

Consideraciones en torno a la evaluación científica

A propósito de la celebración de la 8ª Conferencia Internacional de la Federación de Editores científicos (Barcelona, 9-13 de julio de 1995)

Salvador REGUANT

Departament de Geologia Dinàmica, Geofísica i Paleontologia.

Facultat de Geologia. Universitat de Barcelona

INTRODUCCIÓN

Con el título de “Science, Culture & Communication for the 21st Century” se celebró en Barcelona del 9 al 13 de julio la 8a Conferencia de la IFSE (International Federation of Science Editors), organizada por un Comité presidido por el Profesor Ricard Guerrero Catedrático de Microbiología de la Universidad de Barcelona.

Entre los muy diversos temas que interesan y son propios de esta Federación se discutieron algunos aspectos fundamentales de los actuales sistemas de evaluación científica. Así, por ejemplo, Eugene Garfield del “Institute for Scientific Information” de Filadelfia (USA) leyó una ponencia sobre “The impact of cumulative impact factors” y hubo una sesión sobre el sistema de revisores (Peer Review System), así como varias comunicaciones sobre el rol de las revistas nacionales. Todo el material presentado será publicado en los “Proceedings” correspondientes.

No pretendo resumir las cosas que se dijeron, sino que a propósito de ellas y, en algún caso, haciendo referencia a ciertos planteamientos que allí salieron a la luz, tendría interés en hacer algunas reflexiones sobre la actual situación y funcionamiento de los sistemas de evaluación científica.

En realidad, los sistemas de evaluación actuales en uso en España y, de una manera u otra, en los demás países desarrollados giran en torno a dos fórmulas, frecuentemente interdependientes: el “peer review” o análisis de los resultados para publicar o publicados, y la valoración de los resultados indirectamente a través o de la cualificación de la revista científica en la cual aparecen o del caso, “cuantitativamente” hablando, que de estos resultados hace la comunidad científica. Esta doble valoración indirecta, teóricamente distinta, con frecuencia se plantea de modo que resulta redundante en la práctica.

EL ANÁLISIS DE LOS RESULTADOS CONSEGUIDOS POR UN CIENTÍFICO, O UN EQUIPO CIENTÍFICO, COMO MÉTODO DE EVALUACIÓN

Este método es tan antiguo como la propia ciencia y el “peer review” actual que se produce antes de la publicación de los resultados no es más que una de las fórmulas que se usan en el ámbito científico actual. Las oposiciones y concursos que de un modo u otro preceden a los contratos de acceso a las plazas de profesor de universidad o de investigador son otra fórmula de la misma idea. Podríamos añadir que este proceso se sigue en casi todos los ámbitos laborales en los que se requiere una

cierta excelencia. Así, por ejemplo, los clubs de fútbol fichan aquellos jugadores que han demostrado mejores resultados en sus actuaciones precedentes.

La cierta novedad que se ha producido en los últimos años ha sido la organización sistemática de filtros de juicio antes de publicar los resultados científicos conseguidos en las revistas de casi todo el mundo, es decir, los "peer review". De una manera menos formal esto se venía haciendo de modo que el editor o director de una revista científica aceptaba o no un trabajo para su publicación en función de su criterio o del de algunas personas a quienes consultaba. Este hecho antes, y ahora, ha causado muchos estropicios, pero como se dijo en la Conferencia de Barcelona, parece ser como la democracia que es un mal sistema, pero es el menos malo de los conocidos.

Para mí el hecho de qué se acepte me parece bien. Lo que no me parece bien es que se no se haga constantemente un análisis crítico de los errores que este sistema propicia, para hacer las modificaciones pertinentes, conjunturales o no, que conduzcan a un mejoramiento de actuación y resultados. En realidad, la ciencia es un "producto comercial" que requiere difusión. Sin difusión no hay ciencia. Por consiguiente, el vendedor de este producto, es decir los editores científicos, tienen que valorar su cualidad para que su "tienda" tenga el prestigio requerido para que sea aceptada por los "consumidores" que son los científicos. Es evidente, que la situación en el espacio de la "tienda" es importante. No es lo mismo en el centro de una gran ciudad que en la periferia o en una aldea. En este sentido, se comprende que, sin tener en cuenta su verdadera cualidad, los productos científicos son más apreciados "a priori" si se producen en el seno de los países con una gran densidad de instituciones y personalidades científicas y en la lengua de estos países que representan la punta de lanza de la ciencia contemporánea. Por consiguiente, aquí y ahora, los productos de los países más ricos y los publicados en lengua inglesa son los que juegan con ventaja a la hora de ser aceptados como buenos por definición. Este es un tema muy complejo al que ahora no querría prestar una particular atención, aunque en algún sentido ya le he dedicado algún trabajo (Reguant and Casadellà, 1994).

Volvamos a los fallos del sistema que, aunque no sean muchos, son a veces particularmente clamorosos. En la historia de la introducción y aceptación de la teoría de la tectónica de placas existe una anécdota paradigmática. La tectónica de placas fué concebida una vez se hubo obtenido una interpretación adecuada de las bandas de

anomalías magnéticas que se habían encontrado en el fondo de algunos océanos y que eran paralelas a las dorsales oceánicas. Esta interpretación era que el fondo oceánico se expandía a partir de las dorsales. Lawrence Morley del Servicio Geológico del Canadá y Fred Vine de la Universidad de Cambridge llegaron independientemente a comprender este mecanismo, pero la carta de Morley a "Nature" fué rechazada en junio de 1963 y tampoco tuvo mejor fortuna cuando la presentó al "Journal of Geophysical Research" Por suerte para la ciencia, Vine tuvo mejor fortuna, cuando junto con Matthews consiguió ver publicado su trabajo en "Nature" en el mismo año 1963 (Hallam, 1985). Otros casos se han dado y siguen dándose, algunos de los cuales llegan al hombre de la calle a través de la prensa.

En conclusión, yo pienso que es absolutamente necesario e inevitable que, previo a su publicación o después de la misma, los resultados científicos obtenidos por un investigador o un grupo de investigadores deben ser valorados por otros científicos y, esto tanto, para tener garantías de qué lo que se va a publicar es científicamente sólido, como para una buena administración del dinero, público o privado, a la hora de contratar profesores o investigadores tanto en los centros de investigación estrictamente considerados, ya en la Universidad, ya fuera de ella, como en los ligados a empresas. Esta afirmación debe completarse con un seguimiento crítico constante que corrija los posibles defectos de cualquier forma de revisión y valoración, debidos tanto a la ignorancia, como a la falta de imaginación frente a lo nuevo o, desgraciadamente también, al papanatismo. Este último caso se produce más frecuentemente en países como el nuestro que no están en primera línea en la investigación científica y tienden fácilmente al servilismo (Reguant, 1994).

LA EVALUACIÓN CIENTÍFICA A TRAVÉS DE LA EVALUACIÓN DE LAS REVISTAS O PUBLICACIONES. LOS FACTORES DE IMPACTO

Modernamente se usan criterios, cuantitativamente tratados, que han dado origen a listas de revistas, por los cuales se juzgan los "productos" en consideración de la "caja" que los contiene. Se juzga el trabajo en consideración a la revista en qué se ha publicado, normalmente, añadiendo a esta selección de revistas el número de citas que ha merecido.

Este sistema es particularmente cómodo porque no obliga a entrar en el examen del "producto" científico, sino

que se supone, no sin razón, que las revistas más prestigiosas son muy cuidadosas y serias en la admisión de trabajos para su publicación, como las buenas y acreditadas marcas comerciales procuran sólo vender productos de calidad. Por otra parte, el uso de los números de citas tiene un fundamento correcto y racional, ya que los resultados científicos más válidos acostumbran a ser los más citados.

Ciertamente una proposición que afirma que los trabajos publicados en las “buenas” revistas son de fiar y que el número de citas de un trabajo es un buen criterio de resultados aceptados, es una proposición, en líneas generales, verdadera. Lo que es absolutamente falso es la proposición contraria que dice que los trabajos publicados en las revistas “menos buenas” no son de fiar y que un trabajo poco citado no tiene gran valor científico. En cualquier caso, estos criterios son una aproximación muy grosera y expuestos a muchos más errores y desviaciones que el análisis de los trabajos en sí mismos y que el “peer review”.

Estos errores, o desviaciones, tienen un origen múltiple. En primer lugar el qué está ligado a las facilidades que tienen de publicar en revistas “buenas” los investigadores próximos a las mismas, es decir, los investigadores de los países más potentes en investigación o los investigadores que no tienen problemas en el uso de la lengua inglesa. En el primer caso es una cuestión inevitable que puede ser estudiada por los expertos en comportamiento humano. Es más fácil el contacto entre un americano y una revista americana de gran prestigio que no el que debe establecer un español, por ejemplo, con la misma revista a no ser que tenga algún nexo particular con alguno de sus responsables (*cfr.* Crusafont and Reguant, 1970 para un típico ejemplo de este fenómeno). Por lo que se refiere a la lengua inglesa, prácticamente todas las revistas “buenas” están escritas total o mayoritariamente en inglés. Tanto es así que, con frecuencia, no se lee nada más que inglés y, por consiguiente, no se puede citar lo que no se lee ni se conoce. Este hecho particularmente grave en el caso de los angloparlantes ha sido denunciado por muchos autores (ver por ejemplo, Kennedy *et al.*, 1993 y Reguant, 1993). Cabe preguntarse si las revistas en inglés son más “buenas” porque son realmente más “buenas” o porque son más citadas por causa de la ignorancia lingüística de muchos científicos que no pueden citar trabajos que no sean escritos en inglés.

Un segundo grupo de errores o desviaciones se producen por la consideración admitida por este sistema de evaluación según el cual todas las ciencias funcionan de

una manera análoga a la hora de publicar sus resultados. Esto es sencillamente falso. En realidad la ciencia no es una, sino múltiple, tanto a la hora de proceder en su tarea de descubrimiento de nuevos datos y resultados, como a la hora de establecer la comunicación de los mismos, tanto en el sistema de congresos, simposios o conferencias como en los instrumentos más útiles y eficaces de publicación. Por ello cualquier uniformización introduce planteamientos más o menos desviados en los juicios que, a causa de ella, se deriven. Por ejemplo, para las ciencias observacionales la obtención de datos y su interpretación cualitativa acostumbra a ser mucho más importante que no cualquier tratamiento posterior numérico, hecho que, a veces, resulta hasta inoperante o inútil. Stephen J. Gould (1991; véanse, por ejemplo, las páginas 100, 283 y 284-286) ha denunciado con pasión y contundencia la falacia de considerar la descripción sin elementos cuantitativos como algo poco científico y, en cierta manera, menospreciable en comparación con los métodos cuantitativos tan importantes en otros grupos de ciencias. En la página 100 concluye diciendo: “Me pregunto si conseguiremos librarnos alguna vez del peor legado de la teoría del CI en su interpretación unilineal y hereditaria: la idea de que la inteligencia puede ser captada por un único número y que las personas pueden disponerse a lo largo de una secuencia simple, que va del idiota a Einstein”. Tampoco, a menudo, los mejores trabajos en disciplinas de esta índole se publican en revistas, sino en amplias monografías que no son tenidas en consideración para el cálculo de la mayoría de índices de excelencia científica.

Un tercer grupo de errores o desviaciones tiene una relación estrecha con el anterior, ya que se refiere al excesivo simplismo con qué se toman los datos publicados por los editores de índices de impacto o de ordenamiento de revistas científicas en base a los mismos. En efecto, hay ciencias que trabajan sobre temas de interés universal, y dentro de éste, de mayor importancia para las personas humanas, como son los que se refieren, directa o indirectamente, a la salud o al progreso tecnológico. Es normal, por consiguiente, que los resultados científicos más citados sean los que se refieren a estos temas. Hay decenas de millares de investigadores interesados en el funcionamiento del riñón humano, en sus patologías y disfunciones, pero quizá no hay 10 investigadores que trabajen sobre el aparato excretor de los lagartos. Por consiguiente, en una visión simplista es de más alta calidad cualquier trabajo publicado sobre el riñón humano que sobre el riñón de un reptil. El mismo Eugene Garfield en Barcelona suministró un ejemplo de esta deformación al facilitar a los asistentes una hojas con los artí-

culos más citados entre los producidos por los investigadores barceloneses. Ocho de los diez artículos citados eran artículos de medicina, al lado de 2 de física. Lo mismo sucede en la lista de autores más citados donde los investigadores médicos se llevan la palma. También los centros de investigación médica representan (16 sobre 19) la casi totalidad de instituciones con el mayor número de artículos publicados con éxito de recepción. Yo personalmente tengo en gran estima la gran calidad de la investigación médica en Barcelona, pero me resisto a creer que el 80% de los mejores investigadores sean precisamente investigadores en el campo de la medicina. El mismo Eugene Garfield facilitó una lista ordenada de revistas científicas por "6 year average" de su impacto acumulativo. La lista comprende 120 revistas. La única revista de geología citada está situada en el número 90.

Finalmente, aunque no sea el objetivo principal de estas líneas, hay que decir claramente que en un sistema de evaluación regido por índices de impacto, citas y otros parámetros similares se hace difícil evaluar toda la producción que podríamos englobar en lo que denominamos normalmente "transferencia de tecnología". Una patente, por ejemplo, no es una publicación tradicional en una revista considerada "buena" o no tan buena. No obstante, para producirla se requiere mucho tiempo de trabajo no publicable, se cita tan sólo en circuitos muy especializados y obliga a introducirse en dominios alejados de la especialidad científica. Sin embargo, la repercusión de este tipo de trabajo es particularmente notable y eficaz y de un gran interés para el país en qué se produce y supone, frecuentemente, un alto nivel por parte de los investigadores que son capaces de crear nuevas patentes.

A MODO DE CONCLUSIÓN

Es evidente que el trabajo de investigación sólo puede ser evaluado a través de sus resultados que se concretan en alguna forma de publicación, normalmente en revistas o monografías. Un juicio acerca del valor y seriedad de los diversos modos de evaluar nos obliga a considerar tanto los aspectos positivos como los negativos de los sistemas actualmente en uso. En líneas generales, se producen errores por acción u omisión en cualquier sistema, dada la inevitable imperfección de cualquier actuación humana. Parece, no obstante, más serio y menos expuesto a errores el juicio derivado del examen de la aportación en sí, que no la derivada de la supuesta excelencia del vehículo en el cual la aportación es contenida.

Por desgracia, en España, hemos llegado a una consideración exactamente opuesta. Se da más valor al continente que al contenido. Es una forma más de papanatismo.

AGRADECIMIENTOS

Agradezco a los Profesores Mariano Marzo, Miquel Angel Cuevas y Ricard Guerrero de la Universidad de Barcelona la amabilidad de leer el texto y de sugerirme mejoras importantes.

REFERENCIAS Y NOTAS

- CRUSAFONT-PAIRÓ, M. and REGUANT, S. 1970. The nomenclature of intermediate forms. *Systematic Zoology*, 19 (3): 254-257. En la p. 257 se dice "This paper was translated from the Spanish by G. G. Simpson" En realidad fué la relación de amistad entre el Dr. Crusafont y el ilustre paleontólogo y teórico de la evolución Dr. G. G. Simpson, y no otra circunstancia, la que hizo que este trabajo fuera aceptado, traducido por el propio Simpson y publicado en una de las revistas que ha sido considerada la "más buena" en su especialidad durante muchos años. Simpson era, por su gran prestigio, miembro honorario de la sociedad editora de la revista.
- GOULD, S. J. 1991. La vida maravillosa. Editorial Crítica, Barcelona: 357 pp.
- HALLAM, A. 1985. Grandes controversias geológicas. Editorial Labor, Barcelona: 180 pp.
- KENNEDY, M., READER, S. L. and SWIERCZYNSKI, 1994. Preservation records of micro-organisms: evidence of the tenacity of life. *Microbiology*, 140: 2513-2529.
- REGUANT, S. 1993. What lingua franca?. *Nature*, 361: 107.
- REGUANT, S. 1994. Diez avisos par el científico español. *Microbiología SEM*, 10: 435-438.
- REGUANT, S. AND CASADELLÀ, J. 1994. English as lingua franca in geological scientific publications. A bibliometric analysis. *Scientometrics*, 29 (3): 335-351.