

Perversiones, paradojas y traumas en la evaluación científica

Resumen: Con ocasión de la aparición de un nuevo parámetro de estimación de la repercusión de la actividad científica, el índice h , que probablemente será tenido en cuenta de ahora en adelante en los procesos de evaluación, el autor ofrece una reflexión personal sobre el uso inadecuado de este tipo de parámetros en los procesos de evaluación científica y hace notar cómo esta perversión acaba produciendo sobre la actividad científica el efecto contrario al inicialmente perseguido.

Introducción

Cada una de las dos mitades de la humanidad viene viviendo desde hace siglos bajo el peso del parámetro que tasa y aquilata de forma implacablemente objetiva el valor de su hombría (el tamaño del pene) o su feminidad (la medida del busto). No podía ser que en este siglo de baremos y algoritmos, de dinacuatros, isos y aenores, siguiera sin someterse al metro la valía del científico; que, si en otro tiempo fue personaje bohemio, a veces cortesano y las más montaraz, es ya hoy domesticado funcionario o aspira a serlo en cuanto pueda. El último intento de medida de tan inaprensible cualidad es el índice h .^[1,2]

Vaya por delante que el susodicho me despierta, de entrada, cierta simpatía, ya que se encuentra (junto con la tensión arterial) entre las pocas cosas que no bajan con la edad. Todo tiende a aflojarse, a arrugarse, a desplomarse con el paso de los años, pero sé que siempre podré mostrar en mi escudo heráldico, al lado de mi apellido, un índice h rampante sobre salamandra alquímica (Figura 1).



Figura 1. Mi escudo heráldico.

No hay nada, pues, que me mueva personalmente a no acogerlo con júbilo entre nosotros. Nada, si no fuera porque temo que, igual que la viruela española diezmo la población nativa en los tiempos de la conquista americana, este virus h , diseñado en una sociedad sajona y protestante, produzca entre nosotros cepas mutantes que diezmen la población científica. Las precedentes mutaciones utilitarias de los datos ISI^[3] y de los sexenios de investigación como parámetros de evaluación



Pablo Espinet Rubio

no auguran nada bueno. Puede que entre tanto, Gobierno mediante, se multiplique en aparente paradoja la población de funcionarios científicos, pero es que estos dos conceptos no son la misma cosa.

Precedentes de obligada lectura

"Porque los dioses perciben el futuro, los hombres el presente, y los sabios lo que se avecina". Haciendo buena esta frase de

Filóstrato sobre la que Kavafis construyó un poema,^[4] Antonio Luque y Pascual Román (éste último buen y querido amigo desde los tiempos en los que el único índice que nos interesaba era la lista de libros de lectura prohibida por la Iglesia Católica) se han apresurado a alertar a las gentes que nada oímos sobre *"el clamor secreto de sucesos que se acercan"*.^[4] En su reciente artículo en esta revista, Antonio y Pascual nos han puesto al corriente de las características del nuevo índice, incluyendo algunos comentarios sobre su cálculo, sus virtudes y sus defectos, así como un ejemplo de aplicación a los químicos españoles.^[1] Con la elegancia y bonhomía que le caracterizan, mi buen y admirado amigo José Elguero, universalmente respetado, nos ha regalado a su vez una generosa disección de algunos aspectos de su propia historia como autor de publicaciones científicas, de la que destila un prudente escepticismo frente al significado de los números.^[5] Quisiera yo, en uso de mi ocasional desinhibición expresiva, añadir algunos comentarios y opiniones, por lo que puedan valer mi experiencia como evaluador (más larga e intensa de lo que yo quisiera) y como científico en el sistema, y la observación de mis propias miserias.

Objetivo de la evaluación científica

En una sociedad de libre concurso la Administración del Estado debe procurar justicia en la selección entre candidatos. El objetivo de la evaluación (en nuestro caso científica y académica) es seleccionar a los mejores. No ya para ser justo con ellos, sino también para garantizar el mejor servicio al propio Estado. Pero ¿cómo garantizar esa justicia cuando los méritos no son de determinación tan directa como lo es un registro de tiempo en una carrera o el peso que se levanta en una alzada? La respuesta que se da a ese problema es la utilización en exclusiva de parámetros objetivos. Obsérvese que de entrada se renuncia a aspirar a la *objetividad* como cualidad asociada a las personas considerándola, por lo visto, una virtud humana inalcanzable. Lo que significa que cuando se habla de parámetros *objetivos* se quiere decir, en realidad, parámetros *numéricos*. Universidades, tribunales, ANECA, agencias de evaluación autonómicas y otras instancias seme-

Química Inorgánica, Facultad de Ciencias,
Universidad de Valladolid,
Pº Prado de la Magdalena s/n, 47071 Valladolid
C-e: espinet@qi.uva.es

jantes se han aplicado, en ocasiones con una creatividad digna de mejor cabeza, a la elaboración de fórmulas y plantillas, a veces detalladísimas, para obtener la cifra precisa de nuestro justiprecio.

El proyecto de modificación de la LOU ya recoge la propuesta del Gobierno de sustituir las habilitaciones presenciales por evaluaciones de méritos, como se viene haciendo con las acreditaciones de la ANECA y otras agencias autonómicas, en un proceso en el que las personas encargadas de la evaluación no podrán ser sino meros ábacos encarnados.

El evaluado como personaje imaginario.

Si nadie lo remedia, en este sistema de sordos y mudos (no me resisto a bautizarlo como anecoico), se parte de la base de que hay que juzgar a los concursantes como si fueran entes virtuales o personajes fallecidos, con quienes no cabe el diálogo, la entrevista, ni el intercambio directo de expresión inteligente. El índice h y todos sus compadres (número de publicaciones, docencia impartida, índice de impacto de las revistas, participación en proyectos, número de citas por artículo, "género" y otros varios asuntos "objetivables"), habitualmente mixtificados mediante agrupamientos de revistas en unos pocos niveles, coeficientes correctores y el resto de artimañas niveladoras al uso, pasan a sustituir a nuestros ojos, nuestro cerebro y nuestro criterio profesional. Y el concursante, desprovisto de su alma, su carne, su personalidad, su inteligencia oratoria y dialéctica, sus motivaciones y sus aspiraciones, será como un insecto del jurásico atrapado en la fosilizada gota de ámbar de su currículum (Figura 2). Y no nos queda otra que adquirir el ámbar sin saber muy de cierto si el insecto puede realmente volar o tiene al menos intención de hacerlo.



Figura 2. El concursante.

Objetivo ante la evaluación científica

Olvidamos demasiado a menudo que, aunque los designios de Dios puedan ser inescrutables, las leyes de Newton son inexorables. Deberíamos tenerlas en cuenta para escrutar el futuro derivado de cualquier perturbación del sistema. La tercera ley (o Principio de Acción y Reacción) se aplica en este caso de la siguiente forma: Una persona situada *ante* su evaluación científica orientará su actividad a obtener un resultado positivo de esa evaluación. En otras palabras, un escenario de evaluación condiciona la respuesta y actuación futura de la sociedad evaluada, tanto más cuando ésta conoce perfectamente la selección darwiniana, en la que sobrevive el más apto, entendido como tal el mejor adaptado al escenario.

Nada que objetar siempre que, cerrado el círculo, se haya mantenido por el camino el objetivo inicial; es decir, si adap-

tado el ámbito social de aplicación al procedimiento de evaluación son de verdad los mejores científicos los seleccionados y, en consecuencia, los supervivientes (Figura 3). Mi opinión es que los sistemas de evaluación utilizados, aunque han estimulado la actividad, no llevan al fin inicialmente perseguido.



Figura 3. Somos como niños.

La perversión de los índices.

Supongamos de entrada que los índices elegidos midieran realmente la propiedad que intentamos cuantificar: la calidad científica. Parece claro que este supuesto sólo sería aplicable al número de citas y al índice h , puesto que no necesita discutirse que el índice de impacto de una revista da únicamente una media del éxito de todos sus artículos, que no es aplicable a cada uno de ellos en concreto. Pues bien, aun en el caso supuesto, que representa el óptimo de aproximación a una evaluación numérica, el sistema de medida tiene muchas debilidades:

1) *Las condiciones de aplicación de los índices no se cumplen.* Para cualquiera de estos índices (n° de citas y h) hay que excluir las autocitas, pero esto no se hace. Esta es una primera gigantesca perversión, en la que caen los propios Antonio y Pascual en su artículo (aunque advirtiéndolo explícitamente), pero también cualquier comisión de evaluación que haya llegado a ese improbable nivel de precisión en su trabajo (en este caso sin advertirlo a nadie). Dado que cualquier escrutinio que tenga que aplicar rigurosamente la exclusión de autocitas requiere una cantidad de trabajo considerablemente mayor, se acepta la perversión del índice teniéndolas en cuenta. Así que, en la práctica, se aplica, como mucho, el índice "pervertido".

Incluso en casos en que la autocita no sea una estrategia, es inevitable una incidencia elevada de autocitas cuando se producen muchos artículos y más intensa cuando en el evaluado se de una persistencia prolongada en el tema de trabajo. En otras palabras, la incidencia positiva de la autocita en los indicadores "pervertidos" afecta desigualmente a la población evaluada. Ni los muchos artículos ni la perseverancia temática son malos en sí mismos, pero tampoco son garantía de calidad científica.

Por otro lado, el índice h no está relacionado directamente con el número total de citas. Si se limita la búsqueda de índices h a un ámbito de investigadores con alto número total de artículos o de citas (mucho mayor que el propio número h resultante), se perderán inevitablemente por debajo del corte investigadores con índices h superiores a algunos de los que estén por encima del corte, pero que los han conseguido con un número menor de artículos o de citas totales. Así ocurre en el caso estudio de Antonio y Pascual. Aunque este problema no opera cuando ha de realizarse una clasificación entre una lista concreta de individuos, no debe olvidarse que persiste la incidencia de las autocitas en el índice h "pervertido".

2) *El trabajo de equipo no se puede atribuir a personas concretas.* La actividad investigadora en muchas ciencias (incluida la nuestra) suele ser en equipo. Ocasionalmente se encuentran grupos en cuyas publicaciones aparecen listas de autores relativamente largas, que se repiten y mantienen en el tiempo. Este fenómeno puede ser el resultado de una estrategia del grupo para beneficio de sus componentes, o bien de una auténtica forma de organización del trabajo, en la que cada componente aporta una habilidad o actividad específica. El primer caso es, desde luego, fraudulento en origen, mientras que el segundo puede ser simplemente una legítima optimización del sistema de producción científica. En ambas circunstancias el producto científico puede ser excelente e inobjetable. Sin embargo, es claro que ese resultado global no facilita conocer el valor individual de cada firmante. Tampoco es aceptable una corrección obtusa como dividir el valor total por el número de autores. En definitiva, poco nos dicen esos números mudos acerca de la persona disecada como una mariposa y pinchada con un alfiler en el papel de su CV.

3) *A veces el bisturí está mellado.* Y así sale el corte. Los datos del ISI no corrigen errores materiales en las citas originales, lo que puede tener su importancia. Lo explicaré con un ejemplo sencillo. Imaginemos que uno de nuestros artículos ha recibido 10 citas en total. Sin embargo, debido a un inicial error material alguien equivocó la página de la referencia y él y otros hicieron *copy-paste* de la referencia equivocada hasta alcanzar 5 citaciones con el error. La consecuencia es que en el ISI tendremos dos artículos, cada uno con cinco citas. Obviamente esto aumenta nuestro número de artículos en una unidad, y no altera nuestro número total de citas, pero disminuye nuestro promedio de citas por artículo, y merma la posibilidad de escalar algún puntito en el índice h . Si nuestro h fuese 9, este artículo, con 10 citas reales, no lo haría subir a 10. Puede parecer una tontería, pero invito a cualquiera que tenga un CV un poco largo a que lo examine prestando atención a la repercusión de los errores en las citas, o de la publicación de inoportunas *corrigendas*. Os llevaréis alguna sorpresa.

Como idea malévola, la utilización deliberada de estos defectos puede permitir socavar el trabajo de un enemigo mucho más allá que simplemente ignorarle. Citarle con errores es un torpedo a la línea de flotación de su índice h . Incluir sus trabajos en una revisión, o citarle en la introducción de una publicación para, en las siguientes ocasiones, enterrarlo en la cita de esa publicación nuestra *and references therein* es, sin embargo, un arma de destrucción más masiva que afecta a todos los parámetros. Se viene utilizando.

4) *Cuando los índices se abandonan.* Pero la práctica es aún peor. En los ámbitos habituales de evaluación incluso estos índices "pervertidos" individualizados resultan demasiado

trabajosos. Así, se acaban cursando otros protocolos según los cuales el evaluador se limita a contabilizar el número de publicaciones en revistas "según su índice de impacto", pero previamente se impone que estas se clasifican en tres o en cuatro grupos. Todas las de cada uno de los grupos pasan a cobrar el mismo valor de impacto. ¿Puede imaginarse que sólo se vendieran cuatro tallas de zapatos? Pues eso es lo que se hace frecuentemente al evaluar.

Suficiente. Sólo pretendo argumentar que los parámetros numéricos son datos útiles, pero en sí mismos no son una medida de nada y mucho menos la medida de una persona, de su capacidad, de su madurez, ni de su valor y potencial científico o académico. Hay otros datos que se utilizan, aparte de los índices citados: En cuántos proyectos ha participado, cuántos diplomas ha coleccionado, cuántas clases ha impartido. Tenemos un sistema de evaluación en el que aburrir y desesperar a los alumnos diez años puede tener una valoración el doble de positiva que haberlo hecho solamente durante cinco. Y recibir 200 citas denunciando que nuestro resultado no es reproducible (caso no infrecuente en áreas de lo *bio*) nos puede situar en el primer lugar de la evaluación.

Parece que, en breve, otro de los aspectos con incidencia en la valoración (sobre todo de instituciones) estará relacionado con el llamado "género". Desconozco todavía si, llegado el momento, habrá que incluir como coeficiente de "género" el tipo de índices a los que yo aludía en el primer párrafo de este desahogo.

La perversión de los conceptos.

Pero, ¿es tan grave la perversión de los índices? Desde luego. La aceptación y generalización de esos índices como medida de la ciencia y del científico ha producido ya resultados aberrantes en el desarrollo y proyección de la ciencia española y, en buena parte, de la europea. Veamos algunos.

1) *La desaparición de la Ciencia como objetivo.* Parece que la búsqueda y el riesgo escasean en nuestra investigación. La aplicación de lo que sabemos predomina sobre la búsqueda y penetración en lo que desconocemos. Pero la Ciencia no se puede jugar a la defensiva. Investigar, en Ciencia, no puede ser sino intentar hacer lo que no sabemos hacer. En estos momentos es relativamente fácil publicar en una revista de nivel razonable o alto sólo con resultados previsibles y prescindibles pero bien realizados. La mayor parte de los artículos son así. La competencia de las revistas en su oferta de rapidez, índice de impacto, color, etc. abona la idea de que el fin de la investigación es la publicación en sí misma, no tanto el contenido ni el proceso (el objetivo más importante de la obra alquímica era la transformación del alquimista). Esta estrategia editorial les ha permitido convertirse en el árbitro real de la situación, a veces también en el Petronio de la moda científica. Pero la Ciencia no es siempre prisa, ni mercado, ni policromía. Aceptamos gustosos ese juego garantista¹ frente al inevitable tormento de la inseguridad y la duda en uno mismo que la verdadera búsqueda produce. Ya no buscamos ciencia, buscamos publicaciones. Y, claro está, predominantemente ya no encontramos ciencia, encontramos publicaciones.

¹ Mi corrector de español me advierte de que la Real Academia de la Lengua Española recoge el vocablo *garantizador* y no *garantista*. En este caso la RAE no parece haber atendido todavía el lenguaje del pueblo.

2) *La desaparición del científico como sujeto*. Acomodados a esa situación entregamos nuestra propia conciencia científica al director espiritual que son los índices. Nuestro criterio son los índices, nuestra autoafirmación son los índices, nuestra absolución son los índices. No sabríamos tal vez justificar por qué hacemos lo que hacemos, por qué merece nuestro esfuerzo y el de otros, pero podemos responder que está publicado en el JACS. Y podemos responderlo ante nosotros mismos, si alguna vez nos permitimos una duda de conciencia. Nos entregamos al juego del intento de engaño ajeno y propio, procurando vender nuestros resultados por más de lo que nosotros mismos creemos que valen: "Se la han quedado en...", "La he colado en..." son expresiones que a nadie resultan extrañas al referirse a una publicación.

3) *La desaparición del maestro*. Atrapados en esa trampa que hemos creado e impuesto nos movemos como responsables de quienes se incorporan a nuestro grupo de trabajo, buscando arañar puntos que les den una oportunidad de supervivencia. Y así les alimentamos con el mismo veneno que nos enferma y les educamos en la misma doctrina que nos aliena. En esta vorágine de coleccionismo de números dejamos atrás nuestra obligación de maestros y descuidamos la de profesores. Es posible que seamos todavía mejores docentes que quienes no investigan y mejores investigadores que quienes no enseñan, pero no me cabe duda de que somos mucho peores en ambos aspectos de lo que pudiéramos y debiéramos ser si no estuviéramos corriendo locamente como el Conejo Blanco de Alicia (Figura 4): "*Oh dear, oh dear, I shall be too late*".^[6] Y sacamos del chaleco el reloj que nos marca el índice *h* y apretamos el paso.



Figura 4. El Conejo Blanco.

A modo de conclusiones

Quiero concluir con unas pocas reflexiones, personales y en consecuencia discutibles.

Es ocioso decir a estas alturas que mi opinión sobre la aplicación, tal como se hace, de índices en la evaluación de personas e instituciones es negativa, y mi diagnóstico sobre su influencia en la evolución de la ciencia es pesimista. Estamos aquejados de un grave ataque de *hacheindicitis* y los síntomas febriles son muy evidentes. Pero no queremos reconocerlos porque los índices nos dan asideros para sentirnos seguros frente a nuestra propia incertidumbre, como las religiones lo

hacen frente a la incertidumbre de nuestra vida y de nuestra muerte. Tal vez por eso los índices, como las religiones, quedan a salvo del descrédito mediante análisis intelectual. Lo que me hace presumir que no habré convencido a casi nadie.

Mi experiencia, como investigador y como observador, me indica que se madura y se aprende más en la lucha contra el fracaso que en la explotación del éxito. Por eso puedo en ocasiones conceder más aprecio a la realización de diez publicaciones diferentes que a la de cien veces la misma publicación con distintos collares. No es esto lo que concluyen los índices numéricos, y por eso temo que su mal uso pueda acabar desterrando de la arena científica a los verdaderos buscadores de conocimiento.

La eliminación de pruebas con concurrencia simultánea, presentación personal y discusión, no deja otra herramienta a los evaluadores que la de esos índices. La consecuencia lógica debería ser eliminar a los llamados "profesores de reconocido prestigio" de las comisiones de evaluación, para que no sufran, ya que no hay expectativas de que puedan obrar aplicando el *criterio* que su reconocido prestigio les supone. ¿O es que se necesitan sólo como coartada del procedimiento?

En fin, como ocurrió con motivo del hallazgo de oro en California en 1849, la fiebre del índice ha inflamado a la población investigadora, y provistos de picos y palas, a veces incluso a lomos de un asno, nos hemos lanzado a correr y cavar sin pausa ni seso. No habremos de extrañarnos si lo único que alcanzamos a recolectar es el *oro de los tontos* (Figura 5).²



Figura 5. El oro de los tontos

Referencias

- [1] A. Luque, P. Román, *An. Quim.* **2006**, 102(2), 11–17.
- [2] J. E. Hirsch, *Proc. Nat. Acad. Sci. USA*, **2005**, 102, 16569–16572.
- [3] <http://portal.isiknowledge.com>.
- [4] C. P. Kavafis *Poemas*, 2ª edición, p. 64. Seix Barral, Barcelona, **1994**.
- [5] J. Elguero, *An. Quim.* **2006**, 102(2), 18–24.
- [6] Lewis Carrol, *Alice's Adventures in Wonderland and Through the Looking-Glass*, The Centenary Edition, p. 9–10. Penguin Books, London, **1998**.

² Dada la práctica desaparición de las pantallas de las películas del Oeste, tan abundantes en mi infancia, tal vez no esté de más aclarar que el *oro de los tontos* es la pirita, FeS₂. Me consta que hay piritas auríferas, pero son las menos.