener un gran impacto. Voy a ilustrar este punto con un ejemplo personal. Al solicitar una promoción en la Carrera de Investigador Científico del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) de Argentina, existe un ítem para completar en donde se deben mencionar las publicaciones propias que uno valora como las más trascendentes de los últimos cinco años. Obviamente, yo elijo poner aquellas publicadas en revistas de alto impacto, obviando a las que considero muy buenas, pero por diversas razones han sido publicadas en revistas "menos importantes". Debo confesar que nunca he puesto el artículo de mi producción que considero de mayor impacto. Las razones son relativamente obvias: dicho artículo está publicado en *Ecología Austral* (una revista de las "ligas menores") y trata sobre ciertos vicios en el uso de los términos hipótesis y predicciones en *Ecología* (Farji-Brener 2003). Por ese artículo he recibido decenas de correos con reconocimientos de estudiantes y profesionales completamente desconocidos por mí, y sé que se emplea como material didáctico en muchos cursos de grado y posgrado en Argentina y otros países de

habla hispana. Sin embargo, este artículo posee sólo 9 citas según www.scholar.google.com. Pese a que muchos lo usan, pocos lo citan. En consecuencia, el número de citas de este artículo no representa su real impacto en la comunidad académica. Quizás algunos argumentarán que esto no se aplica a manuscritos estrictamente "científicos". Varias evidencias sugieren lo contrario. En general, los trabajos novedosos necesitan de un tiempo largo para ser aceptados por la comunidad científica y para tener citas acordes a su nivel. Por ejemplo, tanto el trabajo más relevante del siglo XX sobre Biología (el descubrimiento de la doble hélice del ADN) como el artículo fundamental por el cual E.B. Lewis ganó el premio Nobel en 1995 (Lewis 1978) fueron raramente citados en los primeros diez años posteriores a su publicación (Olby 2003). Por lo tanto, el número de citas no necesariamente representa ni su calidad ni su real impacto sobre la comunidad científica, tanto en artículos relacionados con la enseñanza de las ciencias (pero con particular impacto sobre cómo hacer investigación) como en aquellos estrictamente científicos.

Finalmente, la cuarta razón por la cual el número de citas es un mal estimador de la calidad de un manuscrito es un conjunto de motivos ya documentados, que resumo a continuación. Por un lado, es raro que todos los artículos citados sean leídos por completo o citados de forma apropiada (Lawrence 2007). Por ejemplo, se ha establecido que sólo 20% de los trabajos citados han sido leídos (Simkin & Roychowdhury 2003). Una encuesta rápida en mi lugar de trabajo apoya esa idea: los autores leen completas sólo $82 \pm 16\%$ (promedio \pm SD) de las citas que colocan en sus propias publicaciones ($N=26$ encuestados). Además, otro estudio sugiere que 24% de las referencias son citadas de manera inapropiada (Todd et al. 2007). Entonces, ¿cuál es el valor de emplear el número de citas que posee un artículo como sinónimo de su calidad cuando casi la mitad de las mismas provienen de autores que no leyeron el artículo citado o que las mismas fueron utilizadas inadecuadamente? Por otra parte, e independientemente de los argumentos discutidos hasta el momento, se ha demostrado que el número de citas de un artículo depende de factores por completo ajenos a su calidad. Por ejemplo, son mucho más citados los artículos que apoyan las interpretaciones del autor que los cita que aquellos que no, aquellos manuscritos de autores de países

de habla inglesa que los de autores de otros países (aunque el idioma del manuscrito sea el inglés) y los trabajos de muchos autores respecto de trabajos de solo uno (Leimu & Koricheva 2005). En particular, el que se citen más los artículos cuyos resultados apoyan las hipótesis del trabajo que los cita refuerza aun más el argumento de que la pertinencia (en este caso, convencer a los lectores de la validez del estudio en vez de simplemente agradecer una fuente de información) es más relevante que la calidad al momento de citar un artículo. La cantidad de citas también depende del año de publicación del artículo, ya que manuscritos más antiguos poseen una mayor probabilidad de ser citados que aquellos más actuales; y del comportamiento coercitivo de algunos editores, quienes de forma sutil exigen a los autores que citen trabajos de la misma revista donde se pretende publicar (Wilhite & Fong 2012). Finalmente, puede existir una retroalimentación positiva entre algunos grupos de autores o regiones que se citan mutuamente en detrimentos de otros autores o regiones, más allá de la calidad real de los artículos o de su impacto potencial sobre un área determinada del conocimiento. Por ejemplo, los autores norteamericanos citan autores de su misma nacionalidad 32% más que lo esperado por azar (Pasterkamp et al. 2007). Todos estos sesgos ponen en duda, por sí mismos, la validez del uso del número de citas como un buen estimador de la calidad o impacto de una publicación científica.

ERRAR ES HUMANO. CITAR ¿ES DIVINO?

Imaginémonos un mundo en el que a los escritores se los evaluara sólo por la editorial donde publican sus obras (editoriales calificadas, a su vez, por la cantidad de ventas) o por la cantidad de lectores que leen sus libros. O un mundo en donde a los actores se los valorara por la capacidad de espectadores del teatro adonde actúan o por la cantidad de público que presencia sus obras. O un mundo en el cual los directores de cine fuesen principalmente juzgados según el estudio cinematográfico donde filmaron (estudios valorados, a su vez, por la cantidad de películas producidas por año) y por el tiempo que sus películas se mantuvieron en cartel. ¿Qué tipo de escritores, actores y directores de cine se verían beneficiados y cómo estos beneficios impactarían sobre la escritura, la actuación y la cinematografía de un país? En el mundo de la ciencia no hace

falta usar tanta imaginación: la tendencia actual de utilizar parámetros rápidos e inapropiados para valorar el desempeño individual de un investigador ha logrado que, por lo general, los científicos primero piensen en alcanzar rankings altos y luego en hacer buena ciencia (Bergstrom 2010; Farji-Brener & Ruggiero 2010).

En este trabajo he presentado argumentos que hacen dudar de la efectividad del número de citas que posee un artículo como una herramienta válida para evaluar su calidad, su nivel de impacto o el mérito académico del autor. El número de citas es inapropiado porque: a) depende más de la pertinencia que de la calidad, b) depende del tipo de trabajo (revisiones versus estudios de caso) y del tipo de evidencias que aporta (a favor o en contra de las hipótesis del artículo que lo cita), c) es afectado por motivos que nada tienen que ver con la calidad de una investigación, como el número y filiación de los autores, y d) muchas veces, las citas son inapropiadas o se refieren a trabajos no leídos. En consecuencia, dado el sesgo de utilizar el número de citas de un artículo como estimador de su calidad o impacto, es pertinente debatir cuál es la consecuencia de su empleo en los procesos de evaluación y en cuáles circunstancias su uso posee un mayor efecto negativo sobre el sistema científico.

Al emplear el número de citas como indicador de la calidad o del impacto potencial de un artículo podemos, como en la estadística inferencial, cometer dos tipos de errores: rechazar algo supuestamente verdadero o aceptar algo supuestamente falso. En este caso, podemos considerar un artículo con muchas citas como de alta calidad o gran impacto cuando en realidad es de baja calidad o bajo impacto (i.e., sobrevaloración o falso positivo), o considerar de baja calidad o poco impacto un artículo pobremente citado cuando realmente es de alta calidad o posee un buen impacto sobre la comunidad científica (i.e., subvaloración o falso negativo) (Figura 1). La pregunta clave es si ambos tipos de errores poseen la misma relevancia al momento de evaluar investigadores y si hay circunstancias que convierten a un tipo de error más "peligroso" que otro. Por un lado, podemos suponer que la sobrevaloración (i.e., falsos positivos) premia a investigadores que en verdad no lo merecen. Esto significa que este tipo de error "permite" la obtención de becas, trabajos estables, promociones o subsidios

a investigadores que no necesariamente representan lo mejor de su cohorte. En un ambiente sin limitación de recursos, este error no parece tan nocivo para el sistema científico. Sin embargo, la subvaloración puede dejar relegados o afuera del sistema a investigadores prometedores, lo cual podría tener consecuencias más negativas que el caso anterior. En particular, ante una baja oferta de becas, ingresos a carrera de investigador, promociones o subsidios, resulta más perjudicial dejar científicos promisorios relegados o fuera del sistema que incluir o premiar aquellos científicos menos prometedores. En otras palabras, ante una limitación de recursos pareciera más perjudicial para el sistema científico cometer el error de subvalorar al usar el número de citas de un artículo para evaluar el desempeño académico de su autor.





		CITAS	
		Muchas	Pocas
CALIDAD O IMPACTO POTENCIAL	Alto	No error 	Falso negativo 
	Bajo	Falso positivo 	No error 

Figura 1. Diferentes tipos de error que podemos cometer al considerar el número de citas de un artículo como un indicador de su calidad o impacto potencial. Se puede considerar un artículo muy citado como de alta calidad o impacto cuando realmente es de pobre calidad y/o de bajo impacto (sobreevaluación o falso positivo), o considerar a un artículo poco citado como de baja calidad o de poco impacto cuando en verdad es de buena calidad y gran impacto (subvaloración o falso negativo).

Figure 1. Different kinds of possible errors considering the number of citations of a paper as an indicator of its quality and/or potential impact. Basically, a paper with a high number of citations may be qualified as a great impact or high-quality paper when it really is a poor-quality and/or low impact one (i.e., over-qualification or false positive), or may be qualified as a poor impact and/or low-quality paper when it really is a high-quality and/or low impact one (i.e., under-qualification or false negative).

EL EQUILIBRIO ENTRE EL MUNDO IDEAL Y EL MUNDO REAL

Es indudable que los científicos deben ser evaluados para obtener becas, ingresar o promover a mejores trabajos y acceder a subsidios. Para ello, es necesario emplear estimadores de calidad que sean sencillos y eficientes, pero a la vez reales y poco sesgados. Los estimadores actuales, que

pretenden cuantificar de forma reduccionista la calidad de un artículo (e.g., el número de citas) no sólo han sido criticados por ser malas herramientas evaluativas (Moed 2002, 2005; Bornmann et al. 2008), sino por estimular comportamientos ajenos al real quehacer científico (Lawrence 2007; Bergstrom 2010). Su uso, quizás adecuado para ciertas circunstancias, ha sido sobreaprovechado por su relativa facilidad de cálculo para evaluar la producción individual de los investigadores. Tal vez sea hora de que los comités de evaluación consideren que los evaluados son más un conjunto de habilidades (en el cual la originalidad debería tener un peso relevante) que un valor en un ranking (Lawrence 2007). Sin embargo, un miembro de las comisiones de CONICET, que debe evaluar en pocos días 40 solicitudes, cada una con una producción científica promedio de 50 publicaciones, pensará que el planteo de lectura directa de los principales artículos de los candidatos como única herramienta de evaluación es ingenuo e irreal. El arribo a una solución justa ante grandes números de evaluados y poco tiempo para decidir posiblemente exija cuantificar. Yo no propongo la anulación absoluta de los indicadores como una herramienta posible para evaluar grandes números de candidatos. Sin embargo, considero que deberían emplearse sólo cuando sea estrictamente necesario, previo conocimiento y ponderación de los sesgos que estos indicadores presentan. Estas herramientas de evaluación indirecta deberían obligatoriamente complementarse con la prehistórica pero aun útil práctica para valorar la calidad de un manuscrito: su lectura. Dentro de este contexto, el número de citas que posee un artículo (así como otros indicadores controversiales) dejarían de tener tanto valor como estimadores de la producción científica individual.

AGRADECIMIENTOS: Los comentarios y opiniones de Esteban Jobbágy y un revisor anónimo me ayudaron a mejorar la calidad de este manuscrito.

BIBLIOGRAFÍA

BERGSTROM, CT. 2010. Use ranking to help search. *Nature*, 465:870.

BORNMAN, L; R MUTZ; C NEUHAUS & H DANIEL. 2008. Citation counts for research evaluation: standards of good practice for analyzing bibliometrics data and presenting and interpreting results. *Ethics Sci. Environ. Polit.*, 8:93-102.

FARJI-BRENER, AG. 2003. Uso correcto, parcial e incorrecto de los términos hipótesis y predicciones en ecología. *Ecología Austral*, 13:223-227.

- FARJI-BRENER, A G & A RUGGIERO. 2010. ¿Impulsividad o paciencia? Qué estimula y qué selecciona el sistema científico argentino. *Ecología Austral*, **20**:307-314.
- LAWRENCE, P. 2007. The mismeasurement of science. *Current Biology*, **15**:R583-5.
- LEIMU, R & J KORICHEVA. 2005. What determines the citation frequency of ecological papers? *Trends in Ecology and Evolution*, **20**:28-32.
- LEWIS, EB. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature*, **276**:565-570.
- MOED, HF. 2002. The impact-factor debate: the ISI's uses and limits. *Nature*, **415**:731-732.
- MOED, HF. 2005. *Citation Analysis in Research Evaluation*. NY Springer
- OLBY, R. 2003. Quiet debut for the double helix. *Nature*, **421**:402-405.
- PASTERKAMP, P; J ROTMANS; D KLEIJN & C BORST. 2007. Citation frequency: a biased measure of research impact significantly influenced by the geographical origin of research articles. *Scientometrics*, **70**:153-165.
- ROCES, F. 1994. Odor learning and decision-making during food collection in the leaf-cutting ant *Acromyrmex lundi*. *Insectes sociaux*, **41**:235-239.
- SEGLEN, PO. 1997. Why the impact factor of journals should not be used for evaluating research. *Brit. Med. J.*, **314**:497-511.
- SIMKEN, MV & VP ROYCHOWDHURY. 2003. Read before you cite! *Complex Syst.*, **14**:269-274.
- TODD, P; CJ DARREN; D LI & R LADLE. 2007. Citing practices in ecology: can we believe our own words? *Oikos*, **116**: 1599-1601.
- WILHITE, A & E FONG. 2012. Coercitive citation academic publishing. *Science*, **335**:542-543.
- WOOTTON, JT. 1994. The nature and consequences of indirect effects in ecological communities. *Annual Review of Ecology and Systematics*, **25**:443-466.