

Evaluación de la calidad de los artículos y de las revistas científicas: Propuesta del factor de impacto ponderado y de un índice de calidad

Gualberto Buela-Casal
Universidad de Granada

El factor de impacto y otros índices bibliométricos son utilizados en la actualidad en diversos países para evaluar la producción y/o la calidad de la investigación científica. Sin embargo, rara vez se interpreta correctamente el factor de impacto (o el factor de prestigio). Éstos presentan varias limitaciones importantes que deben ser consideradas, una es que se refieren sólo a las citas de los dos o tres años anteriores (respectivamente) y la otra es que no se tiene en cuenta el «impacto o el prestigio» de las revistas donde se producen las citas, dando el mismo valor a cualquier cita, independientemente de la revista en la que se produce. Para paliar esta limitación se proponen dos índices: el *factor de impacto medio de las revistas donde se producen las citas* (FIMRC) y el *factor de impacto ponderado* (FIP). Además, se proponen también otros índices que resultan útiles para analizar la interacción entre las revistas: el *porcentaje de interacción parcial de citas* (PIPC) y el *porcentaje de interacción mutua de citas* (PIMC). En este artículo se explican los fundamentos y la forma de calcularlos. A continuación se comenta el sistema actual utilizado para evaluar artículos, el sistema de revisión «por iguales»: los sesgos del director, la cualificación de los revisores y los criterios establecidos para evaluar. Además, se establece una serie de argumentos por los que no debe evaluarse la calidad de un artículo sólo en función de la revista en la que fue publicado. Por último, se presenta una alternativa para evaluar la calidad de los artículos y de las revistas científicas, centrándose en: qué se debe evaluar, quién debe evaluar y qué criterios se pueden usar. En función de esto se hace una propuesta de cómo se podría crear un índice de calidad, independientemente del *factor de impacto* o del *factor de prestigio*.

Evaluating quality of articles and scientific journals. Proposal of weighted impact factor and a quality index? The factor of impact and other bibliometric indices are currently used in several countries to evaluate the type and quality of scientific production. However, the impact factor (or prestige) rarely receives an accurate interpretation. Available impact factors display a number of shortcomings: they only refer to citations in the previous two or three years, and they do not take into account the impact or prestige of the periodicals where citations appear, so that every citation is given the same value, regardless of the periodical where it appears. In order to overcome these limitations, two indices are proposed: a *mean impact factor of the journals where citations appear* (FIMRC) and a *weighted impact factor* (FIP). Additionally, other useful indices are suggested for the analysis of interaction between periodicals: a *percentage of partial interaction of citations* (PIPC), and a *percentage of mutual interaction among citations* (PIMC). This paper explains their details a procedures for their calculation. Several problem areas are discussed, namely, peer review, the policy of publications, qualification of referees, and assessment criteria. It is also argued that quality of studies should not only be primarily evaluated in terms of the periodical where they are published. Finally, an alternative is offered for the assessment of quality of scientific articles and journals on three bases: *what* is to be evaluated, *who* is to be an evaluator, and possible *criteria* for evaluation. These considerations lead to a proposal for a quality index aside of impact or prestige.

El *factor de impacto* de las revistas es el índice bibliométrico más utilizado para evaluar y comparar la producción de los países, véase por ejemplo, la comparación de España con respecto a la

producción científica mundial (Aguirre de Cárcer, 2001), donde la producción científica española representa el 2,7% de la producción mundial correspondiente al período 1996-2000; en otros estudios se pone de manifiesto que España ocupa el sexto lugar con respecto a los países de la Unión Europea y el undécimo en relación a la producción mundial (Sánchez, 1999). En el ámbito de la psicología en España también se han realizado estudios sobre la producción científica usando como criterio el factor de impacto de las revistas donde publicaron los psicólogos españoles (véase Alcain y Carpintero, 2001). Sin embargo, el uso de este índice bi-

bliométrico presenta importantes inconvenientes, que por una parte se deben a la incorrecta interpretación que se hace de estos índices (véase Buela-Casal, 2001, 2002) y por otra a limitaciones «intrínsecas» del propio índice, como por ejemplo, que el factor de impacto se base exclusivamente en las citas de un período de dos años (Buela-Casal, 2002), y que para su cálculo no se tenga en consideración el «impacto» o «prestigio» donde se producen las citas.

En la mayoría de los países para evaluar la producción y la calidad de las investigaciones científicas se recurre a índices bibliométricos. Éstos se usan de forma indiscriminada para evaluar investigadores, centros de investigación, universidades, países, etc. Estos índices se basan en el número de citas que reciben las revistas científicas (en un período determinado y en función de los artículos que publican), el *factor de impacto* y el *factor de prestigio* son algunos de estos índices.

No habría ningún problema si los índices bibliométricos se usaran realmente como lo que son, es decir, indicadores del nivel de difusión entre la comunidad científica. El problema aparece cuando estos factores son utilizados como índices de calidad, y en función de ello se considera que un artículo tiene calidad en función del «impacto» o «prestigio» que tiene la revista en la que fue publicado. Esto está siendo cada vez más cuestionado en la actualidad. Para Pelechano (2000) esto es confundir la ciencia con la sociología de la ciencia, y critica que lo que comenzó siendo una determinada y muy específica manera de entender el «impacto» de las publicaciones científicas (sin leerlas) terminó siendo la manera de valorar las aportaciones científicas. De manera similar, Sternberg (2001) dice que hay que diferenciar entre lo que se publica y donde se publica, no todo lo que se publica en una misma revista tiene la misma calidad. La correlación entre el «impacto» de un artículo y el «impacto» de la revista en la que fue publicado está muy lejos de ser perfecta (Sternberg y Gordeeva, 1996). Buela-Casal (2001) concluye que el «impacto» de una revista no es un índice de calidad de la profesión o de la relevancia social o de la implementación que suponen las investigaciones publicadas en esa revista. Tal como plantean Werner, Pelicioni y Chiattonne (2002), ¿de qué sirve que se publiquen propuestas de programas de intervención en ámbito de la psicología de la salud si luego éstas no se implementan?

Por otra parte, hay que tener en cuenta que para evaluar la investigación en las áreas aplicadas, ésta no siempre tiene como resultado la publicación, puesto que muchos investigadores, empresas o industrias prefieren patentar los resultados de sus investigaciones que publicarlos en revistas científicas (Gómez y Bordons, 1996).

Las reflexiones críticas sobre los actuales sistemas de evaluación de la investigación, de los artículos y de las revistas científicas son más frecuentes cada día (véase Pelechano, 2002a, 2002b, 2002c) y desde múltiples perspectivas: desde el punto de vista de la experiencia de los directores de revista (Siguan, 2002), desde la experiencia de un revisor (Bobenrieth, 2002), desde la perspectiva de un investigador (Sternberg, 2002, 2003), en función de la relación entre la ciencia y la política (Polaino, 2002), en relación al sistema de producción científica (sexenios de investigación) de los investigadores españoles (Echeburúa, 2002), analizando los criterios para evaluar la calidad universitaria (Chacón, Pérez-Gil, Hologado y Lara, 2001), en relación a los fundamentos científicos (Bornas y Noguera, 2002), desde la perspectiva del contexto de la cultura española (Carpintero, 2001; Pérez Álvarez, 2002), en fun-

ción del idioma y del lugar de edición de la revista en la que se publica (Buela-Casal, 2001; Van Leeuwen, Moed, Tijssen, Visser y Van Raan (2001), reflexionando sobre las diferencias entre la cantidad y la calidad (Gil-Roales y Luciano, 2002), criticando el mal uso que se hace de criterios como «la opinión de la mayoría», del factor de impacto y del factor de prestigio para evaluar la calidad de la investigación (Buela-Casal, 2002), analizando las limitaciones de los indicadores bibliométricos (Aleixandre y Porcel, 2000; Amin y Mabe, 2000; Bordons, Fernández y Gómez, 2002; Gómez y Bordons, 1996; Seglen, 1993, 1997); evaluando el efecto que tiene sobre las revistas científicas españolas y sobre su futuro (Díaz et al, 2001; Bordons, Felipe y Gómez, 2002; Jiménez-Contreras, Faba y Moya, 2001) y hasta desde el planteamiento de si se trata de un pseudoproblema, sociologismo o ideologismo (Pelechano, 2002b).

Los distintos índices bibliométricos como factor de impacto, factor de prestigio, factor de impacto equivalente, etc. pueden ser manipulados de forma intencionada o no por medio de un incremento «artificial» del número de citas. Recientemente, Buela-Casal (2002) ha descrito diversos procedimientos que se pueden usar para aumentar las citas y que ha denominado como «Los Diez Mandamientos para incrementar las citas» (véase cuadro 1).

<i>Cuadro 1</i> Los diez mandamientos para incrementar las citas	
1.	Incrementar la difusión de la revista.
2.	Incluir la revista en el mayor número posible de bases de datos.
3.	Publicar artículos polémicos.
4.	Publicar revisiones.
5.	Publicar en idioma inglés.
6.	Publicar artículos sobre temas de actualidad.
7.	Publicar artículos de autores muy citados.
8.	Establecer acuerdos con medios de comunicación.
9.	Recomendar que se citen trabajos publicados en la misma revista.
10.	Facilitar el acceso a los artículos por internet.
Fuente: Buela-Casal (2002)	

Otro aspecto importante que no se suele cuestionar es el sistema de evaluación de los artículos científicos, lo que se conoce como la evaluación «por iguales». Aquí se analiza de forma detallada las limitaciones que presenta este procedimiento, tal como se practica en la mayoría de las revistas. Asimismo, se analizan las distintas razones por las cuales debe erradicarse la mala costumbre de evaluar la calidad de los artículos sólo a través del lugar en el que se publica.

Por último, se presenta una alternativa para evaluar los artículos y las revistas científicas, enfocando dicha evaluación en tres elementos: qué criterios se podrían utilizar, qué se debe evaluar y quién debe evaluar. Algunos de estos aspectos no son totalmente novedosos, aunque sí la forma de plantearlos y de ponerlos en práctica.

El criterio del «*factor de impacto*» y del «*factor de prestigio*»

El interés por clasificar o «medir» la investigación científica no es algo reciente, una de las primeras clasificaciones fue propuesta por Gross y Gross (1927). Sin embargo, el criterio de medir el «impacto» de las publicaciones científicas fue planteado por Garfield (1955) y publicado en la revista *Science*, no obstante, el cri-

terio «*impact factor*» fue utilizado por primera vez para cuantificar las publicaciones en la edición de 1963 del *Science Citation Index (SCI)*. Inicialmente este índice se publicaba en un suplemento del SCI, con el nombre *Journal Citation Reports (JCR)*, y en la actualidad se ha convertido en la publicación más importante del *Institute for Scientific Information (ISI)*.

El *Journal Citation Reports* se edita anualmente y se compone de dos secciones, una denominada *Sciences* y otra *Social Sciences*. En cada una de estas secciones se clasifican las revistas en dos categorías, una por orden alfabético y otra por materias (dentro de ésta se clasifica de mayor a menor factor de impacto). Aunque el índice principal y más conocido es el *factor de impacto*, también se incluyen otros: total citas realizadas en el año, índice inmediato y vida media de las citas (para una descripción de estos índices véase Buela-Casal, 2002).

El *factor de impacto* se calcula dividiendo el número total de citas que reciben en un año los artículos publicados en una revista en los dos años anteriores entre el número de artículos publicados en esa revista en esos dos años.

El *factor de prestigio*. Éste ha sido propuesto recientemente, por una empresa canadiense (*Factorprestige*), como una alternativa al *factor de impacto* del ISI. Sin duda, el *factor de prestigio* presenta importantes ventajas como por ejemplo utilizar cuatro bases de datos: ciencias biomédicas, tecnológicas, agro-geoambientales y sociales. En estas bases se recogen más de seis mil revistas que se clasifican en 859 categorías (mientras que el ISI utiliza 212 categorías). El uso de estas cuatro bases de datos permite una comparación más racional de las revistas, pues se compara cada revista con otras relacionadas en la misma área.

Otra novedad importante que presenta el *factor de prestigio* es que para calcular éste no se incluyen las citas procedentes de los artículos de revisiones. Se ha estimado que las revisiones a su vez reciben tres veces más citas que un artículo original. Además, la mayor parte de las revisiones son publicadas por invitación del director de una revista a un autor «importante», lo cual rompe en cierto modo las «reglas del juego» en la selección de los artículos.

El *factor de prestigio* se calcula dividiendo el número total de citas que reciben en un año los artículos originales publicados en una revista en el mismo año y en los dos años anteriores entre el número de artículos originales publicados en esa revista en esos tres años (cuando se trata de revistas biomédicas se tienen en cuenta otras dos variables: artículos clínicos o artículos básicos). Los resultados son convertidos mediante un algoritmo en una escala con rango de 0 a 1.000.

Factorprestige ofrece otros índices bibliométricos específicos que permiten una mejor cuantificación de las publicaciones: factor de impacto equivalente, factor de revisión, índice de revisión, percentil, índice de interacción, índice de publicación, etc. Sin embargo, de forma directa o indirecta todos estos índices se basan en el número de citas (véase Buela-Casal, 2002 para una descripción de estos índices).

Es importante tener presente que el factor de impacto y el factor de prestigio, aunque tratan de medir lo mismo, no son comparables, y ello por varios motivos: utilizan períodos de citas diferentes (dos años el primero y tres el segundo), utilizan rangos diferentes, el primero incluye revisiones y el segundo no, etc. (para más información véase Buela-Casal, 2001).

Una limitación importante que presentan el factor de impacto y el factor de prestigio es dar el mismo valor a cualquier cita, y ello indistintamente de la revista en la que se produce la cita, una po-

sible solución a esta limitación es el uso del factor de impacto ponderado, tal como se describe en el siguiente apartado.

Propuesta del factor de impacto ponderado y de otros índices bibliométricos

Tal como fue propuesto por el *Institute for Scientific Information* el factor de impacto se calcula en función del número total de citas pero no tiene en cuenta en qué tipo de revistas se producen las citas. Una limitación importante es dar el mismo valor a todas las citas, da igual que éstas se produzcan en una revista que tenga factor de impacto bajo, como por ejemplo, *Scientist*, que recibió 348 citas (en el año 2000) y tiene un factor de impacto de 0,347; que en una revista como *Science* que en el año 2000 recibió 274.443 citas y tiene un factor de impacto de 23,872 (*Institute for Scientific Information*, 2000). Es evidente que una cita en esta revista es más importante o tiene más «impacto» que una cita en *Scientist*. En resumen, al igual que no es lo mismo publicar en *Science* que en *Scientist*, tampoco es lo mismo una cita en una u otra de estas revistas.

La argumentación puede aplicarse de igual forma al factor de prestigio, que aunque es mucho menos conocido es considerablemente mejor que el factor de impacto (para una comparación de estos índices véase Buela-Casal, 2002). No obstante, presenta el mismo inconveniente de dar el mismo valor a una cita sin tener en cuenta el nivel de la revista en la que se produce.

Uno de los objetivos del presente artículo es hacer una propuesta de un índice modificado del factor de impacto y del factor de prestigio, que pondere el valor de las citas en función del impacto y/o prestigio de las revistas donde se producen las citas. Para ello, es necesario proponer el uso de otro índice bibliométrico como el *Factor de impacto medio de las revistas donde se producen las citas*, que a su vez será necesario para calcular el *Factor de impacto ponderado*. Por otra parte, se proponen otros dos índices bibliométricos que son de gran utilidad para comparar dos o más revistas entre sí en relación a la interacción de citas comunes. Un ejemplo en el que se analiza la interacción entre distintas revistas es el reciente estudio comparativo entre las cuatro revistas de Psicología con factor de impacto editadas en castellano (Buela-Casal, Carretero-Dios y De los Santos-Roig, 2002).

Propuesta de índices bibliométricos

– *Factor de impacto medio de las revistas donde se producen las citas (FIMRC)*: Se refiere al promedio del factor de impacto ponderado de las revistas donde fue citada la revista. Se calcula multiplicando el factor de impacto de cada revista (en ese año) en la que se citan artículos de los años anteriores de la revista analizada por el número de artículos citados en cada revista, la suma total se divide por el número total de artículos citados.

Este índice bibliométrico es uno de los más importantes, pues indica el factor de impacto medio de las revistas donde se citan los artículos de una revista. Cuando este factor es superior al factor de impacto de la revista indica que las citas de la revista se producen en revistas con mayor factor de impacto que la revista citada. Cuanto mayor sea el factor de impacto medio de las revistas mayor influencia o efecto tiene esa revista sobre la comunidad científica, pues los artículos son citados en revistas que a su vez son muy citadas. En la tabla 1 puede verse las revistas en las que fue

citada *Psicothema* en el año 2000, éstas tienen un factor de impacto desde 0 hasta más de 3, y es evidente que no es lo mismo una cita en una revista con factor de impacto 0,000 que con 3,858, dado que esta última a su vez es muy citada, por lo que una cita de *Psicothema* en *Exp. Neurol.* necesariamente tendrá más importancia que una cita en una revista sin factor de impacto, puesto que muchos investigadores que han leído y citado el artículo de *Exp. Neurol.* conocen al menos la existencia del trabajo publicado en *Psicothema*, por lo que aumenta la probabilidad de que sea más consultado y más citado. Por el contrario, una cita en una revista con 0 factor de impacto indica que los artículos publicados en esa revista no fueron citados por otros investigadores, lo cual no potencia futuras citas del artículo en cuestión.

Revista	Factor de Impacto
EXP NEUROL	3,858
J CHILD PSYCHOL PSYC	2,940
BRAIN RES	2,526
BEHAV BRAIN RES	2,263
MED CARE	2,535
NEUROCHEM RES	1,858
PHARMACOL BIOCHEM BE	1,732
PSYCHON B REV	1,580
BEHAV THER	1,494
PROG NEUROPSYCHOPH	1,078
PERS INDIV DIFFER	0,920
INT J CLIN EXP HYP	0,897
AGRESSIVE BEHAV	0,815
ADDICT RES	0,596
J BEHA THER EXP PSY	0,567
MATH Z	0,502
PSICOTHEMA	0,473
SALUD MENTAL	0,329
PERCEPT MOTOR SKILL	0,308
SOC BEHAV PERSONAL	0,227
ACTAS ESP PSIQUIATRI	0,098
PSICOL CONDUCT	0,000
THEOR COMPUT SCI	0,000
FACTOR DE IMPACTO MEDIO	0,592

Nota: éste no se calcula directamente, sino a partir del número de citas que recibió *Psicothema* (de los artículos publicados en los dos años anteriores) en cada una de estas revistas.

Un inconveniente que hay que tener en cuenta en este índice es que se produce en cierto modo un fenómeno de «regresión a la media», pues las revistas con un factor de impacto muy alto serán citadas necesariamente en revistas con igual y menor factor de impacto, es decir, que a partir de un determinado factor de impacto, el factor de impacto medio de las revistas donde se producen las citas será menor que su propio factor de impacto, lo cual es normal, pero no debe ser mal interpretado.

– *Factor de impacto ponderado (FIP)*: Éste se calcula a partir del factor de impacto (FI) de una revista y del factor de impacto medio de las revistas donde se producen las citas (FIMRC). Se obtiene calculando el promedio de ambos factores.

$$\text{FIP} = \frac{\text{FIMRC} + \text{FI}}{2}$$

El factor de impacto ponderado presenta la ventaja de que integra información sobre las citas que recibe una revista y del factor de impacto que tienen las revistas que la citan. Es evidente que no es lo mismo que una revista reciba un determinado número de citas procedentes de revistas de bajo factor de impacto que de alto factor de impacto. Este índice corrige el efecto de la «regresión a la media» comentado en el índice anterior (FIMRC), pues al promediar el factor de impacto medio de las revistas donde se producen las citas y el factor de impacto de la propia revista impide que se produzca ese efecto. El ejemplo en el caso de *Psicothema* para el año 2000 sería:

$$\frac{0,592 (\text{FIMRC}) + 0,473 (\text{FI})}{2} = 0,532 (\text{FIP})$$

– *Factor de prestigio medio de las revistas donde se producen las citas (FPMRC)*. Se refiere al promedio del factor de prestigio ponderado de las revistas donde fue citada la revista. Se calcula multiplicando el factor de prestigio de cada revista (en ese año) en la que se citan artículos del año en curso y de los años anteriores de la revista analizada por el número de artículos citados en cada revista, la suma total se divide por el número total de artículos citados.

– *Factor de prestigio ponderado (FPP)*. Éste se calcula a partir del factor de prestigio (FP) de una revista y del factor de prestigio medio de las revistas donde se producen las citas (FPMRC). Se obtiene calculando el promedio de ambos factores.

$$\text{FPP} = \frac{\text{FPMRC} + \text{FP}}{2}$$

– *Porcentaje de interacción parcial de citas (PIPC)*. Se refiere al tanto por ciento de artículos citados de una revista sobre otra revista (citas entre sí en un sentido) entre las que se analiza la interacción. El porcentaje de interacción parcial se calcula obteniendo el total de citas procedentes de una o más revistas en un año y calculando el porcentaje en función del número de citas recibidas por esas revistas en el mismo año. A continuación se presenta como ejemplo el PIPC de la revista *Psicothema* sobre la revista *Personality and Individual Differences* y viceversa:

Citas de *Psicothema* en el año 2000= 144

Citas de *Pers Indiv Differ* en el año 2000= 2.836

Citas de *Psicothema* en *Pers Indiv Differ*= 39

Citas de *Pers Indiv Differ* en *Psicothema*= 1

PIPC de *Psicothema* sobre *Pers Indiv Differ*= 1,37%

PIPC de *Pers Indiv Differ* sobre *Psicothema*= 0,69%

Estos porcentajes de interacción significan que la revista *Psicothema* representa el 1,37% de las citas que recibe *Pers Indiv Differ* en un año, lo cual es una cantidad considerable si se tiene en cuenta que la primera es una revista general y editada preferentemente en castellano, mientras que la segunda es una revista especializada en Psicología de la personalidad y diferencias individuales y editada en inglés. En cuanto a la relación inversa entre estas revistas, se puede observar que *Pers Indiv Differ* representa el 0,69% de las citas que recibe *Psicothema*.

– *Porcentaje de interacción mutua de citas (PIMC)*. Se refiere al tanto por ciento de artículos citados entre dos o más revistas (citas entre sí en ambos sentidos) entre las que se analiza la interacción, pero no se incluyen las autocitas. El porcentaje de interacción se calcula obteniendo el total de citas comunes entre dos o más revistas en un año y calculando el porcentaje en función del número de citas recibidas por esas revistas en el mismo año. A continuación se presenta el porcentaje de interacción mutua de citas para las dos revistas comentadas anteriormente:

Total citas de *Psicothema* y *Pers Indiv Differ* (2000)= 2.980
 Total citas comunes= 40
 PIMC de *Psicothema* y *Pers Indiv Differ*= 1,34%

Un PIMC entre las anteriores revistas de 1,34% indica que ambas revistas comparten más de un uno por ciento de citas comunes. No obstante, si se analiza la interacción parcial, tal como se ha comentado anteriormente, se pone de manifiesto que este porcentaje de interacción mutua se debe fundamentalmente a la influencia de *Psicothema* sobre *Pers Indiv Differ*. Es evidente que los porcentajes de interacción parcial y mutua deben ser interpretados también en relación al parecido temático de las revistas que se comparan.

El sistema de revisión «por iguales» para evaluar la calidad: ventajas e inconvenientes

Las revistas científicas utilizan un sistema de evaluación conocido como «evaluación por pares o iguales». Según este sistema los trabajos que llegan a una revista son enviados a otros investigadores supuestamente especialistas en la materia objeto de investigación. Normalmente los revisores (que pueden oscilar entre dos y seis) reciben el trabajo anónimo, una vez revisado se remite al director de la revista, quien en función de las opiniones de los revisores y de la suya propia toma una decisión final. Los revisores son también anónimos para los investigadores que someten el trabajo para revisión. Este sistema parece tener muchas ventajas, como: una revisión «imparcial» por no conocer los investigadores, una revisión realizada por especialistas, una coherencia temática de la revista pues el director y revisores también deciden sobre la idoneidad del trabajo en relación a la revista. Sin embargo, si se analiza con detalle todo el proceso podemos concluir que también presenta varios inconvenientes:

Los sesgos del director

Aunque cada revista tenga un área temática delimitada, el director siempre tiene un margen para potenciar la publicación de determinados temas o áreas, y por tanto dificultar la publicación de otras. Para ello hay varios medios:

- a) Una es la selección del comité de revisores, el director elige directamente en unos casos y aconseja o influye en otros sobre quién compone el comité. Se suele considerar que esto afecta más a las revistas europeas e iberoamericanas, donde hay una mayor tradición de que el director se mantenga de forma vitalicia o por muchos años en el cargo, a diferencia de las revistas estadounidenses, en las cuales el director cambia periódicamente. Pero en realidad, lo único que afectaría es que el sesgo es más variable o más constante, pues por el hecho de cambiar al director no se elimina el sesgo.

- b) El director decide a qué revisores envía el trabajo y sabe que no todos son igual de estrictos, por ello, la decisión de a quién envía el trabajo ya influye en alguna medida sobre su posible publicación.
- c) La decisión final de la publicación la toma el director, y éste se puede encontrar con informes contradictorios o al menos diferentes, por tanto debe tomar decisiones sobre si enviarlo a otros revisores o inclinarse por algunos de los que ya han evaluado.
- d) Algunas revistas suelen tener una mayor cantidad de trabajos bien evaluados de los que pueden publicar. El director decide cuáles de los trabajos bien evaluados serán definitivamente aceptados, y sin duda, el director tendrá preferencia por unos u otros temas.
- e) El «efecto tornado de citas», que consiste en que los autores muy citados tienen más facilidades para publicar pues potencian las citas de la revista. No hay que olvidar que el director sí sabe quiénes son los autores, lo que sin duda puede influir en su decisión. Por ejemplo, entre un trabajo bien evaluado de un autor muy conocido y uno desconocido seguramente se inclinará por el primero.
- f) La tendencia a publicar estudios en los que se encuentran efectos o correlaciones y a rechazar aquellos que siendo metodológicamente correctos no se encuentran resultados positivos. Esto es menos frecuente en la actualidad, pero se sigue dando y fue una práctica habitual en décadas pasadas (véase ejemplo en el cuadro 2). Esto supone un considerable sesgo en los estudios publicados.

Los comentarios anteriores no tienen la intención de una crítica destructiva del trabajo de los directores, sino que pretenden llamar la atención sobre el hecho de que un director influye en mayor o menor medida en la decisión final de la publicación de un trabajo. Los sesgos del director son implícitos a este sistema, pero no por ello anulan el sistema.

Cuadro 2

Ejemplo de la tendencia a publicar estudios en los que se encuentran efectos y a rechazar los que no encuentran

El autor de este artículo, cuando era un estudiante de doctorado, envió un trabajo para su posible publicación en una revista incluida en el *Journal Citation Reports*. El director de la revista devolvió el trabajo diciendo que aunque era metodológicamente correcto no podía aceptarlo para su publicación pues en realidad no demostraba ningún efecto de la variable independiente sobre las dependientes. El autor se quedó sorprendido, pues la muestra del estudio era diez veces superior a la de otros estudios similares, algunos publicados en la misma revista, y en los que sí se encontraban efectos. La conclusión era clara, lo importante no era ni usar una metodología adecuada ni el tamaño de la muestra, lo realmente importante era «encontrar efectos». Así las cosas, el joven investigador decidió archivar el trabajo y no enviarlo a ninguna otra revista, pues ya bastante atrevimiento había sido enviarlo a una revista de «prestigio».

Pasaron bastantes años y ya siendo profesor el autor del trabajo, con motivo de un cambio de universidad, encuentra en el fondo de un cajón de su despacho el artículo que había olvidado totalmente. Ya instalado en la nueva universidad decidió actualizar la revisión teórica, exclusivamente eso. Envío el trabajo a una revista incluida en el *Journal Citation Reports*, pero con un factor de impacto **cuatro veces superior** al que tenía la revista que lo había rechazado varios años antes. El trabajo fue aceptado sin ninguna crítica y fue publicado. Alguien podría pensar que la madera del cajón había tenido el mismo efecto que la madera de roble sobre el vino cuando envejece en los toneles. Pero ni la mesa era de roble, ni esa explicación parecía la más adecuada. Había otras posibles explicaciones: falta de fiabilidad entre revisores, intencionalidad por parte de revisores especialistas en el tema que sí encontraban efectos, sesgos de los directores, etc. Sin embargo, el autor atribuyó la aceptación al cambio de tendencia que se había dado en las revistas al considerar también aquellos trabajos en los que no se encuentran efectos. Esto no implica que esa atribución fuese la correcta, pero en cualquier caso, qué importa, si lo importante era publicar el trabajo, al menos así pensaba en aquel momento el autor.

La cualificación y la objetividad de los revisores

Los revisores no son tan cualificados, independientes y objetivos como se podría creer y ello queda de manifiesto por lo siguiente:

- a) La elección de los revisores no es perfecta, en algunas revistas se nombra directamente por el director, para lo cual no sólo se sigue el criterio de ser un especialista, sino otros como prestigio del revisor, amistad con el director, etc. En otros casos, como por ejemplo en las revistas de la APA se hacen periódicamente anuncios para que se presenten candidatos que cumplan las siguientes condiciones: haber publicado previamente en revistas con sistema de revisión, ser lector habitual de cinco o seis revistas de un área, estar especializado en un tema y disponer de suficiente tiempo para realizar las revisiones. Por tanto, es evidente que no se selecciona a los mejores posibles, sino a los que se presentan.
- b) Los revisores no son más cualificados que los autores. De hecho, ocurre que en algunas ocasiones los autores son más conocidos, pues han publicado más que los revisores, por tanto, al menos se puede cuestionar la autoridad que tiene el revisor para juzgar el trabajo del autor. Un buen ejemplo es el caso de dos autores que en el anexo de su artículo critican y dan consejos a los revisores de la misma revista para que mejoren el sistema de revisión (véase Montero y León, 2001).
- c) Los revisores no son mejores cuando revisan que cuando investigan. Si los revisores también hacen investigaciones, que luego son juzgadas por otros «iguales» y que podrían ser rechazadas, se entra en una contradicción: se consideran cualificados para evaluar pero, sin embargo, sus trabajos pueden ser rechazados.
- d) Los revisores aprenden «por experiencia». Los revisores no han tenido una formación o aprendizaje previo de cómo revisar un artículo, por tanto, revisan los trabajos según su opinión y experiencia. Cuando un revisor evalúa su primer artículo, ¿con qué criterio lo hace?, ¿cuándo y dónde aprendió?
- e) Falta de fiabilidad entre revisores. No son excepciones en las que los revisores discrepan parcial o totalmente ante un mismo trabajo. Esto se explica en parte por lo comentado en el punto anterior, pero lo importante es que el sistema de revisión «por iguales» no es fiable (véase en el cuadro 3 un ejemplo de un caso real de falta de fiabilidad entre revisores).
- f) Los sesgos de los revisores. Además de los sesgos particulares de cada revisor, el hecho de que sean especialistas en el tema y concededores de las teorías sobre un campo concreto de trabajo implica un cierto sesgo en aceptar trabajos coherentes con la situación actual y rechazar trabajos innovadores. Esto es una limitación para aquellos investigadores más creativos.
- g) El anonimato de los revisores facilita críticas intencionadas, desproporcionadas o crueles. Si los revisores son especialistas en un área de trabajo es que han realizado investigaciones y publicaciones en ese campo. Si un trabajo sometido a revisión crítica o los resultados contradicen total o parcialmente sus trabajos es más probable que el revisor tienda a rechazar el trabajo, lo cual facilita el anonimato. Véase en el cuadro 4 un ejemplo de un caso real de una crítica despiadada apoyándose en el anonimato. El revisor anónimo llegó a escribir que el texto parecía escrito por un «abogado charlatán» y que se equiparaba con el trabajo de un alumno

Cuadro 3
Ejemplo de falta de fiabilidad entre revisores

A continuación se recogen los comentarios de los revisores de dos revistas incluidas en *Journal Citation Reports* y con el mismo factor de impacto sobre un mismo artículo que fue sometido para revisión. Como se puede ver, mientras los revisores de la revista «A» desacreditan y rechazan de manera radical el artículo, los revisores de la revista «B» aceptan el artículo y además hacen comentarios bastante positivos (se trata de un caso real cedido por un investigador que prefirió mantener el anonimato).

Revista A: El artículo fue rechazado siguiendo las recomendaciones de los revisores:

Revisor 1: «El manuscrito lo considero de pobre calidad, plagado de problemas y no adecuado para su publicación en esta revista».

Revisor 2: «El artículo presenta limitaciones metodológicas y conceptuales importantes, que limitan considerar su pertinencia para publicación».

Revisor 3: «No tiene sentido abundar más sobre un artículo de tan lamentable nivel. Mi única duda es por qué un artículo así se envía a evaluación y no se decide a nivel del Comité Editorial su rechazo».

Revista B: El artículo fue aceptado con mínimos cambios sobre aspectos formales, y algunos comentarios de los revisores fueron:

– «El artículo está claramente escrito y tanto el texto como las figuras presentan la información de manera directa, resumida y didáctica».

– «Los autores han presentado de manera lúcida los resultados presentados en el artículo».

– «Los autores han hecho un muy buen resumen de los resultados».

Éste es un claro ejemplo de la falta de fiabilidad entre evaluadores, pues en la revista A no se rechaza el trabajo por no adecuarse a la línea de la revista, sino por la opinión de los revisores sobre la calidad que tienen los trabajos. Este ejemplo también sirve para poner de manifiesto que el factor de impacto no se corresponde con el nivel de calidad, pues dos revistas teniendo el mismo factor de impacto, una rechaza el trabajo y otra lo publica.

Nota: sin duda alguna los revisores de la revista A también podrían ser un buen ejemplo de las revisiones crueles.

de primer curso universitario, el autor anónimo resultó ser Robert Sternberg, actual Presidente de la *American Psychological Association* (APA), quien decidió hacer públicas las crueles críticas y que se resumen en el cuadro 4 (Sternberg, 2003).

Los parámetros y criterios establecidos sobre qué aspectos evaluar

Las revistas tienen unos formularios de revisión que son enviados a los revisores, junto con el trabajo. Estos formularios tienen la función de establecer los parámetros y/o criterios que se deben seguir en la evaluación del trabajo. Pero si se analizan esos criterios se puede comprobar que no aseguran totalmente calidad. Por ejemplo, algunos de los más frecuentes son: relevancia del tema tratado, rigor metodológico, claridad expositiva, aportaciones del trabajo, uso correcto del lenguaje, adecuación de la bibliografía, etc. Sin embargo, salvo con algunas excepciones no se evalúa la validez interna y externa, la utilidad, la implementación, la originalidad y/o innovación, etc.

Es cierto que algunos de los parámetros evaluados en el proceso de revisión, tales como «relevancia del tema», «rigor metodológico», «aportaciones del trabajo» necesariamente están relacionados con la calidad del trabajo. Sin embargo, el problema es que se evalúan de forma muy general y queda al arbitrio de los revisores la aplicación de estos parámetros. Por tanto, la cuestión no es que no evalúan calidad, sino que el problema está en la forma de evaluar, es demasiado genérica, lo cual facilita la subjetividad de los revisores.

Cuadro 4
Ejemplo de revisiones «cruelles»

SOBRE LA CORTESÍA EN LAS REVISIONES CIENTÍFICAS
Robert J. Sternberg

Muchos de nosotros hemos dedicado nuestros esfuerzos mejor intencionados a escribir artículos para revistas científicas o propuestas de beca y recibido revisiones crueles de los mismos. Recientemente he recibido una de estas revisiones de un artículo co-escrito y enviado a una publicación que refiriéndose al mismo afirmaba que sonaba como si lo hubiera escrito un «abogado charlatán», y calificaba partes del mismo como «absurdas» o «galimatías». Se comparaba su argumentación con la de «ensayos de primer curso universitario», y recomendaba que el autor, quien «se encuentra seriamente fuera de su elemento en este tema... evite aventurarse en áreas que exceden su competencia profesional».

Otros comentarios de la revisión no diferían mucho de éstos. Afortunadamente, la despiadada revisión era «confidencial», aunque por supuesto, el editor, otros revisores, y no se sabe quién más, vieron esta revisión confidencial. También hubo una segunda revisión del artículo mucho más favorable.

No sé si otros revisores hacen lo mismo, aunque tengo entendido que cuando se trata de revisar artículos, no existe campo en que los revisores ataquen de forma más cruel que la de algunos psicólogos. De lo que tengo constancia es que más de una vez en mi carrera profesional me ha asombrado la crueldad con la que se atacaba mi propio trabajo y el de otros, generalmente bajo el manto del anonimato. Me ha sorprendido tanto lo escrito por los revisores como que los editores (o los presidentes de las comisiones que asignaban las becas) remitieran las revisiones de esta forma. Tales revisiones son, en mi opinión, inaceptables, con independencia de la calidad (o falta de calidad) de los manuscritos examinados.

Los profesores universitarios emplean bastante tiempo enseñando a los estudiantes de postgrado a escribir artículos, pero no dedican nada a enseñarles a revisar artículos (o libros, o propuestas de beca). Quizá piensan que hacer revisiones es una destreza natural de las personas. Es evidente que no y es probable que se necesite una formación reglada.

Las revisiones crueles son negativas para todos por varias razones:

Ética fundamental. En relación con los principios éticos, las revisiones crueles violan la regla de oro: actuar con los demás como queremos que actúen con nosotros. Nadie quiere ser tratado de manera irrespetuosa ni insultante, sea quien sea.

Equilibrio. En realidad es inusual que un artículo o propuesta de beca carezca por completo de algún valor positivo. La tarea de un revisor es proporcionar una evaluación equilibrada de un artículo o propuesta de beca, señalando el valor positivo así como los defectos en lo examinado. Las revisiones despiadadas son desequilibradas y por tanto con ellas los revisores no cumplen su cometido. Los comentarios que éstos proporcionan resultan desequilibrados y sesgados. Incluso si los artículos han de ser criticados severamente, tales críticas se pueden transmitir con cortesía.

Exageración. Es infrecuente que más de uno de los revisores sea del todo negativo. A menudo otros revisores son incluso positivos. Los artículos y propuestas son pocas veces tan inaceptables como consideran estos revisores. Al hacer estas críticas despiadadas los revisores normalmente fuerzan la verdad.

Mantener la credibilidad para producir un cambio. En general no conseguimos cambiar la conducta de las personas degradándolas ni insultándolas. Al contrario, perdemos credibilidad precisamente ante aquellos cuya conducta queremos cambiar. Como consecuencia, estas personas no estarán dispuestas a escucharnos. El contenido personal de la revisión reducirá su eficacia en lugar de aumentarla.

Socavar el rendimiento. Aunque no siempre, los investigadores con mayor experiencia normalmente toman con calma las revisiones despiadadas. Ya las han recibido antes y saben que las volverán a recibir. Además suelen ser titulares y catedráticos de modo que no tienen que preocuparse porque éstas comprometan sus puestos de trabajo. Sin embargo, los investigadores jóvenes que acaban de empezar quedan sorprendidos, descorazonados y hasta pierden la esperanza de conseguir el éxito en el campo de estudio porque no se dan cuenta de que el problema está en el revisor, no en el material revisado. Puede que el trabajo efectivamente requiera una considerable reelaboración, pero aun así, no hay necesidad de atacarlo cruelmente (Sternberg, 2003). Reproducido con permiso de R. Sternberg.

La calidad de un artículo no debe evaluarse sólo por el impacto o prestigio de la revista en la que se publica

Sternberg (2001) y Buela-Casal (2002) proponen quince razones según las cuales es un error dar más importancia al «dónde» se publica un artículo, que al artículo en sí mismo:

1. Dar más importancia a la revista donde se publica que a lo que se publica. Es más fácil cuantificar las citas o evaluar dónde se publican que leer el artículo, pero el impacto de la revista no es un sustituto de la evaluación crítica del trabajo.
2. El conservadurismo de las revistas más prestigiosas. Normalmente las revistas más prestigiosas son más conservadoras y, por ello, los revisores suelen ser más estrictos para comprobar si el trabajo se adecúa a las normas más convencionales (ver cuadro 5).

Cuadro 5
Ejemplos de conservadurismo de las revistas más prestigiosas

«Las revistas más prestigiosas son a menudo, en mi experiencia, más conservadoras. Yo opino esto después de haber sido director de dos revistas, director asociado de otras dos y revisor de muchas revistas con un rango desde bajo prestigio hasta muy alto. En cierto sentido, las revistas más prestigiosas tienden a ser algo más conservadoras. Cuanto más rigurosa es la revisión más se preocupa el revisor de que el artículo se ajuste a lo convencionalmente establecido. Pero las normas convencionales pueden servir o no en casos particulares. De hecho, cuando yo fui presidente de la *Society for General Psychology*, fundé una revista, *Journal of General Psychology*, debido a mi creencia de que las revistas punteras pueden ser más conservadoras de lo que sería ideal» (Sternberg, 2001, p. 3).

«El caso de mayor resonancia es quizás el del premio Nobel Rosalyn Yalow, cuyo artículo fundamental acerca de la radioinmunología sufrió el rechazo de dos prestigiosas revistas, una de las cuales se justificó con la curiosa y ambigua argumentación: Las personas verdaderamente imaginativas y creativas, sostenían los encargados de las reseña, no pueden ser juzgadas por sus iguales, porque no tienen» (Di Trocchio, 1993, p. 137).

3. Dificultades a las publicaciones de investigaciones interdisciplinarias. Teniendo en cuenta que las revistas de mayor prestigio suelen estar encuadradas en campos tradicionalmente definidos, resulta difícil encontrar revistas con alto prestigio y que sean interdisciplinarias, por lo que este tipo de trabajos suelen ser «penalizados», pues tampoco resulta fácil publicarlos en las revistas generales. Un buen ejemplo de esto es la psiconeuroinmunología, y especialmente el nacimiento de ésta con el experimento de Ader y Cohen, el cual encontró importantes dificultades para su publicación.
4. Dificultades de la investigación no paradigmática. Si una investigación no se enmarca en los paradigmas convencionales de investigación tendrá más dificultades para ser publicada, pues los revisores tienden a ser convencionales a la investigación, por lo que los investigadores tienden a investigar en los paradigmas convencionales, y los que no lo hacen suelen ser penalizados con dificultades en la publicación. Aquí también podemos recurrir como ejemplo a la investigación de Ader y Cohen, una vez que consiguieron publicarla, luego no resultó difícil publicar trabajos posteriores en importantes revistas (*Science, Lancet, Annual Review of Psychology*).
5. Inconvenientes de publicar en libros y otros tipos de publicación que no sean revistas. Por ejemplo en España (y en otros países europeos) para la evaluación y promoción de los investigadores y profesores universitarios sólo se valoran las publicaciones en revistas, y especialmente las que están incluidas en el *Journal Citation Reports*.
6. La profecía autocumplida. Se refiere a que los artículos publicados en revistas de prestigio tienden a ser más citados que artículos publicados en revistas de más bajo prestigio, por lo que consecuentemente tiende a incrementar o mantener el prestigio de las primeras.

7. El «efecto Mateo». «Porque a cualquiera que tuviese, le será dado y tendrá más; y al que no tuviese, aun lo que tiene le será quitado». Esto se aplica a que las revistas que tienen alto prestigio tienden a recibir más y mejores artículos que las revistas con bajo prestigio.
8. No todos los artículos publicados en una revista tienen el mismo «impacto». Una crítica importante que hay que hacer a los distintos índices bibliométricos es la adjudicación del mismo «impacto» a cada uno de los artículos publicados en una misma revista, pues los factores de impacto y de prestigio se calculan de forma general para la revista. Y es evidente que unos artículos reciben más citas que otros, el sistema es tan injusto que los artículos que reciben pocas citas penalizan a los muy citados. Un buen ejemplo son algunas revistas que publican las estadísticas anuales de un trastorno o enfermedad, ese artículo suele incrementar indirectamente el factor de impacto de los demás artículos publicados en el mismo año.
9. La elección de los autores sobre a qué revista someter el trabajo influye en el nivel de impacto que tendrá. Imaginemos que hay un artículo de excelente calidad, que sería susceptible de ser publicado en una revista del alto prestigio, pero los autores deciden enviarlo a una revista de prestigio medio (por diversos motivos puede ocurrir: prisa por publicar, desconocimiento del sistema, la revista es más afín, etc.), supongamos que se publica. Pues bien, ¿este artículo por publicarse en esta revista ya tiene menos calidad?
10. El sistema de revisión «por iguales» no garantiza calidad. Tal como se ha comentado anteriormente, este procedimiento de evaluación está influido por múltiples sesgos del director y de los revisores que sin duda afectan seriamente a la evaluación y por tanto no hay una relación total con la calidad.
11. No todos los artículos rechazados en una revista tienen poca calidad. Se dan casos de que algunas revistas reciben muchas propuestas de trabajos y el resultado puede ser que hay más artículos bien evaluados de los que se pueden publicar, por lo que el director tiene que rechazar trabajos bien evaluados. Éstos terminan siendo publicados en otras revistas y quizá con menor impacto. Pero la pérdida de impacto es consecuencia de la cantidad de trabajos y no de la calidad.
12. Los artículos publicados en revistas con «impacto» ni siquiera tienen garantía de veracidad. En la historia de las publicaciones científicas se dieron bastantes casos de publicaciones falsas o al menos manipuladas en parte. Y esto afecta por igual a revistas de alto o bajo prestigio.
13. El número de citas puede ser manipulado por distintos procedimientos. Existen múltiples estrategias por medio de las cuales se puede incrementar el número de citas de una revista, que es lo mismo que incrementar el factor de impacto o de prestigio, y ello independientemente de la calidad de los artículos publicados en la revista.
14. Que una revista tenga *factor de impacto* y *factor de prestigio* no sólo depende de que reciba citas, sino de que su director y/o institución de la cual depende realice los trámites necesarios para que pueda ser incluida en estos sistemas de cuantificación de citas, de hecho, hay muchas revistas que nunca realizaron esos trámites formales, pero, como es lógico, esto no se relaciona con la calidad.
15. El idioma en que se edita una revista influye en el *factor de impacto* y en el *factor de prestigio*, pues el idioma influye en el número de citas que recibe un artículo. Así, en la actualidad, la mayoría de los investigadores lee y publica preferentemente en inglés, por tanto, las revistas editadas en inglés serán más citadas que las editadas en otro idioma; y es evidente que nadie defendería que el idioma en que se publique influye en la calidad de la investigación.

Una alternativa para evaluar la calidad de los artículos y de las revistas. Un índice de calidad

En primer lugar habría que establecer qué se entiende por calidad de los artículos y revistas científicas. Parece lógico que debe ser algo más que sean muy citadas, en principio habría que plantearse qué criterios se deben seguir para evaluar, qué características deben tener los evaluadores y qué se debe evaluar (los artículos y las revistas).

Criterios para evaluar

Existe un importante repertorio de pautas para la evaluación de artículos científicos, que se recogen en varios libros sobre investigación y/o estilo. Sin embargo, pocos se centran en los aspectos importantes en relación a la calidad, pues en su mayoría se limitan a los aspectos formales sin cuestionar el sistema general de evaluación.

A continuación se presenta una serie de parámetros, algunos ya fueron propuestos por Sternberg (1988) y por Buéla-Casal (2002), para evaluar artículos científicos en Psicología y otras ciencias de la salud, aunque la mayoría de estos criterios son aplicables a otros campos de conocimiento. Un artículo y/o una revista que cumpla los siguientes criterios o la mayoría de ellos puede considerarse como un artículo o revista de calidad, aunque evidentemente deben cumplirse otros criterios relacionados con los aspectos formales y de estilo.

1. Aportar resultados sorprendentes que tienen sentido en algún contexto teórico.
2. Aportar resultados de gran importancia teórica o práctica.
3. Las ideas planteadas son nuevas e interesantes y pueden aportar un nuevo enfoque para tratar un viejo problema.
4. La interpretación que se hace de los resultados es inequívoca.
5. Se consigue integrar en un marco nuevo y más simple resultados que antes implicaban un marco más complejo y fragmentado.
6. Se desacreditan ideas anteriores que parecían incuestionables.
7. Se presentan investigaciones en las que hay un paradigma especialmente ingenioso o novedoso.
8. La investigación tiene suficiente validez interna debido a que se utiliza un diseño y una metodología adecuada.
9. La investigación tiene suficiente validez externa dado que los resultados y/o la teoría presentados son generalizables.
10. En las investigaciones se describe de forma suficiente el método y procedimiento para que otros investigadores puedan replicarlas.
11. Los resultados teóricos o prácticos tienen un alto grado de implementación.

12. En la investigación se aportan resultados teóricos o prácticos que son útiles a la sociedad.
13. Se especifica de forma clara el tipo de estudio de que se trata (León y Montero, 1997; Montero y León, 2001, 2002).
14. Se realiza una adecuada formulación e interpretación de las hipótesis (Borges, San Luis, Sánchez y Cañadas, 2001).
15. En los estudios clínicos se utilizan técnicas y programas de tratamientos de eficacia contrastada (Bados, García y Fuste, 2002; Echeburúa y Corral, 2001; Fernández-Hermida y Pérez Álvarez, 2001; Hamilton y Dobson, 2001; Méndez, 2001; Pérez Álvarez y Fernández-Hermida, 2001).
16. Los instrumentos de evaluación utilizados tienen calidad y garantías científicas (Muñiz y Fernández-Hermida, 2000; Prieto y Muñiz, 2000).
17. En los estudios de casos se siguen unas normas adecuadas para la presentación del caso (Bucla-Casal y Sierra, 2002).
18. En los estudios instrumentales sobre construcción de instrumentos se sigue la metodología adecuada (Muñiz, 1996).
19. En los estudios instrumentales sobre la adaptación de instrumentos se sigue una metodología adecuada (Muñiz y Hambleton, 1996).
20. Se cumplen las normas y los principios deontológicos (Almeida y Bucla-Casal, 1997; Colegio Oficial de Psicólogos, 1987; Muñiz, 1997).

Éstos son algunos de los posibles parámetros, pues podrían incluirse otros, y deben aplicarse en función del tipo de trabajo de que se trate. Pero lo más importante es que dichos parámetros se apliquen con los mismos criterios, los cuales deben ser facilitados a los revisores, tal como se explica más adelante (véase también Bobenrieth, 2002).

Qué evaluar: los artículos y las revistas

La evaluación de las publicaciones científicas debería realizarse a dos niveles (artículos y revistas), siguiendo los criterios mencionados anteriormente. En un primer nivel, tal como se suele realizar en la actualidad, aunque sería conveniente que los comités fuesen mejor seleccionados y quizá debería replantearse la cuestión de que las evaluaciones sean anónimas. En un segundo nivel se pueden evaluar las revistas por un comité mixto de expertos profesionales. Cada revista se evaluaría anualmente y recibiría una puntuación sobre la calidad. La evaluación debería realizarse independientemente del idioma en que esté editada, la institución que la financia, la sociedad que representa, o cualquier otro factor que pueda suponer un seudoprestigio a la revista. Este índice sin duda tendría mucha más relación con la calidad que el mero factor de impacto o factor de prestigio, que en realidad no son más que índices del nivel de difusión.

Quién debe evaluar

Los comités de las revistas deberían ser mejor seleccionados, pues sería conveniente contar no sólo con especialistas en el tema, sino también con expertos en metodología y diseño de investigación. Cada revisor debería recibir un manual de funcionamiento con los parámetros y criterios de evaluación y cómo aplicarlos (véase por ejemplo, Bobenrieth, 2002). Esto sin duda aumentaría la validez de la evaluación y la fiabilidad entre evaluadores. Los centros o empresas dedicadas a la evaluación de las publicaciones

científicas, como por ejemplo: *Institute for Scientific Information*, *Prestigefactor* o *Latindex*, tendría que disponer de comités de expertos profesionales para evaluar anualmente la calidad de las revistas. Estos comités estarían compuestos por expertos en las áreas en las que se clasifican las revistas, expertos en epistemología, expertos en metodología y diseño de investigación. Incluso se podrían establecer cursos de formación para formar especialistas en evaluación de la calidad de las publicaciones científicas. La evaluación de estos comités es ajena a la dirección de las revistas y tiene la posibilidad de detectar y evaluar los sesgos de los directores y de los revisores de revistas. Esta revisión por comités de expertos también facilitaría el descubrimiento de plagios e investigaciones falsas, pues los mismos expertos revisan todas las revistas de un campo de conocimiento, es evidente que no se eliminarían por completo, pero ayudaría a reducirlos. También habría que replantearse si los revisores deben ser anónimos o no, pues aunque el anonimato tiene ciertas ventajas y por eso se ha instaurado, hay que reconocer que tiene importantes inconvenientes, un buen ejemplo son las revisiones desproporcionadas o despiadadas que realizan algunos revisores escondiéndose en el anonimato. Este tipo de revisiones, tal como ha analizado Sternberg (2002), genera indefensión especialmente en los investigadores más jóvenes y no suponen ninguna contribución positiva al proceso de evaluación de las publicaciones científicas. Diversos estudios demostraron que las revisiones son más específicas y más constructivas cuando los revisores firman sus críticas. El anonimato es una «receta» para la falta de responsabilidad en las críticas (Shashok, 1997). Por tanto, parece que el anonimato tiene más inconvenientes que ventajas, por lo que en el futuro debería plantearse la identificación de los revisores.

Cómo construir un índice de calidad de las revistas

El índice de calidad sería posible si se realiza por revisores profesionales debidamente formados, se utilizan unos criterios que evalúen realmente calidad (por ejemplo, los criterios propuestos anteriormente), y se siga un manual de funcionamiento para establecer la forma de aplicar esos criterios. Cada miembro del comité de expertos evaluaría cada uno de los artículos editados por una revista en un año, asignando puntuaciones según los criterios establecidos. La puntuación final de cada revista se obtiene calculando el promedio de puntuaciones otorgadas por los revisores y el resultado se transformaría en una escala única (por ejemplo, de cero a cien). Un sistema como éste permitiría obtener:

- a) Índice de calidad para cada artículo.
- b) Índices de calidad anual de cada revista.
- c) Análisis de las variaciones intraanuales de los índices de calidad según los distintos números editados en el mismo año.
- d) Análisis de las variaciones entre años de los índices de calidad según los distintos volúmenes editados por la revista en distintos años. Lo cual permite establecer índices de crecimiento o decremento de las revistas en función de la calidad.
- e) Índices específicos para cada uno de los criterios establecidos en la revisión, por ejemplo: índice de validez externa/interna, índice de innovación, índice de implementación, índice de utilidad, etc.
- f) Una comparación de la calidad entre revistas de un mismo área temática y entre revistas de distintas áreas dado que se usa la misma escala.

- g) Una clasificación de las revistas según categorías relacionadas con los índices. Esto permite a los investigadores orientarse sobre qué revistas son más adecuadas según su investigación: las que potencian la utilidad y/o implementación, las que priman la validez interna y/o externa, las que prefieren investigaciones innovadoras, etc.
- h) Análisis de la relación entre índices de calidad e índices basados en citas (*factor de impacto*, *factor de prestigio*, etc.).

Por otra parte, los distintos índices de calidad no estarían afectados por el número de citas y por los factores que afectan al *factor de impacto* y al *factor de prestigio*, y por tanto no se confundiría calidad con difusión.

Discusión

En este trabajo se ha discutido una de las limitaciones que presenta el factor de impacto y el factor de prestigio en los estudios bibliométricos. Esta limitación se refiere al hecho de asignar el mismo valor a una cita independientemente del «impacto» o «prestigio» que tiene la revista donde se produce la cita. Una solución a este problema es la aplicación de factores ponderados, que además del número de citas tienen en cuenta el tipo de revista donde se citan los artículos.

Para realizar comparaciones entre revistas se propone el uso de los porcentajes de interacción de citas. Estos índices bibliométricos permiten analizar la influencia de las revistas entre sí, en función del porcentaje de citas comunes.

El porcentaje de interacción propuesto aquí no debe confundirse con el índice de interacción propuesto por Factor de Prestigio, puesto que este último se refiere a aquellas publicaciones más «interactivas», como son las cartas al editor o similares. El porcentaje de interacción parcial o mutua es muy útil para comprobar la influencia de las revistas entre sí.

Hay que tener en cuenta que para poder comparar el factor de impacto o de prestigio entre dos revistas o más revistas no sólo deben pertenecer a un mismo campo temático, sino que además hay que tener la seguridad de que las revistas analizadas en un mismo año son revistas «que citan» y «que son citadas». Es decir, que algunas revistas pueden ser citadas pero por distintos motivos pueden no ser editadas ese año (ejemplos: revistas cesadas, suspendidas, cambio de nombre, etc.). Los factores de impacto o de prestigio en estos casos no serían comparables, pues una revista que «es citada» pero que «no cita» tiene normalmente un factor de impacto o de prestigio más bajo dado que no se producen las autocitas, las cuales suelen ser un porcentaje importante de las citas de una revista. Sin embargo, en los índices de interacción basados en los porcentajes de citas comunes no es necesario tener esto en cuenta, pues los análisis se realizan siempre para un mismo año en que las revistas analizadas fueron publicadas.

Hay que resaltar que tanto estos índices bibliométricos tratados en este trabajo como cualquier otro que se base en el número de citas deben ser interpretados como parámetros cuantitativos de citas, que sirven para cuantificar la producción y la difusión de las publicaciones científicas. Y aunque estos índices puedan tener alguna relación con la calidad de los artículos y de las revistas, no deben ser utilizados como sustitutos de otros parámetros más relacionados con la calidad (Bobenrieth, 2002; Buela-Casal, 2002; Buela-Casal y Sierra, 2002; Montero y León, 2001, 2002).

Los sistemas utilizados para calcular el factor de impacto o el factor de prestigio se basan en las citas y artículos publicados en los dos o tres últimos años (respectivamente). Si reflexionamos sobre esto nos podemos preguntar porqué se utiliza este criterio, sin duda, parece más razonable un período de tiempo mucho mayor, pues si un artículo resulta «clave» en un campo de conocimiento será citado durante muchos años. Un período de dos años no es igual de adecuado para las distintas áreas científicas, dado que el «envejecimiento» de las publicaciones no es igual en cada área. Así, mientras que en biología molecular se citan trabajos muy recientes, en otras áreas como la anatomía o la botánica donde generalmente se citan artículos de diez años de antigüedad (Gómez y Bordons, 1996).

Un aspecto importante que hay que tener presente es que las costumbres de citar no son igual en las distintas disciplinas, así, mientras que en ciencias sociales se cita una media de treinta trabajos en cada artículo, en ingeniería se citan diez y en matemáticas sólo cinco, y esto, lógicamente, afecta al factor de impacto. De igual forma, el número de investigadores y de revistas que hay en cada área determina en gran medida el factor de impacto (Aleixandre y Porcel, 2000).

Por otra parte, hay que tener presente que no existe una relación directa entre número de citas y calidad, es decir, que un artículo sea más o menos citado depende de muchos otros factores y, como mucho, la calidad sería uno de esos factores. Recientemente, Buela-Casal (2002) describió diversos procedimientos que pueden ser usados para incrementar las citas de las revistas y ninguno de ellos se refiere a la publicación de artículos de más calidad. Si las citas no se corresponden con la calidad, por lógica tampoco se corresponden los distintos índices bibliométricos (tales como factor de impacto, factor de impacto equivalente, factor de prestigio, etc.) con la calidad. Por ejemplo, algunas citas son críticas negativas de un trabajo y paradójicamente se contabilizan como indicadoras de calidad. Hay muchas razones por las que un autor cita un trabajo y muchas de esas razones no tienen relación con la calidad del trabajo citado (Aleixandre y Porcel, 2000). No debemos olvidar que el propio Gardfiel (1970, 1996), fundador del factor de impacto, advirtió que no hay una relación absoluta entre el factor de impacto y la calidad. Cada día hay más autores que critican la relación entre calidad e índices bibliométricos. Sternberg y Gordeeva (1996) comentan que no hay una relación perfecta entre el impacto de un artículo y el impacto de la revista. Seglen (1997) argumenta que la tasa de citas que recibe un artículo determina el factor de impacto de la revista y no viceversa. Pelechano (2000) resalta que el índice de impacto está influido por el apoyo económico que tienen las revistas y por las modas por publicar determinados temas. De hecho, Garfield (1996) reconoce que para calcular el factor de impacto no se incluyen cientos de revistas normalmente editadas en países del tercer mundo, y por tanto con menos recursos económicos. Por ello, no son adecuadas las comparaciones que se hacen en cuanto a productividad científica y a factor de impacto entre países no equiparados en cuanto a idioma y revistas incluidas en el JCR. Para este tipo de comparaciones se recomienda utilizar el «índice de actividad» y el «factor de impacto relativo» (Bordons, Fernández y Gómez, 2002). Buela-Casal (2001) argumenta que, entre otros factores, el idioma en que se publica afecta al impacto, y lógicamente el idioma no puede dar más o menos calidad. Otros factores importantes son aspectos formales de las revistas (por ejemplo, re-

gularidad y puntualidad en la edición) que nada tienen que ver con la calidad del trabajo; o incluso otros más superfluos como que la dirección de la revista no realice los trámites para que sea incluida en los distintos índices bibliográficos. De manera similar, Sternberg (2001) concluye que el lugar de publicación no es un apoderado válido para la calidad y el impacto de la calidad, por tanto, sólo en casos excepcionales puede darse más importancia a la revista en la que se publica que al propio contenido del artículo.

Otro problema importante, que habitualmente no suele considerarse, es el problema de las limitaciones del proceso de revisión de los artículos por el sistema de evaluación «por iguales». Tal como se ha analizado de forma pormenorizada, los sesgos tanto por parte de los revisores como de los directores presentan serias dudas de que se trate de un sistema válido y fiable. De hecho, Sokal (1996a, 1996b) puso de manifiesto con la publicación de un artículo sin ningún sentido, que el filtro de los revisores no siempre funciona. En los últimos años cada vez más autores se cuestionan tanto la calidad como la veracidad de las publicaciones como equivalente directo del prestigio que tengan las revistas científicas en las que fueron publicadas (Benach de Rovira y Tapia, 1995; Bobenrieth, 2000; Buena-Casal, 2001, 2002; Colom, 2000; Di Trocchio, 1993; Pelechano, 2000; Seglen, 1993, 1997; Sokal y Bricmont, 1999; Sternberg, 1999, 2002, 2003).

Los cuestionamientos anteriores no deben entenderse como una crítica absoluta según la cual este sistema de evaluación de los artículos no tiene ningún valor. Estas críticas deben entenderse en su justa medida, pues, sin duda, si este tipo de revisión no sirviera para nada, tendríamos que concluir que los informes de investigación publicados hasta nuestros días tienen muy pocas garantías, lo cual nadie defendería. Lo importante es considerar que la revisión no es perfecta, y que dentro de las posibilidades de que disponemos hasta la fecha es mejorable, tal como se ha planteado en este trabajo. Sin embargo, debemos tener presente que tanto la labor de los directores de las revistas y de los revisores es, en la mayoría de los casos, altruista, tanto unos como otros dedican tiempo y esfuerzo a unas tareas no remuneradas, que en cierto modo podríamos considerar como altruista. Por ello, difícilmente los directores pueden exigir más dedicación y cualificación a los revisores, aunque sí podría mejorarse la evaluación de los artículos según se ha propuesto en este trabajo. Mientras tanto, debemos tener presente que el sistema funciona con importantes deficiencias, las cuales fueron puestas de manifiesto en diversos estudios (Shashok, 1997). Por ejemplo, Howard y Wilkinson (1998) demostraron que los revisores y los directores coinciden más en aquellos artículos que rechazan que en los que se consideran susceptibles de publicación. Por tanto, el papel del director es fundamental, dado que es quien toma la decisión final y no está claro cómo éste establece los límites de adecuación y de calidad de un artículo (Crammer, 1998).

En este artículo se propone un índice de calidad como una alternativa al proceso de evaluación de las publicaciones científicas, tanto para la evaluación de artículos como de revistas. El índice de calidad nada tiene que ver con las citas, pero sí con la «validez científica» del trabajo.

No hay ningún inconveniente en que se sigan utilizando los criterios de *factor de impacto*, *factor de prestigio* y otros índices bibliométricos, siempre y cuando se utilicen como lo que realmente son: índices basados en la cantidad de citas (véase por ejemplo Sierra y Buena-Casal, 2001). Éstos, aunque aparentemente se ob-

tienen de una forma más objetiva que una evaluación de la calidad, no evalúan calidad. El problema fundamental es que se produce un aparente argumento lógico sin ningún fundamento. Y éste es que los revisores y directores de las revistas tratan de evaluar la calidad, pero una vez publicado el artículo esto se olvida y se equipara calidad con número de citas, lo cual no tiene un fundamento demostrado. Por ello, es necesario crear procedimientos para evaluar la calidad de las revistas.

La propuesta comentada en este artículo sobre la creación de un posible índice de calidad, aunque no sea la solución definitiva al menos serviría para iniciar una mejora en el proceso de evaluación de la calidad de los artículos y de las revistas científicas, que cada día está más desprestigiado. Es evidente que se trataría de un sistema más costoso pues implicaría revisores especializados y profesionales, pero es una forma de evaluar realmente calidad y aunque nunca sería un sistema perfecto, es sin duda mucho mejor que los índices que se limitan a contabilizar citas. Si los parámetros y los manuales de aplicación propuestos para evaluar las revistas se aplicaran también por los revisores de los artículos, sin duda incrementaría la validez y fiabilidad de las revisiones de los artículos.

Por último, un aspecto muy importante a considerar en relación al futuro de las revistas en castellano es que éstas no deben ser evaluadas por medio del factor de impacto, al menos según se hace actualmente en el Institute for Scientific Information. Este sistema de evaluación de revistas está muy sesgado con respecto al idioma de publicación. Un buen ejemplo es el hecho de que sea obligatorio que el título de los artículos, las palabras clave y el resumen sean obligatoriamente en inglés, sea cual sea el idioma en el que se edita la revista; y esto es uno de los requisitos para estar en el JCR. Otro ejemplo es el sesgo que generan algunos revisores cuando revisan un artículo escrito en inglés y critican que algunas de las referencias están en otro idioma (y esto no es un fenómeno aislado).

La presión por publicar en inglés está llevando al uso del castellano a una situación muy preocupante, por un lado cada día son más los investigadores de España y de Latinoamérica que publican sus trabajos en revistas editadas en inglés, debido a que tienen mayor factor de impacto; y por otro, cada año más revistas españolas se editan en inglés. En la actualidad, de las revistas españolas incluidas en el JCR, el 45% se editan sólo en inglés, el resto o en ambos idiomas o sólo en castellano, y esto es una forma de supervivencia, pues las que sólo se editan en castellano tienen dificultades para seguir editándose (Díaz et al, 2001; Bordons, Felipe y Gómez, 2002; Jiménez-Contreras, Faba y Moya, 2001). Es evidente que es necesaria una importante reflexión sobre la situación actual, es imprescindible un apoyo institucional a las revistas científicas que editen en castellano, crear un sistema de evaluación de la calidad para las revistas, por ejemplo, como el índice de calidad propuesto en este artículo, y consecuentemente un cambio de actitud de los investigadores españoles para que envíen también sus mejores trabajos a revistas editadas en castellano.

Agradecimientos

El autor agradece a Robert Sternberg sus comentarios, ideas y material aportado para este trabajo; y a José Ramón Fernández Hermida y a José Muñiz por sus reflexiones, que han contribuido a mejorar la versión final de este artículo.

Referencias

- Alcain, M.D. y Carpintero, H. (2001). La Psicología en España a través de las revistas internacionales: 1981-1999. *Papeles del Psicólogo*, 78, 11-20.
- Aguirre de Cárcer, A. (2001). La contribución de España a la producción científica mundial sigue con su lenta escalada. *ABC* (26-5-2001), 32.
- Aleixandre, R. y Porcel, A. (2000). El factor de impacto y los cómputos de citas en la evaluación de la actividad científica y las revistas médicas. *Trastornos Adictivos*, 1, 264-271.
- Almeida, L. y Buela-Casal, G. (1997). Deontología y evaluación psicológica. En G. Buela-Casal, G. y Sierra, J.C. (Dir.), *Manual de Evaluación Psicológica. Fundamentos, Técnicas y Aplicaciones* (pp. 239-247). Madrid: Siglo XXI.
- Amin, M. y Mabe, M. (2000). Impact factor: use and abuse. *Perspectives in Publishing*, 1, 1-6.
- Bados, A.; García, E. y Fusté, A. (2002). Eficacia y utilidad clínica de la terapia psicológica. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 2, 477-502.
- Benach de Rovira, J. y Tapia, J. (1995). Mitos o realidades: a propósito de la publicación de trabajos científicos. *Mundo Científico*, 154, 124-130.
- Bobenrieth, M.A. (2000). Mitos y realidades en torno a la publicación científica. *Medicina Clínica de Barcelona*, 114, 339-341.
- Bobenrieth, M.A. (2002). Normas para la revisión de artículos originales en ciencias de la salud. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 2, 509-523.
- Bordons, M.; Felipe, A. y Gómez, I. (2002). Revistas científicas españolas con factor de impacto en el año 2000. *Revista Española de Documentación Científica*, 25, 49-73.
- Bordons, M.; Fernández, M.T.; Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country. *Scientometrics*, 55, 196-206.
- Borges, A.; San Luis, C.; Sánchez, J. A. y Cañadas, I. (2001). El juicio contra la hipótesis nula. Muchos testigos y una sentencia virtuosa. *Psicothema*, 13, 173-178.
- Bornas, X. y Noguera, M. (2002). Bases científicas de la Terapia de Conducta: nuevas propuestas para un viejo problema. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 2, 9-24.
- Buela-Casal, G. (2001). La Psicología española y su proyección internacional. El problema del criterio: internacional, calidad y castellano y/o inglés. *Papeles del Psicólogo*, 79, 53-57.
- Buela-Casal, G. (2002). La evaluación de la investigación científica: el criterio de la opinión de la mayoría, el factor de impacto, el factor de prestigio y «Los Diez Mandamientos» para incrementar las citas. *Análisis y Modificación de Conducta*, 28, 455-476.
- Buela-Casal, G.; Carretero-Dios, H. y De los Santos-Roig, M. (2002). Análisis comparativo de las revistas de Psicología en castellano con factor de impacto. *Psicothema* 14, 837-852.
- Buela-Casal, G. y Sierra, J.C. (2002). Normas para la redacción de un caso clínico. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 2, 525-532.
- Carpintero, H. (2001). La Psicología y el problema de España: Una cuestión de Psicología social. *Psicothema*, 13, 186-192.
- Chacón, S.; Pérez-Gil, J. A.; Holgado, F.P. y Lara, A. (2001). Evaluación de la calidad universitaria: validez de contenido. *Psicothema*, 13, 294-301.
- Colegio Oficial de Psicólogos (1987). *Código deontológico*. Madrid: Colegio Oficial de Psicólogos.
- Colom, R. (2000). Algunos «mitos» de la Psicología: entre la ciencia y la ideología. *Psicothema*, 12, 1-14.
- Crammer, J.L. (1998). Invited commentaries on: Peer review and editorial decision-making. *British Journal of Psychiatry*, 173, 114-115.
- Di Trocchio, F. (1993). *Le bugie della scienza. Perché e come gli scienziati imbrogliono*. Milán: Arnoldo Moldadori Editore.
- Díaz, M.; Asensio, B.; Llorente, G.; Moreno, E.; Montori, A.; Palomares, F.; Palomo, J.; Pulido, F.; Senar, J.C. y Tellería, J.L. (2001). El futuro de las revistas científicas españolas: un esfuerzo científico, social e institucional. *Revista Española de Documentación Científica*, 24, 306-314.
- Echeburúa, E. (2002). El sistema de sexenios en la evaluación de la actividad investigadora en el ámbito de la Psicología clínica: una primera reflexión. *Análisis y Modificación de Conducta*, 28, 391-404.
- Echeburúa, E. y Corral, P. (2001). Eficacia de las terapias psicológicas: de la investigación a la práctica clínica. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 1, 181-204.
- Fernández-Hermida, J.R. y Pérez Álvarez, M. (2001). Separando el grano de la paja en los tratamientos psicológicos. *Psicothema*, 13, 337-344.
- Garfield, E. (1955). Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, 122, 108-111.
- Garfield, E. (1970). Citation index for studying science. *Nature*, 227, 669-671.
- Garfield, E. (1996). How can impact factors be improved? *British Medical Journal*, 313, 411-413.
- Gil-Roales, J. y Luciano, M.C. (2002). A la calidad por *the quantity* (porque la cantidad no vale). Algunas reflexiones sobre los criterios de evaluación de la evaluación de la calidad de la investigación psicológica. *Análisis y Modificación de Conducta*, 28, 431-454.
- Gómez, I. y Bordons, M. (1996). Limitaciones en el uso de los indicadores bibliométricos para la evaluación científica. *Política Científica*, 46, 21-26.
- Gross, P.L. y Gross, E.M. (1927). College libraries and chemical education. *Science*, 66, 385-389.
- Hamilton, K.E. y Dobson, K.S. (2001). Empirically supported treatments in psychology: implications for international promotion and dissemination. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 1 (35-51).
- Howard, L. y Wilkinson, G. (1998). Peer review and editorial decision-making. *British Journal of Psychiatry*, 173, 110-113.
- Institute for Scientific Information (2001). *Journal Citation Reports*. Philadelphia: ISI.
- Jiménez-Contreras, E.; Faba, C. y Moya, A. (2001). El destino de las revistas científicas nacionales. El caso español a través de una muestra (1950-90). *Revista Española de Documentación Científica*, 24, 147-161.
- León, O.G. y Montero, I. (1997). *Diseño de investigaciones*. Madrid: McGraw Hill.
- Méndez, F.X. (2001). Prólogo. En G. Buela-Casal y J.C. Sierra, *Manual de evaluación y tratamientos psicológicos* (pp. 21-24). Madrid: Biblioteca Nueva.
- Montero, I. y León, O.G. (2001). Usos y costumbres metodológicas en la Psicología española: un análisis a través de la vida de *Psicothema* (1990-1999). *Psicothema*, 13, 671-677.
- Montero, I. y León, O.G. (2002). Clasificación y descripción de las metodologías de investigación en Psicología. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 2, 503-508.
- Muñiz, J. (1996). *Psicometría*. Madrid: Universitas.
- Muñiz, J. (1997). Aspectos éticos y deontológicos de la evaluación psicológica. En A. Cordero (Dir.), *La Evaluación Psicológica en el año 2000* (pp. 307-345). Madrid: TEA Ediciones.
- Muñiz, J. y Fernández-Hermida, J.R. (2000). La utilización de los tests en España. *Papeles del Psicólogo*, 76, 41-49.
- Muñiz, J. y Hambleton, R.K. (1996). Directrices para la traducción y adaptación de tests. *Papeles del Psicólogo*, 66, 63-70.
- Pelechano, V. (2000). *Psicología sistemática de la personalidad*. Barcelona: Ariel.
- Pelechano, V. (2002a). Presentación del Monográfico 2002: La valoración de la actividad científica. *Análisis y Modificación de Conducta*, 28, 311-316.
- Pelechano, V. (2002b). ¿Valoración de la actividad científica en Psicología? ¿Pseudoproblema, sociologismo o idealismo? *Análisis y Modificación de Conducta*, 28, 323-362.
- Pelechano, V. (2002c). Epílogo del Monográfico 2002: La valoración de la actividad científica. *Análisis y Modificación de Conducta*, 28, 477-479.
- Pérez Álvarez, M. (2002). La Psicología en el contexto de la cultura española. *Análisis y Modificación de Conducta*, 28, 405-430.

- Pérez Álvarez, M. y Fernández-Hermida, J.R. (2001). El grano y la criba de los tratamientos psicológicos. *Psicothema*, 13, 523-529.
- Polaino, A. (2002). Ciencia, política y política de la investigación. *Análisis y Modificación de Conducta*, 28, 363-390.
- Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.
- Sánchez M. (1999). *Análisis comparativo de la producción científica en la Unión Europea, España y el Consejo Superior de Investigaciones Científicas*. Madrid: Consejo Superior de Investigaciones Científicas.
- Seglen, P. O. (1993). How representative is the journal impact factor? *Research Evaluation*, 2, 143-149.
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314, 498-502.
- Shashok, K. (1997). Responsabilidades compartidas en la revisión de los originales por expertos. *Revista de Neurología*, 25, 1.946-1.950.
- Sierra, J.C. y Buela-Casal, G. (2001). Past and present of behavioural assessment. *Revista Internacional de Psicología Clínica y de la Salud /International Journal of Clinical and Health Psychology*, 1 (225-236).
- Siguan, M. (2002). Evaluando manuscritos: la experiencia de un director. *Análisis y Modificación de Conducta*, 28, 317-322.
- Sokal, A. (1996a). Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity. *Social Text*, 46/47, 217-252.
- Sokal, A. (1996b). Transgressing the boundaries: An afterword. *Dissent*, 43, 93-96.
- Sokal, A. y Bricmont, J. (1999). *Imposturas intelectuales*. Barcelona: Paidós (Orig. 1998).
- Sternberg, R.J. (1988). *The psychologist's. A guide to scientific writing for students and researchs*. Nueva York: Cambridge University Press.
- Sternberg, R.J. (1999). A propulsion model of types of creative contributions. *Review of General Psychology*, 3, 83-100.
- Sternberg, R.J. (2001). Where was it published? *Observer*, 14, 3.
- Sternberg, R.J. (2002). On civility in reviewing. *Observer*, 15, 3, 34.
- Sternberg, R.J. (2003). There is no place for hostile review. *Revista Internacional de Psicología Clínica y de la Salud. International Journal of Clinical and Health Psychology*, 3 (en prensa).
- Sternberg, R. y Gordeeva, T. (1996). The anatomy of impact: What makes an article influential? *Psychological Science*, 8, 69-75.
- Van Leeuwen, T. N.; Moed, H.F.; Tussen, R.J.W.; Visser, M.S. y Van Raan, A.F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparison of national research performance. *Scientometrics*, 51, 335-346.
- Werner, R.; Pelicioni, M.C. y Chiatone, H. (2002). La Psicología de la Salud Latinoamericana: hacia la promoción de la salud. *Revista Internacional de Psicología Clínica y de la Salud /International Journal of Clinical and Health Psychology*, 2 (153-172).