



Fraude en las ciencias: para el fortalecimiento de las revistas científicas (promoviendo actitudes, prácticas y comunidades científicas)

Carlos Guerrero de Lizardi

Resumen

El fraude en las ciencias, en todas, está ampliamente documentado. Se trata de mejorar la maquinaria de producción y difusión de la ciencia que actualmente arroja, en México y el mundo, productos que se presentan como científicos pero que no lo son. Reviso las políticas de las revistas, y en general las métricas de los sistemas de ranqueo. Argumento que debemos modificar nuestras ideas acerca de la evaluación de los investigadores, las revistas y los posgrados. En breve, el sistema estimula malas prácticas. Esbozo algunas ideas en torno a las políticas públicas requeridas para promover mejores prácticas por parte de la comunidad científica.

Palabras clave: fraude en las ciencias; replicación; métricas de evaluación; nuevas políticas.

Abstract

Fraud in science is documented. It is about improving the machinery for the production and dissemination of science that currently produces, in Mexico and the world, products that present

themselves as scientific but are not. I review the journal policies, and in general the metrics of the ranking systems. I argue that we must modify our ideas about the evaluation of researchers, journals, and postgraduates. In short, the system encourages bad practices. I outline some ideas around the public policies required to promote best practices by the scientific community.

Keywords: fraud in science; replication; evaluation metrics; new policies.

Introducción

El fraude en las ciencias en general (Crocker y Cooper 2011), y en la economía en particular (Duvendack, Palmer-Jones y Reed 2015), está ampliamente documentado. En cualquier universidad en el mundo algunos tantos profesores investigadores, con toda intención o como producto de malas prácticas científicas, intentan hacer pasar resultados no científicos como científicos. Y en muchas ocasiones lo logran.

A propósito, quiero señalar que dejaré fuera de mi análisis un "pecado menor", a saber, el plagio, y las diez modalidades de trabajo poco original.

Las prácticas no científicas se observan fuera y dentro de la academia. Un ejemplo relativo a los institutos de estadística lo escribieron Lequiller y Blades (2014, 44) en su libro publicado, en segunda edición, por la OCDE:

National accounts could better be called 'national accounts statistics' because without this qualifier users may think they are as reliable as the business accounts of a company. This is not true. In particular, while GDP for technical reasons is often expressed in millions of units of the National currency, users should be aware that they are very, very far from being accurate at the level of millions [...] It is not even possible to give a summary figure of the accuracy of the GDP. Indeed, national accounts, and in particular GDP, are not the result of a single big survey for which one might compile a confidence interval. They are the result of combining a complex mix of data from many sources, many of which require adjustment to put

them into a national accounts database and which are further adjusted to improve coherence, often using non-scientific methods.

Promoviendo actitudes, prácticas y comunidades científicas

La línea de defensa contra el fraude en las ciencias es una y solo una: la replicación –o regla de oro de las ciencias, según la Academia de Ciencias de los Estados Unidos–. Cualquier producto que busque la etiqueta de científico debe aprobar el ejercicio de replicación.

En el caso de las revistas científicas por el momento vivimos en el peor de los mundos. La pieza clave, esto es, el proceso de dictaminación, es deficiente. Como autor, miembro de comités, editor, y coordinador académico de programas PNPC, me queda claro que en vistas a su mejoramiento el primer paso burocrático-formal-meramente-cuantitativo consistiría en que el trabajo de dictaminación fuese reconocido por los programas de estímulos internos y externos; decisión, digamos, rápida de tomar, pero que no se ha instrumentado, por ejemplo, por el Sistema Nacional de Investigadores (SNI) del CONAHCYT.

El segundo paso, encaminado al genuino mejoramiento de la calidad del proceso de dictaminación, implicaría modificar significativamente las conductas de los agentes en juego, ni más ni menos. Berk *et al.* (2017, 231-2) identificaron algunas malas prácticas:

[...] referees feel the need to demonstrate their intelligence or industriousness to editors by identifying problems in papers. The result is that in many cases reviewers inflate minor blemishes to the status of major flaws... Another cause for concern is the level of disagreement amongst referees. As Welch (2014) documents in a study of eight prominent journals in economics and finance, in cases with multiple referees on the same paper, the probability that one referee recommends inviting a revision rather than rejecting, conditional on another referee doing so, is only marginally higher than the unconditional probability.

Furthermore, the correlation of referee recommendations at a major finance conference was only 0.28. This pattern of disagreement suggests a high level of arbitrariness in the review process.

Berk *et al.* (2017) de hecho trataron de explicar "racionalmente" la actitud del dictaminador para justificar su sesgo. Yo creo simplemente que se trata de un caso de empoderamiento dañino (llamado problema de agente-principal en Economía). Ahora bien, desde el sentido común, el objetivo y la solución están a la mano: el proceso de dictaminación y replicación debería servir, en primer lugar, para identificar y rechazar productos no científicos, y en segundo, para mejorar a los genuinamente científicos; pero en la práctica su instrumentación es imposible dadas las reglas del juego vigentes –que, dicho sea de paso, dan forma a los incentivos "perversos" de los agentes inmiscuidos en la producción y difusión de la ciencia–. Más adelante seguiremos hablando de esto.¹

El reto es gigantesco, y al momento la comunidad científica nuestra (revistas, profesores-investigadores, el Sistema Nacional de Investigadores y sus cuerpos colegiados –en el caso de la UNAM, el sistema de estímulos PRIDE–, y las propias burocracias universitarias y no universitarias) no cuenta, pienso, con la templanza para dar los primeros pasos en la dirección correcta. Simplemente todos somos engranes de una maquinaria que no

1 Mi caso emblemático es el siguiente. En 2003 Clive Granger recibió el Nobel de Economía por su aportación conocida como análisis de cointegración (que continúa representando la frontera del conocimiento en el análisis de series de tiempo). Su documento seminal fue rechazado repetidamente por la revista emblemática bajo la dirección de Angus Deaton, quien también ya fue galardonado con el Nobel, en otro año, en 2015. Granger (2010, 3) mismo platicó las vicisitudes para publicarlo: "*Econometrica* rejected the paper for various reasons, such as wanting a deeper theory and some discussion of the testing question and an application. As I knew little about testing I was very happy to accept Rob's offer of help with the revision. I re-did the representation theorem and he produced the test and application, giving a paper by Granger and Engle which evolved into a paper by Engle and Granger, whilst I was away for six months leave in Oxford and Canberra. This new paper was submitted to *Econometrica* but it was also rejected for not being sufficiently original. I was anxious to submit it to David Hendry's new econometrics journal but Rob wanted to explore other possibilities. These were being considered when the editor of *Econometrica* asked us to re-submit because they were getting so many submissions on this topic that they needed our paper as some kind of base reference."

se detiene, en México y en el mundo. Y repito, como economista, las causas me saltan a la vista: las reglas del juego vigentes, que estructuran el sistema de estímulos y de ranqueos, distorsionando el comportamiento de los agentes involucrados.

Mis palabras están basadas en mi experiencia como autor, miembro de comités, editor, y coordinador académico de programas PNP. Y probablemente estoy equivocado en la medida que la realidad es mucho más compleja que mi mirada. Dicho lo anterior no veo cómo entender el proceder de un científico quien se atrevió a falsear sus resultados por décadas, contaminando en el camino a un buen número de tesis, coautores, revistas y a las universidades involucradas. Todo lo anterior refiere el caso documentado por Crocker y Cooper (2011). En Economía ocho de cada diez *papers* no pasan la prueba de la replicación (Duvendack, Palmer-Jones y Reed 2015). E intuía que todo lo anteriormente escrito es apenas la punta del *iceberg* respecto a comportamientos no científicos.

Debemos modificar sustancialmente nuestras ideas acerca de la evaluación de los profesores investigadores, de las revistas, de los programas de posgrado, y de las burocracias (CONAHCYT, UNAM, BANXICO, INEGI, SHCP, SE, CONEVAL, etc.). Para valorar correctamente el papel de la medición de resultados en general, y de los productos científicos en particular, propongo como telón de fondo un par de citas de Muller (2018):

Used properly, measurement, as we'll see, can be a good thing... (But measurement) can also distort, divert, displace, distract, and discourage. While we are bound to live in an age of measurement, we live in an age of mismeasurement, over-measurement, misleading measurement, and counter-productive measurement... And measurement may provide us with distorted knowledge –knowledge that seems solid but is actually deceptive. (241-7)

For if doctors or hospitals are remunerated by government agencies or private insurers based on their success rates in keeping patients alive, then such measurements should create incentives for better care... What could go wrong? A good deal, as we have already seen. When their scores are used as a basis of reward and

punishment, surgeons, as do others under such scrutiny, engage in creaming, that is, they avoid the riskier cases. When hospitals are penalized based on the percentage of patients who fail to survive for thirty days beyond surgery, patients are sometimes kept alive for thirty-one days, so that their mortality is not reflected in the hospital's metrics. In England, in an attempt to reduce wait times in emergency wards, the Department of Health adopted a policy that penalized hospitals with wait times longer than four hours. The program succeeded—at least on the surface. In fact, some hospitals responded by keeping incoming patients in queues of ambulances, beyond the doors of the hospital, until the staff was confident that the patient could be seen within the allotted four hours of being admitted. (252-63)

En breve, ante un proceso de medición-evaluación, los agentes ajustan su comportamiento para satisfacer sólo formalmente los criterios utilizados por sus certificadores, perdiendo de vista (distorsionando, o peor, echando a la basura) los objetivos genuinos-sustanciales del proceso en juego (llámese ranqueo de una universidad, programa de posgrado PNPC, revista, o miembro del SNI).

Para hacer efectiva la replicación los autores deberían compartir todas y cada una de las decisiones tomadas en el proceso de producción de conocimiento, así como todos los insumos utilizados. Todo lo anterior no sólo mejoraría sustancialmente el contenido de las publicaciones sino, de más trascendencia, representaría un genuino acto de comunidad científica. Pero los autores no están dispuestos a hacerlo, y por razones obvias. A este comportamiento se le llama, en Economía, "autoselección" o "autoexclusión". Esta autoexclusión es un reconocimiento explícito de que sus resultados son, en algún grado, fraudulentos. A propósito, desde la producción y difusión del conocimiento aquí salta a la vista otra característica de la Ciencia Abierta.

Por el lado de las revistas se debería contar con un equipo técnico que sea capaz de acompañar a los dictaminadores para realizar la replicación y, así, corroborar los resultados aducidos por los autores. La replicación permitiría mejorar sustancialmente los documentos publicados, y aún otorgarles un "sello de garantía". Ninguna revista en México tiene los recursos para hacer-

lo, y en el mundo de la economía sólo conozco a unas pocas (dicho sea de paso, de varios cientos de revistas) que ya caminan por esta vereda, por lo menos en el terreno discursivo (que no efectivo). Tampoco los directores de revistas están dispuestos a hacerlo, entre otros motivos porque las revistas sirven a acotadas comunidades "científicas" en México y el mundo. En la misma dirección, las decisiones de qué y a quién publicar reflejan francamente criterios extraacadémicos, entre otros, la ideología y la pertenencia a redes.

Por el lado del SNI los pares deberían reconocer que los productos científicos son variopintos, desde las revisiones de la literatura hasta las aportaciones teóricas, pasando por los productos aplicados y un largo etcétera según la ciencia de la que estemos hablando. El SNI debería ponderarlos en su justa medida ya que un buen documento –aquel que aporta genuinamente a la ciencia en alguna de las direcciones comentadas– puede representar años de trabajo comprometido. Algo similar ocurre en el sistema de estímulos de la UNAM, entre muchos otros. Pero el actual criterio es de cantidad (decimos, coloquialmente, "por kilo") y no de calidad, en México y el mundo.

A propósito, las agencias internacionales de clasificación de universidades, los cuestionarios levantados por algunos periódicos en México, y los propios repositorios e índices de revistas (nacionales e internacionales) tampoco cumplen cabalmente con su trabajo. Son agentes interesados que forman parte, en algún sentido, de la comunidad científica. La más reciente compra-venta del WoS ascendió a ¡3,500 millones de dólares! Y las universidades europeas desde hace años enfrentan a su propio pulpo omnipresente Elsevier-Scopus –convertido en repositorio, editor de revistas e indizador–.

La clasificación, en el mundo y en nuestros países, de profesores-investigadores, de revistas, de posgrados, y de las propias universidades, gira en torno a la producción científica. Y de la clasificación obtenida dependen prestigios, ingresos, becas, y presupuestos. Y así, en esa medida, todos nos convertimos en engranes de esa maquinaria ligada a la producción y difusión del conocimiento.² Pero la suerte no está echada.

2 Al respecto Afonso (2014) del King's College de Londres escribió un documento titulado "How academia resembles a drug gang", y Guerrero (2022) otro titulado "Desenmascarando a la autonomía: una estimación del coeficiente de Gini puma para el año 2020".

Breve nota teórica-histórica sobre el origen de la evaluación de las políticas públicas

El estudio de las políticas sociales tiene un fuerte componente ideológico. A muchos científicos se les ha enseñado que su sentido se limita a reparar las "fallas de mercado", y que su evaluación es imprescindible por las, aún peores, "fallas de gobierno" (Weimer y Vining 2005). Se parte así del supuesto de que el funcionario público realiza su trabajo deficientemente, y que más vale que sea sometido a evaluaciones externas realizadas por organismos autónomos. Esta ideología ha permeado en muchos profesionistas y agentes sociales y políticos.

Dicho lo anterior debemos señalar que la evaluación externa, y tanto o más importante, la autoevaluación, representan piezas clave de los programas sociales y de las políticas de desarrollo social (Aguilar 2003a, 2003b, 2003c). Puesto con otras palabras, las evaluaciones deben contribuir genuinamente al mejoramiento de los programas y las políticas, y la medición no es un componente imprescindible en todos los casos. Baste recordar que los agentes ajustan su comportamiento a las (imperfectas) evaluaciones (externas) por lo que su efecto puede ser inocuo y, peor aún, contraproducente.

Propuesta de políticas públicas para promover actitudes, prácticas y comunidades científicas

Se trata de mejorar la maquinaria de producción y difusión de la ciencia que actualmente arroja, en México y el mundo, productos que se presentan como científicos pero que no lo son. Esta maquinaria está hecha de piezas (las universidades públicas y privadas, los centros de investigación, las revistas, el CONACYT y muchas otras agencias públicas, etc.), de engranes (los profesores-investigadores, los editores de revistas, los comités de pares, los administradores, etc.), y tiene reglas (entre otras, las que alteran negativamente nuestros comportamientos como profesores-investigadores). Para lograrlo se requiere de una política de estado que promueva la replicación de cualquier pro-

ducto que busque la etiqueta de científico, originada en todas y cada una de las piezas y por cualquiera de los engranes.

Subrayo que la puesta en marcha de una política de estado correspondería inicialmente al Ejecutivo federal, acompañado de los otros dos niveles de gobierno (incluidas las universidades estatales), y de los otros dos poderes. Naturalmente el liderazgo del CONAHCYT resultaría clave para su diseño, implementación y evaluación. Simplemente, mientras que el Ejecutivo federal no dé el banderazo de salida, la maquinaria, seguirá produciendo los mismos resultados. Por cierto, también, naturalmente, otras instancias públicas tendrían, en su momento, mucho que decir y hacer, entre otras INEGI, BANXICO, CONEVAL, y las universidades nacionales (UNAM, IPN, UAM, etc.).

Cierro con una atenta petición al CONAHCYT que se desdoble repetidamente. Por todo lo dicho anteriormente creo que es conveniente que se reduzca la clasificación de revistas y de posgrados PNPC, así como el número de categorías dentro del SNI.

Referencias

- Afonso, Alexandre. 2014. "How academia resembles a drug gang". *SSRN Electronic Journal*. <https://dx.doi.org/10.2139/ssrn.2407748>
- Crocker, Jennifer y M. Lynne Cooper. 2011. "Addressing scientific fraud". *Science* 334, núm. 6060 (diciembre): 1182. <https://doi.org/10.1126/science.1216775>
- Duvendack, Maren, Richard W. Palmer-Jones, y W. Robert Reed. 2015. "Replications in economics: A progress report". *Econ Journal Watch* 12, núm. 2 (mayo): 164-191.
- Granger, Clive W. J. 2010. "Some thoughts on the development of cointegration". *Journal of Econometrics* 158, núm. 1 (septiembre): 3-6. <https://doi.org/10.1016/j.jeconom.2010.03.002>
- Guerrero, Carlos. 2022. "Desenmascarando a la autonomía: una estimación del coeficiente de Gini puma para el año 2020". *SciELO Preprints* núm. 4376. <https://doi.org/10.1590/SciELOPreprints.4376>
- Berk, Jonathan B., Campbell R. Harvey y David Hirshleifer. 2017. "How to write an effective referee report and improve the scientific review process". *Journal of Economic*

Perspectives 31, núm. 1 (invierno): 231-244. <https://doi.org/10.1257/jep.31.1.231>

Lequiller, François. y Derek Blades. 2014. *Understanding National Accounts: Second Edition*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264214637-en>

Muller, Jerry Z. 2018. *The Tyranny of Metrics*. Estados Unidos: Princeton University Press.

Reseña curricular

Carlos Guerrero de Lizardi. Doctor en Economía por la Universidad Autónoma de Madrid. Profesor en la UNAM (cgdl@unam.mx). Miembro del SNI. Especialista en medición. Fue director de *Investigación Económica* del número 292 al 303, periodo en el cual la revista quintuplicó su factor de impacto en el JCR, y en el SJR se consolidó como Q3. Una mejora editorial fue la puesta a disposición de las bases de datos utilizadas y rutinas estadísticas aplicadas en los artículos publicados. *Investigación Económica* ingresó a la "Replication Network", siendo la única revista publicada en español adscrita a la misma. Actualmente es editor adjunto de la *Revista Mexicana de Economía y Finanzas*.