

ASSESSING THE QUALITY OF ARTICLES AND SCIENTIFIC JOURNALS: PROPOSAL FOR WEIGHTED IMPACT FACTOR AND A QUALITY INDEX

Gualberto Buela-Casal
University of Granada

The impact factor and other bibliometric indices are currently used in several countries to assess the scientific production and its quality. However, the impact (or prestige) factor is rarely interpreted accurately. Available impact factors present 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 journals where citations appear, so that every citation is given the same value, regardless of the periodical in which it is published. In order to overcome these limitations, two indices are proposed: a mean impact factor of the journals where citations appear (MIFJC) and a weighted impact factor (WIF). Additionally, other useful indices are suggested for the analysis of interaction between periodicals: a percentage of partial interaction of citations (PPIC), and a percentage of mutual interaction among citations (PMIC). This paper explains them in detail and outlines procedures for their calculation. Several problem areas are discussed, namely, peer review, publication policies, the credentials of referees, and assessment criteria. It is also argued that the quality of studies should not be primarily evaluated in terms of the journal in which 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 assessed, who should assess it, and possible criteria for assessment. These considerations lead to a proposal for a quality index independent of impact or prestige.

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.

The impact factor of journals is the most widely used bibliometric index for assessing and comparing the scientific production of different countries. On the basis of this factor, we find, for example, that Spanish scientific production represents 2.7% of world production for the period 1996-2000 (Aguirre de Cárcer, 2001); other studies indicate that Spain occupies sixth place in the

ranking of European Union countries and eleventh in the world ranking (Sánchez, 1999). In the field of psychology in Spain, studies have also been carried out on scientific production, using as criterion the impact factor of the journals in which Spanish psychologists publish (see Alcain & Carpintero, 2001). However, the use of this bibliometric index presents serious disadvantages, due on the one hand to their incorrect interpretation (see Buela-Casal, 2001, 2002), and on the other to "intrinsic" limitations of the index itself, such as the fact that the impact factor is based exclusively on citations from a two-year period (Buela-Casal, 2002), and that its calcu-

The original Spanish version of this paper has been previously published in *Psicothema*, 2003, Vol. 15. No 1, 23-35

.....
Correspondence concerning this article should be addressed to Gualberto Buela-Casal, Facultad de Psychology, Universidad de Granada, 18071 Granada (Spain). E-mail: gbuela@ugr.es

lation does not take into account the “impact” or “prestige” of the journals in which the citations appear.

In the majority of countries the production and quality of scientific research is assessed by means of bibliometric indices. These are used indiscriminately for assessing researchers, research institutes, universities, countries, and so on. Such indices are based on the number of citations received by scientific journals (in a given period and in accordance with the articles they publish); the *impact factor* and the *prestige factor* are examples of these indices.

There would be no problem if bibliometric indices were used as what they actually are, that is, as indicators of the level of circulation among the scientific community. The problem emerges when these factors are used as indices of quality, so that, in turn, it is considered that an article has quality on the basis of the “impact” or “prestige” of the journal in which it was published. This type of inference is now being called more and more into question. For Pelechano (2000) it is to confuse science with the sociology of science, and he argues that what began as a quite specific way of understanding the “impact” of scientific publications (without reading them) ended up constituting a form of judging scientific contribution. Likewise, Sternberg (2001) points out the need to differentiate between what is published and where it is published: not everything published in the same journal is of the same quality. The correlation between the “impact” of an article and the “impact” of the journal in which it was published is far from perfect (Sternberg & Gordeeva, 1996). Buela-Casal (2001) concludes that a journal’s “impact” is no indicator of the level of professionalism, social relevance or implementation of the research published in that journal. As Werner, Pelicioni and Chiattonne (2002) argue, what good is the publication of proposals for intervention programmes in the field of health psychology if they then fail to be implemented?

Furthermore, it should be borne in mind in relation to the assessment of research in applied areas that the result in this case is not always publication, since many researchers, companies or industries prefer to patent the results of their work rather than publish them in scientific journals (Gómez & Bordons, 1996).

Critical reflections on current systems of assessing scientific research, articles and journals are becoming ever more frequent (see Pelechano, 2002a, 2002b, 2002c), and derive from a wide variety of perspectives: from that of journal directors (Siguan, 2002), from that

of reviewers (Bobenrieth, 2002), from that of researchers (Sternberg, 2002, 2003), based on the relationship between science and politics (Polaino, 2002), in relation to Spanish researchers’ system of scientific production (research *sexenios*) (Echeburúa, 2002), via the analysis of criteria for assessing the quality of universities (Chacón, Pérez-Gil, Holgado & Lara, 2001), in relation to basic scientific principles (Bornas & Noguera, 2002), from the perspective of the Spanish cultural context (Carpintero, 2001; Pérez Álvarez, 2002), in relation to the language and place of publication of the journal (Buela-Casal, 2001; Van Leeuwen, Moed, Tijssen, Visser and Van Raan (2001), from reflections on the differences between quantity and quality (Gil-Roales & Luciano, 2002), from critical views on inappropriate use of criteria such as “the majority opinion”, the impact factor and the prestige factor in assessing research quality (Buela-Casal, 2002), from analysis of the limitations of bibliometric indicators (Alexandre & Porcel, 2000; Amin & Mabe, 2000; Bordons, Fernández & Gómez, 2002; Gómez & Bordons, 1996; Seglen, 1993, 1997), based on assessments of the effect on Spanish scientific journals and their future (Díaz et al, 2001; Bordons, Felipe & Gómez, 2002; Jiménez-Contreras, Faba & Moya, 2001), and even from the claim that it is a question of a pseudoproblem, a sociologism or an ideologism (Pelechano, 2002b).

The different bibliometric indices, such as impact factor, prestige factor, equivalent impact factor, and so on, can be manipulated, intentionally or otherwise, by means of an “artificial” increase in the number of citations. Recently, Buela-Casal (2002) described various procedures that can be used for raising the number of citations, and which he refers to as “The Ten Commandments for increasing citations” (see Box 1).

Box 1 The Ten Commandments for increasing citations
<ol style="list-style-type: none">1. Increase the circulation of the journal.2. Include the journal in the greatest possible number of databases.3. Publish controversial articles.4. Publish reviews.5. Publish in English.6. Publish articles on topical issues.7. Publish articles by widely-cited authors.8. Make deals with the communications media.9. Recommend citation of works published in the same journal.10. Facilitate access to articles via Internet.
Source: Buela-Casal (2002)

Another important aspect that is rarely called into question is the system of assessment of scientific articles, known as “peer review”. Here we analyze in detail the limitations of this procedure, as employed in the majority of journals. Likewise, we examine the different reasons why it is necessary to eradicate the unfortunate custom of assessing the quality of articles solely on the basis of where they are published.

Finally, we present an alternative for assessing articles and scientific journals, focusing on three aspects: what criteria can be used, what should be assessed, and who should do the assessing. This is not the first time these aspects have been discussed, but the novelty here lies in the way they are considered and the practical proposals made in relation to them.

THE “IMPACT FACTOR” AND “PRESTIGE FACTOR” CRITERIA

Interest in classifying or “measuring” scientific research is not a recent phenomenon: one of the first classifications was proposed by Gross and Gross (1927). Nevertheless, the criterion of measuring the “impact” of scientific publications was suggested by Garfield (1955), and published in the journal *Science*, and “*impact factor*” was first used for quantifying publications in the 1963 edition of the *Science Citation Index (SCI)*. This index was initially published in a supplement of the SCI, under the name *Journal Citation Reports (JCR)*, and it has now become the most important publication of the *Institute for Scientific Information (ISI)*.

Journal Citation Reports is published annually, and made up of two sections, one called *Sciences* and the other *Social Sciences*. In each of these sections journals are classified in two categories, one alphabetically and another by subject (and within the latter they are ranked from highest to lowest impact factor). Although the principal and most well-known index is that of *impact factor*, others are also included: total citations made over the year, immediate index and mean life of citations (for a description of these indices, see Buéla-Casal, 2002).

The *impact factor* is calculated by dividing the total number of citations that the articles published in a journal in the two previous years receive in one year by the number of articles published in that journal in those two years.

The *prestige factor*. This has been proposed recently by a Canadian company (*Factorprestige*), as an alternative to the ISI’s *impact factor*. The *prestige factor* undoubtedly presents considerable advantages, such as its use of

four databases: biomedical, technological, agro-geo-environmental and social sciences. These databases include over six thousand journals classified in 859 categories (whilst the ISI uses 212 categories). The use of these four databases permits a more rational comparison of journals, as each one is compared with other, related ones in the same field.

Another important novelty of *prestige factor* is that its calculation does not take into account citations from review articles. It has been estimated that reviews receive three times more citations than an original article. Moreover, the majority of reviews are published as a result of the journal director’s invitation to an “important” author, thus somehow breaking the “rules of the game” for selection of articles.

The *prestige factor* is calculated by dividing the total number of citations that the original articles published in a journal in that same year and in the two previous years receive in one year by the number of original articles published in that journal in those three years (in the case of biomedical journals two further variables are taken into account: clinical articles or basic articles). The results are converted by means of an algorithm on a scale of 0 to 1000.

Factorprestige offers other specific bibliometric indices that permit better quantification of publications: equivalent impact factor, review factor, review index, percentile, interaction index, publication index, and so on. Nevertheless, directly or indirectly, all of these indices are based on the number of citations (see Buéla-Casal, 2002, for a description of these indices).

It is important to bear in mind that impact factor and prestige factor, despite attempting to measure the same thing, are not comparable, for various reasons: they use different citation periods (two years the former and three the latter), they use different ranks, the former includes reviews and the latter does not, and so on (for more information, see Buéla-Casal, 2001).

An important limitation of impact factor and prestige factor is that they give the same value to all citations, and regardless of the journal in which the citation appears. A possible solution to this situation involves the use of the weighted impact factor, as will be described in the following section.

PROPOSAL OF WEIGHTED IMPACT FACTOR AND OTHER BIBLIOMETRIC INDICES

As proposed by the *Institute for Scientific Information*,

the impact factor is calculated according to the total number of citations, but it does not take into account the type of journals in which the citations appear. One of its serious limitations is that it gives the same value to all citations, regardless of the fact that they may appear in a journal with low impact factor, such as *Scientist*, which received 348 citations (in the year 2000) and has an impact factor of 0.347, or in a journal such as *Science*, which in the year 2000 received 274,443 citations and has an impact factor of 23,872 (*Institute for Scientific Information*, 2000). Clearly, a citation in this latter journal is more important, or has more “impact” than a citation in *Scientist*. In sum, just as it is not the same to publish in *Science* as in *Scientist*, nor is it the same to be cited in one of these journals as in the other.

The same argument applies to the prestige factor, which although much less well-known, is a considerable improvement on the impact factor (for a comparison of these indices, see Buela-Casal, 2002). Nevertheless, it has the same unfortunate feature of giving equal value to all citations, failing to take into account the status of the journals in which they appear.

One of the objectives of the present article is to propose a modified index of impact factor and prestige factor, which weights the value of citations according to the impact and/or prestige of the journals in which the citations appear. To this end, it is necessary to propose the use of another bibliometric index, such as *mean impact factor of the journals where citations appear* (MIFJC), which will in turn be necessary for calculating the *weighted impact factor* (WIF). Furthermore, we propose another two bibliometric indices that are of great utility for comparing two or more journals with one another in relation to the interaction of common citations. An example of analysis of the interaction between different journals is provided by the recent comparative study of the four psychology journals with impact factor published in Spanish (Buela-Casal, Carretero-Dios & De los Santos-Roig, 2002).

PROPOSAL OF BIBLIOMETRIC INDICES

– *Mean impact factor of the journals where citations appear* (MIFJC): This refers to the average of the weighted impact factor of the journals in which the journal was cited. It is calculated by multiplying the impact factor of each journal (in that year) that cites articles from the two previous years of the analyzed journal by the number of articles cited in each journal, the total sum

being divided by the total number of articles cited.

This bibliometric index is one of the most important, as it indicates the mean impact factor of the journals that cite the articles of given journal. A value of this factor higher than that of the journal’s impact factor indicates that the journal is cited in journals with an impact factor greater than its own. The higher the mean impact factor of the journals, the greater influence or effect that journal has on the scientific community, since the articles are cited in journals that are themselves widely cited. Table 1 shows the journals in which *Psicothema* was cited in the year 2000, whose impact factors range from 0 to more than 3, and it is clearly not the same to be cited in a journal with an impact factor of 0.000 than in one with 3.858, since in the latter case the journal is itself widely cited. Thus, a citation of *Psicothema* in *Exp. Neurol.* will necessarily have greater importance than a citation in a journal without impact factor, since many researchers who have read and cited the article from *Exp. Neurol.* are at least

Table 1
Journals in which *Psicothema* was cited in the year 2000

JOURNAL	IMPACT FACTOR
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
AGGRESSIVE 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
MEAN IMPACT FACTOR	0.592

Note: This is not calculated directly, but based on the number of citations received by *Psicothema* (in articles published in the previous two years) in each one of these journals.

aware of the existence of the work published in *Psicothema*, leading to an increase in the probability of it being more often consulted and more frequently cited. On the other hand, a citation in a journal with 0 impact factor indicates that the articles published in that journal were not cited by other researchers, and this does not favour future citations of the article in question.

A disadvantage that should be borne in mind in this index is that there is a degree of “regression to the mean”, since journals with a very high impact factor will necessarily be cited in journals with similar or lower impact factor; that is, starting from a certain impact factor, the mean impact factor of the journals where citations appear will be lower than their own impact factor, and while this is normal, it should not be misinterpreted.

– *Weighted impact factor (WIF)*: This is obtained by taking the impact factor (IF) of a journal and the mean impact factor of the journals where citations appear (MIFJC) and calculating the average of the two factors.

$$WIF = \frac{MIFJC + IF}{2}$$

The advantage of the weighted impact factor is that it combines information on the citations received by a journal and the impact factor of the journals that cite it. It is clearly not the same for a journal to receive a certain number of citations from journals with low impact factor as it is to receive them from those with high impact factor. This index corrects the “regression to the mean” effect mentioned in relation to the previous index (MIFJC), since, on averaging the mean impact factor of the journals in which the citations appear and the impact factor of the journal itself, this effect is prevented from occurring. The example in the case of *Psicothema* for the year 2000 would be:

$$\frac{0.592 (MIFJC) + 0.473 (IF)}{2} = 0.532 (WIF)$$

– *Mean prestige factor of the journals where citations appear (MPFJC)*. This refers to the average of the weighted prestige factor of the journals in which the journal was cited. It is calculated by multiplying the prestige factor of each journal (in that year) that cites articles from the current year and the two previous years from the analyzed journal by the number of articles cited in each journal; the sum is then divided by the total number of articles cited.

– *Weighted prestige factor (WPF)*. This is calculated from the prestige factor (PF) of a journal and the mean prestige factor of the journals in which the citations appear (MPFJC). It is obtained by calculating the average of the two factors.

$$WPF = \frac{MPFJC + PF}{2}$$

– *Percentage of partial interaction of citations (PPIC)*. This refers to the percentage of articles cited in one journal from another journal (citations between them in one direction), so that their interaction can be analyzed. Partial percentage of interaction is calculated by taking the total of citations from one or more journals in one year and calculating the percentage according to the number of citations received by those journals in the same year. Below we show, by way of example, the PPIC of the journal *Psicothema* with respect to the journal *Personality and Individual Differences*, and vice-versa:

- Citations of *Psicothema* in the year 2000= 144
- Citations of *Pers Indiv Differ* in the year 2000= 2,836
- Citations of *Psicothema* in *Pers Indiv Differ*= 39
- Citations of *Pers Indiv Differ* in *Psicothema*= 1
- PPIC of *Psicothema* with respect to *Pers Indiv Differ*= 1.37%
- PPIC of *Pers Indiv Differ* with respect to *Psicothema*= 0.69%

These percentages of interaction mean that the journal *Psicothema* represents 1.37% of the citations received by *Pers Indiv Differ* in a year, which is a considerable quantity if we bear in mind that the former is a general journal published mostly in Spanish, while the second is a journal specializing in the psychology of personality and individual differences, and published in English. As regards the inverse relationship between these journals, it can be seen that *Pers Indiv Differ* represents 0.69% of the citations received by *Psicothema*.

– *Percentage of mutual interaction of citations (PMIC)*. This refers to the percentage of articles cited among two or more journals (citations between them in both directions), so that their interaction can be analyzed, but self-citations are excluded. Percentage of interaction is calculated by taking the total of common citations among two or more journals in one year and calculating the percentage according to the number of citations received by these journals in that same year. Below we show the percentage of mutual interaction of citations for the two journals referred to in the previous example:

- Total citations of *Psicothema* and *Pers Indiv Differ* (2000)= 2,980
- Total common citations= 40
- PMIC of *Psicothema* and *Pers Indiv Differ*= 1.34%

A PMIC between these two journals of 1.34% indicates that they have more than one percent of citations in common. However, if we analyze the partial interaction, as discussed above, we find that this percentage of mutual interaction is due basically to the influence of *Psicothema* on *Pers Indiv Differ*. Obviously, percentages of partial and mutual interaction must also be interpreted in relation to the thematic similarity between the journals compared.

THE “PEER REVIEW” SYSTEM FOR ASSESSING QUALITY: ADVANTAGES AND DISADVANTAGES

Scientific journals use an assessment system known as “peer review”. In this system, manuscripts sent to a journal are remitted to other researchers supposedly specializing in the subject of the article received. These reviewers (whose number ranges from two to six) usually receive an anonymous manuscript. Once they have made their review, they send it to the director of the journal, who in accordance with the opinions of the referees and his or her own opinion makes the final decision on whether to accept it for publication or reject it. The reviewers are also anonymous for the researchers who submit the work. Such a system would appear to have many advantages: an “impartial” review given the anonymity of the authors, a review carried out by specialists in the field, thematic coherence, since the director and referees also decide on the suitability of the text for that particular journal, and so on. Nevertheless, detailed analysis of the process leads us to the conclusion that it also has some disadvantages, some of which we discuss below.

Directors’ biases

Although each journal has a defined thematic field, the director always has some degree of freedom for favouring the publication of manuscripts on certain themes or areas, and thus for hindering that of others. Such bias is achieved by various means:

- a) One of these is the selection of the review committee: in some cases directors select its members directly, and in others they at least have considerable influence over the committee’s make-up. It is

usually considered that European and Ibero-American journals are more affected by this bias, since these have more tradition of directors maintaining their position for life or for many years, in contrast to the case of American journals, where directors change periodically. However, such differences only really affect whether the bias is more variable or more constant, since changing the director does not eliminate the bias.

- b) The director decides to which reviewers to send the work, in the knowledge that not all of them are equally strict, so that this decision has a clear influence on the manuscript’s possibilities of publication.
- c) The final decision on publication of an article is taken by directors, who may have received different or even contradictory reviews of it, so that it falls to them to choose whether to send it to other referees or opt for some of those already received.
- d) Some journals frequently have more studies with favourable reviews than they can publish. Directors decide which of these suitable works to publish, and will undoubtedly have preference for certain topics over others.
- e) The “citation tornado effect”, which refers to the fact that widely-cited authors have more likelihood of being published, since their articles will increase the degree to which the journal is cited. It should be borne in mind that the director does know the identity of the author, and this will undoubtedly influence his or her decision. For example, between a work of suitable quality by a well known author and an equally suitable one by an unknown author, directors will surely opt for the former.
- f) The tendency to publish studies that find effects or correlations and to reject those that, while methodologically correct, do not obtain positive results. This was a customary practice in the past, and although it is less frequent today, it nevertheless still occurs (see the example in Box 2), and is the source of considerable bias in relation to articles published.

The above comments are not intended as destructive criticism of the work of directors; rather, they are aimed at drawing attention to the fact that a director influences to some degree or other the final decision on the publication of an article. The director’s biases are implicit in the system, though this does not mean they invalidate the system.

The credentials and objectivity of reviewers

Reviewers are not as qualified, independent and objective as it might be believed, as shown by the following:

- a) The selection of reviewers is by no means perfect. In some cases they are named directly by the director, and although the criterion of using specialists is adhered to, others also come into play, such as the reviewer's prestige, friendship with the director, and so on. In other cases, such as that of the APA journals, advertisements periodically appear requesting applications from candidates who fulfil the following conditions: having previously published in journals with review systems, being an habitual reader of five or six journals in a field, being a specialist in an area and having sufficient time to work on reviews. In this latter case, it is clearly not the best possible reviewers who are selected, but rather those who apply.
- b) Reviewers are not better qualified than the authors. Indeed, in some cases the authors are better known,

as they have published more work than the reviewers, so that we can at least question the reviewer's authority for judging the work of the author. A good example is the case of two authors who in the appendix to their article criticize and advise the reviewers of the same journal with a view to improving the review system (see Montero & León, 2001).

- c) Reviewers are not better when they review than when they carry out research. If reviewers also perform studies, which are subsequently assessed by other "peers" and may be rejected, a contradiction arises: they are considered qualified to assess, but at the same time their work can be turned down.
- d) Reviewers learn "by experience". They have had no previous instruction or training in how to review an article, so that they review manuscripts on the basis of their opinion and experience. When reviewers assess their first article, with what criteria do they do so? When and where did they learn?

Box 2

Example of the tendency to publish studies that find effects and to reject those that do not find them

The author of the present article, when he was a doctoral student, submitted a study for publication to a journal included in *Journal Citation Reports*. The director of the journal returned the work, saying that although it was methodologically correct, he could not accept it for publication because in truth it did not demonstrate any effect of the independent variable on the dependent variables. The author was surprised, as the sample size of his study was ten times larger than that of similar studies, some published in the same journal, which did indeed find effects. The conclusion was clear: what was important was not to use appropriate methodology or sample size, but to "find effects". Consequently, the young researcher decided to file the study away and not send it to any other journals: it had already taken quite a lot of daring to submit it to a "prestige" journal.

Several years passed, and the author was now a university lecturer. During a move from one university to another, he found at the bottom of a drawer in his office the article, which he had totally forgotten about. Once he was settled at his new university, he decided to update the theoretical review, but did not change the rest of the manuscript at all. He submitted the study to a journal included in *Journal Citation Reports*, but with an impact factor **four times greater** than that of the journal which has rejected it several years earlier. The work was accepted without any type of criticism and published. One might be inclined to think that the wood in the drawer had had the same effect as the oak on wine when it matures in casks. But neither was his desk made of oak, nor did this explanation seem to be the right one. There were other possible explanations: lack of reliability between reviewers, intentionality on the part of reviewers specializing in the field who had indeed found effects, director's biases, and so on. However, the author attributed the work's acceptance to a change in tendency that had occurred in journals, which were now more prepared to accept also those studies in which no effects were found. This does not imply that his attribution was correct, but, what did it matter? – the main thing was to get the work published. Or at least that's what the author thought at the time.

Box 3

Example of lack of reliability between reviewers

Below we present the comments of the reviewers for two journals included in *Journal Citation Reports* and **with identical impact factor** on the same article that was submitted for review. As it can be seen, while the Journal "A" reviewers discredit and reject the article in no uncertain terms, the Journal "B" reviewers accept it and, moreover, make quite positive comments (this is a real case provided by a researcher who prefers to remain anonymous).

Journal A: The article was rejected following the reviewers' recommendations:

Reviewer 1: "*I find the manuscript of poor quality, riddled with problems and unsuitable for publication in this journal*".

Reviewer 2: "*The article has considerable methodological and conceptual limitations, which make it difficult to consider for publication*".

Reviewer 3: "*It makes no sense to make further comment on an article of such a poor standard. I can only ask myself why such an article is sent for review, rather than simply being rejected by the Editorial Committee*".

Journal B: The article was accepted subject to minor modifications to formal aspects. The reviewers' comments included the following:

- "*The article is clearly written, and both the text and the figures present the information in a direct, well-summarized and didactic manner*".
- "*The authors have presented the results in a very clear fashion*".
- "*The authors have made an excellent summary of the results*".

This is a clear example of lack of reliability between reviewers, since those of Journal A reject the work not because it is unsuitable for that type of journal, but because of the reviewers' opinion on its quality. This example also serves to indicate that the impact factor does not correspond to quality level, as, of two journals with the same impact factor, one rejects the manuscript and the other publishes it.

Note: Clearly, the comments of the Journal A reviewers could also be used as examples of hostile or cruel reviews.

- e) Lack of reliability between reviewers. It is far from exceptional to find partial or total disagreement between different reviewers with regard to the same article. This is partly explained by the previous point, but what is most important is that the “peer review” system is not reliable (see Box 3 for a real example of a case of lack of reliability between reviewers).
- f) Reviewer bias. Apart from the particular biases of each reviewer, the fact that they are specialists in the topic and conversant with the theories in a given field implies a certain bias towards accepting works in line with the current situation and rejecting innovative studies. This represents a restriction on the most creative researchers.

- g) The anonymity of reviewers gives rise to deliberate, exaggerated or hostile criticisms. If reviewers are specialists in a field they will have published research in it. If a work submitted to critical review or its results are in total or partial contradiction to their work, reviewers will most likely tend to reject it, and this is made easier by the anonymity. Box 4 shows an example of a real case of a merciless criticism taking advantage of the cloak of anonymity. The anonymous reviewer even went as far as to write that the manuscript appeared to be written by “a charlatan attorney”, and considered it worthy of no more than a first-year university student. The anonymous author was none other than Robert Sternberg, current President of the American Psychological Association.

Box 4
Example of “cruel” reviews

ON CIVILITY IN REVIEWING
Robert J. Sternberg

Many of us have put in our best-faith efforts in writing journal articles or grant proposals, only to receive savage reviews. I recently received a savage review of an article I co-wrote and submitted to a journal that referred to the submitted article as sounding like it was written by a “charlatan attorney” and that referred to parts of the article as “absurd” and as “gibberish.” It compared the argumentation to that in “freshman-level term papers,” and recommended that the author, who is “seriously out of his/her element with this topic...refrain from venturing into areas that exceed his/her professional competence.”

Other comments in the review were not dissimilar to these comments. Fortunately, the savage review was “confidential,” although of course, this confidential review was seen by the editor, other reviewers, and who knows who else. There was a second review of the article as well, which was very favorable.

Whether reviewers in other fields do the same, I do not know, although I have heard that when it comes to reviewing, reviewers in no field attack their own in quite the savage way psychologists sometimes do. What I do know is that more than once in my career, I have been stunned by the savagery with which my own work, and the work of others, has been attacked, usually under the cloak of anonymity. I have been surprised both that the reviewers wrote what they wrote and that the editors (or heads of granting panels) sent back the reviews in such form. Such reviews are, I believe, unacceptable, regardless of the quality (or lack thereof!) of the manuscripts being reviewed.

Professors spend some amount of time teaching students in graduate school how to write articles, but they spend little or no time teaching them how to review articles (or books or grant proposals). Perhaps they like to think that reviewing is a skill that just comes naturally to people. Apparently, it doesn't. Perhaps explicit training is needed.

Savage reviews are harmful to all concerned, and for several reasons:

Fundamental Ethics. As a matter of fundamental ethics, savage reviews violate the Golden Rule-to act toward others as we would have them act toward us. No one wants to be treated in a disrespectful and even insulting way, no matter who the individual is.

Balance. Really, it is the rare article or grant proposal that has no fundamental positive value at all. The job of a reviewer is to provide a balanced evaluation of an article or grant proposal, pointing out the positive value of as well as the flaws in whatever is reviewed. Savage reviews are unbalanced. Thus, savage reviewers are not really doing their job. They are providing unbalanced one-sided accounts of an article or proposal rather than a balanced one. Even if articles need to be criticized rather severely, such criticism can always be communicated in a tactful way.

Stretching the truth. It is rare that more than one of a set of reviews is savage. Not infrequently, other reviews even are positive. The papers and proposals are rarely as bad as these reviewers make them out to be. In savaging a paper, the reviewers usually are stretching the truth.

Maintaining Credibility in Order to Effect Change. In general, you do not get people to change their behavior by demeaning or insulting them. Rather, you lose credibility in the eyes of precisely the people whose behavior you want to change. As a result, these people may be disinclined to listen to what you say. The personal content of the review reduces rather than increases its effectiveness.

Undermining self-efficacy. Senior investigators usually, although not always, take occasional savage reviews in stride. They have gotten such reviews before; they know they most likely will get them again. Besides, they often have tenure, so do not have to worry that their employment is in jeopardy. But junior investigators just starting out are often taken aback. They may become quite discouraged and even give up hope of ever achieving success in the field. They do not yet realize that the problem is in the reviewer, not in the material he or she reviewed. The material may in fact be in need of considerable work, but even if it is, there is no need savagely to attack it. (Sternberg, 2003).

Reproduced with the permission of R. Sternberg.

Association (APA), who decided to make these harsh criticisms public. They are summarized in Box 4 (Sternberg, 2003).

The established parameters and criteria on the aspects to be assessed

Journals use review forms that are sent to referees together with the manuscript to be reviewed. The purpose of these forms is to establish the parameters and/or criteria to be followed in assessing the work. However, if we examine these criteria, it is clear that they do not totally ensure quality. For example, some of the most frequent are: relevance of the topic dealt with, methodological rigour, clarity of exposition, contributions of the study, correct use of language, appropriateness of the bibliography, and so on. But, with rare exceptions, there is no assessment of such important aspects as internal and external validity, utility, implementation, originality or innovation.

It is true that some of the parameters assessed in the review process, such as “relevance of the topic”, “methodological rigour” or “contributions of the study”, are necessarily related to the quality of the work. Even so, the problem is that they are assessed in a quite general way, and it is left to the reviewers’ discretion to apply these parameters. Thus, the point is not that they fail to assess quality; the problem resides in the way the assessment is made, which is far too general, thus bringing reviewers’ subjectivity into play.

THE QUALITY OF AN ARTICLE SHOULD NOT BE ASSESSED ONLY IN RELATION TO THE IMPACT OR PRESTIGE OF THE JOURNAL IN WHICH IT IS PUBLISHED

Sternberg (2001) and Buéla-Casal (2002) propose fifteen reasons why it is a mistake to give more importance to “where” an article is published than to the article itself:

1. It is easier to quantify citations or to make an assessment based on the publications cited in a work than it is to read the article, but the impact of the journal is not a substitute for critical evaluation of the work.
2. The conservatism of the most prestigious journals. Normally, the most prestigious journals are more conservative, so that reviewers tend to check more strictly that the work is in line with the most conventional norms (see Box 5).

3. Difficulties for the publication of interdisciplinary research. Bearing in mind that the most prestigious journals tend to be established within traditionally defined fields, it is difficult to find high-prestige journals that are interdisciplinary, so that studies of this type are usually “penalized”, since it is also difficult for them to gain acceptance by generalist journals. A good example is the case of psychoneuroimmunology, and especially the early work by Ader and Cohen, who found it extremely difficult to publish their experimental research.
4. Difficulties for non-paradigmatic research. Studies that do not fall into the conventional research paradigms have less likelihood of being published, since reviewers tend to be conventional in their approach to assessing scientific work, so that researchers tend to work within the conventional paradigms, and those who fail to do so generally find it difficult to publish their reports. In this case we can turn again to the example of Ader and Cohen’s research: once they had managed to publish their initial reports they found no difficulties in publishing subsequent work in prestigious journals (Science, The Lancet and Annual Review of Psychology, for example).
5. Disadvantages of publishing in books and other types of publication other than journals. For example, in Spain (and other European countries), for the assessment and advancement of university researchers and teachers only work published in journals is taken into account, and particularly those included in *Journal Citation Reports*.

Box 5

Examples of the conservatism of the most prestigious journals

“The most prestigious journals are often, in my experience, more conservative. I say this after having been director of two journals, associate director of another two and a reviewer for many journals ranging from low to very high prestige. Somehow, the most prestigious journals tend to be somewhat more conservative. The more rigorous the review, the more the reviewer is concerned for the article to fit with what is conventional. But the conventional norms can be helpful or unhelpful, depending on the particular case. In fact, when I was President of the *Society for General Psychology*, I founded a journal, the *Journal of General Psychology*, given my belief that the leading journals may be more conservative than was truly desirable” (Sternberg, 2001, p. 3).

“The most remarkable case is perhaps that of Nobel Laureate Rosalyn Yalow, whose seminal article on radioimmunology was rejected by two prestigious journals, one of which justified itself with the following curious and ambiguous argumentation: Truly imaginative and creative people cannot be judged by their peers, because they have none” (Di Trocchio, 1993, p. 137).

6. The self-fulfilling prophecy. This refers to the fact that articles published in prestigious journals tend to be more widely cited than articles published in journals of lower prestige, so that there is a tendency for the prestige of the former to be increased or maintained.
7. The “Matthew Effect”. “For unto every one that hath shall be given, and he shall have abundance; but from him that has not shall be taken away even that which he hath”. This is applicable to the fact that journals with high prestige tend to receive more and better articles than low-prestige journals.
8. Not all the articles published in a journal have the same “impact”. An important criticism to be made of the different bibliometric indices is the attribution of the same “impact” to all the articles published in the same journal, given that the impact and prestige factors are calculated in a general way for the journal. Furthermore, it is clear that some articles receive more citations than others, and the system is so unfair that the articles which receive few citations penalize the widely-cited ones. A good example is provided by those journals that publish annual statistics for a disorder or disease: the article in question tends to increase indirectly the impact factor of the other articles published in the same year.
9. Authors’ choice of the journal to which their work is sent influences the impact it will have. Let us imagine that there is an article of excellent quality, wholly acceptable for publication in a high-prestige journal, but that the authors decide to send it to a journal with medium prestige (this may occur for a variety of reasons: urgency of finding a publisher, lack of knowledge of the system, the journal’s field is more appropriate, etc.), and let us suppose that it is published. So, does the fact of its being published in that journal reduce its quality?
10. The “peer review” system does not guarantee quality. As mentioned earlier, this assessment procedure is influenced by a series of biases on the part of the director and the reviewers that undoubtedly have a serious effect on the assessment, so that there is a far from perfect correspondence with quality.
11. Not all the articles rejected by a journal are of poor quality. There are cases in which journals receive

many applications, so that there may be more acceptable articles available than can be published. The director is thus obliged to reject good work, which ends up being published in other journals that may have lower impact. But the loss of impact is a consequence of the quantity of work, and not of its quality.

12. Articles published in journals with “impact” do not even have a guarantee of truth. In the history of scientific publication there have been a not inconsiderable number of false (or at least partly manipulated) reports. And this affects journals of both high and low prestige.
13. The number of citations can be manipulated in various ways. There is a host of strategies through which the number of citations of a journal can be increased (which is the same as increasing the impact or prestige factor), independently of the quality of the articles published in the journal.
14. Whether or not a journal has an impact factor and prestige factor depends not only on its receiving citations; it is also necessary for the journal’s director and the institution backing it to apply for its inclusion in these systems of citation statistics; indeed, there are many journals that have never carried out this application process, but this obviously has no relation to quality.
15. The language in which a journal is published influences the impact factor and prestige factor, since the language affects the number of citations an article receives. Currently, the majority of researchers read and publish preferentially in English, so that journals published in English will be more widely cited than those published in other languages; and clearly nobody would argue that the language of publication influences the quality of the research.

AN ALTERNATIVE FOR ASSESSING THE QUALITY OF ARTICLES AND JOURNALS: A QUALITY INDEX

First of all it is necessary to establish what we understand by quality in relation to articles and scientific journals. It would seem obvious that it should correspond to more than simply their being widely cited. In principle, we should consider the criteria to be followed in assessment, the requisite characteristics of assessors, and what is to be assessed (articles and journals).

Assessment criteria

There is a substantial repertoire of guidelines for the assessment of scientific articles, which can be found in various books on research and style. However, few of them focus on the important aspects in relation to quality, the majority dealing only with formal aspects, without questioning the general system of assessment.

Below we present a series of parameters, some of which were already proposed by Sternberg (1988) and by Buela-Casal (2002), for assessing scientific articles in psychology and other health sciences, even though the majority of these criteria are also applicable to other fields of knowledge. An article or journal that fulfils the following criteria (or at least the majority of them) can be considered as a quality article or journal, though it should obviously also meet other criteria related to formal and stylistic aspects.

1. Contributes surprising results that make sense in some theoretical context.
2. Contributes results of great theoretical or practical importance.
3. The ideas discussed are novel and interesting, and can give rise to a new approach to an old problem.
4. The interpretation made of the results is unequivocal.
5. Creates a new and simpler framework for results that were previously conceived within a more complex and convoluted framework.
6. Discredits previous ideas that appeared unquestionable.
7. Presents research involving an especially ingenious or novel paradigm.
8. The study has sufficient internal validity, thanks to appropriate design and methodology.
9. The study has sufficient external validity, given that the results and/or theory presented are generalizable.
10. The report provides an adequate description of the method and procedure so that other researchers can replicate them.
11. Theoretical or practical results have a high degree of implementation.
12. The study presents theoretical or practical results that are useful to society.
13. The type of study is clearly specified (León & Montero, 1997; Montero & León, 2001, 2002).
14. Hypotheses are adequately formulated and interpreted (Borges, San Luis, Sánchez & Cañadas, 2001).
15. Clinical studies use treatment techniques and programmes of proven effectiveness (Bados, García & Fuste, 2002; Echeburúa & Corral, 2001; Fernández-Hermida & Pérez Álvarez, 2001; Hamilton & Dobson, 2001; Méndez, 2001; Pérez Álvarez & Fernández-Hermida, 2001).
16. The assessment instruments used are high-quality and have scientific guarantees (Muñiz & Fernández-Hermida, 2000; Prieto & Muñiz, 2000).
17. Case studies adhere to the appropriate norms for presentation of the case (Buela-Casal & Sierra, 2002).
18. Instrumental studies on the construction of instruments follow the appropriate methodology (Muñiz, 1996).
19. Instrumental studies on the adaptation of instruments follow the appropriate methodology (Muñiz & Hambleton, 1996).
20. Deontological norms and principles are adhered to (Almeida & Buela-Casal, 1997; Colegio Oficial de Psicólogos, 1987; Muñiz, 1997).

These are some of the possible parameters, others of which may also be included. What is most important is that they are applied according to the type of work in question. Also of paramount importance is that they are applied with the same criteria, which should be provided to reviewers, as explained below (see also Bobenrieth, 2002).

What to assess: articles and journals

The assessment of scientific publications should be carried out at two levels (articles and journals), following the criteria mentioned above. At a first level, as it is generally carried out now, though it would be advisable to exercise more rigour in the selection of committee members, and perhaps to reconsider the question of reviewer anonymity. At a second level, journals could be assessed by a mixed committee of expert professionals. Each journal would be assessed annually, and would receive a score for its quality. The assessment would be made independently of the language in which the journal is published, the institution backing it, the society it represents, and any other factor that may confer pseudoprestige on it. This index would undoubtedly correspond much more closely with quality than the mere impact factor or prestige factor, which in reality are nothing more than indices of the level of circulation.

Who should assess

Journal committees should be more carefully selected, as it would be advantageous to employ not only specialists in the field, but also experts in research methodology and design. Each reviewer should receive a manual with the assessment parameters and criteria and how to apply them (see, for example, Bobenrieth, 2002). This would undoubtedly increase the validity of assessment and the reliability between assessors. Institutions or companies devoted to the assessment of scientific publications, such as the *Institute for Scientific Information*, *Factorprestige* or *Latindex*, would be required to use committees of expert professionals to make annual assessments of the quality of journals. These committees would be made up of experts in the fields in which the journals are classified, experts in epistemology, and experts in research methodology and design. Courses could even be set up for the training of specialists in assessment of the quality of scientific publications. Assessment by these committees must be independent of the directors of the journals and their boards, and have the authority to detect and assess bias in directors and reviewers. This review by committees of experts would also make it easier to discover plagiarism and false reports, since the same experts would review all the journals in a particular field; such abuse could clearly not be totally eradicated, but would certainly be reduced. It would also be necessary to reconsider the issue of reviewer anonymity; although it brings certain advantages, it must be recognized that it also has important drawbacks, such as the fact that some reviewers take advantage of it to deal out harsh or exaggerated criticism. Such unnecessarily hostile criticism, as Sternberg (2002) argues, generates feelings of helplessness, especially in younger researchers, and makes no positive contribution to the process of assessment of scientific publications. Diverse studies have shown that reviews tend to be more specific and more constructive when reviewers put their signature to them. Anonymity is a "recipe" for lack of responsibility in critical reviews (Shashok, 1997). It would seem, then, that reviewer anonymity brings more disadvantages than advantages, so that in future it may be advisable to identify those who assess.

How to construct a quality index for journals

The quality index would be made possible through the involvement of professional reviewers with proper training, through the use of criteria that truly assessed quality (such as those proposed above), and through adhe-

rence to a manual that set down the way these criteria should be applied. Each member of the expert committee would assess each one of the articles published by a journal in the year, assigning scores according to the established criteria. The final score for each journal would be obtained by calculating the average of scores awarded by the reviewers, and the result would be represented on a single scale (for example, from 0 to 100). Such a system would permit us to obtain:

- a) Quality index for each article.
- b) Annual quality indices for each journal.
- c) Analysis of intra-annual variations of quality indices in accordance with the different editions in a single year.
- d) Analysis of variations between years of quality indices in accordance with the different volumes published by the journal in different years. This would permit the calculation of growth or downturn of journals in terms of quality.
- e) Specific indices for each of the criteria established in the review, such as: internal/external validity index, innovation index, implementation index, utility index, and so on.
- f) Comparison of quality between journals in the same thematic field and between journals from different fields, since the same scale would be used.
- g) Classification of journals according to categories related to the indices. This would help to guide researchers about which journals are most suitable for their work: those that favour utility and/or implementation, those that put a high priority on internal and/or external validity, those that encourage innovative research, and so on.
- h) Analysis of the relationship between quality indices and indices based on citations (impact factor, prestige factor, etc.).

Furthermore, the different indices of quality would not be affected by number of citations and by the factors that affect the *impact factor* and the *prestige factor*, and therefore quality would not be confused with circulation.

DISCUSSION

In the present work we have discussed one of the limitations of impact factor and prestige factor in bibliometric studies. This limitation refers to the fact of assigning the same value to a citation regardless of the "impact" or "prestige" of the journal in which it appears. A solution to this problem involves the application of weighted fac-

tors, which, in addition to number of citations, also take into account the type of journal in which the articles are cited.

In order to make comparisons between journals we propose the use of percentages of interaction of citations. These bibliometric indices allow analysis of the influence of journals on one another, according to the percentage of citations in common.

The percentage of interaction proposed here should not be confused with the interaction index proposed by *Factorprestige*, since the latter refers to the more “interactive” publications, such as letters to the editor and the like. Percentage of partial or mutual interactions is of great use for assessing the influence of journals on one another.

It should be borne in mind that in order to compare the impact or prestige factor between two or more journals, not only should they belong to the same thematic field, but we must also be sure that the journals analyzed in the same year are journals “that cite” and “that are cited”. That is, some journals may be cited, but for different reasons may not be published that year (having gone out of business, temporarily suspended publication or changed their name, for example). The impact or prestige factors in these cases would not be comparable, since a journal that “is cited” but that “does not cite” normally has a lower impact or prestige factor, given that there are no self-citations (which usually constitute a considerable percentage of the citations in a journal). However, in interaction indices based on percentages of common citations it is not necessary to take this into account, since the analyses are always carried out for a year in which the journals in question were published.

It should be emphasized that the bibliometric indices discussed in this work, as indeed any such index based on number of citations, are to be interpreted as quantitative parameters of citations, which serve to quantify the production and circulation of scientific publications. And although these indices may be related in some way to the quality of the articles and the journals, they should not be used as substitutes for other parameters more directly associated with quality (Bobenrieth, 2002; Buela-Casal, 2002; Buela-Casal & Sierra, 2002; Montero & León, 2001, 2002).

The systems used for calculating impact factor or prestige factor are based on the citations and articles published in the last two or three years (respectively). Reflecting on this, one might ask why this criterion is used: it would cer-

tainly seem more reasonable to use a much longer period of time, since, if an article is “key” in a field of knowledge it will be cited over many years. A period of two years is not equally appropriate for the different scientific areas, since the “ageing” of publications is not the same in each field. Thus, while in molecular biology articles cited tend to be very recent, in other areas, such as anatomy or botany, works cited are generally at least ten years old (Gómez & Bordons, 1996).

Another important aspect to bear in mind is that citation habits differ among disciplines. Thus, while in the social sciences articles cite an average of thirty works, in engineering the figure is ten, and in mathematics the average falls to just five. This obviously influences the impact factor. Likewise, the numbers of researchers and journals in each area greatly affect the impact factor (Aleixandre & Porcel, 2000).

Furthermore, there is clearly no direct relationship between number of citations and quality. That is, the fact that an article is cited more or cited less depends on many factors, quality being, at most, just one of these. Recently, Buela-Casal (2002) described diverse procedures that can be used for increasing the citations of journals, none of which involved the publication of articles of more quality. If the citations do not correspond to quality, it follows logically that nor do the different bibliometric indices (impact factor, equivalent impact factor, prestige factor, etc.) correspond to quality. For example, some citations relate to negative criticism of a work, but paradoxically are counted as indicators of quality. There are all sorts of reasons why an author cites an article, and many of them have nothing to do with the quality of the cited work (Aleixandre & Porcel, 2000). Indeed, Garfield (1970, 1996) himself, who introduced the concept of impact factor, warned that there is no absolute relationship between impact factor and quality. More and more authors are criticizing the relationship between quality and bibliometric indices. Sternberg and Gordeeva (1996) remark that there is no perfect relationship between the impact of an article and the impact of the journal. Seglen (1997) argues that the citation rate of an article determines the impact factor of the journal, and not vice-versa, while Pelechano (2000) stresses the fact that the impact index is influenced by the journal’s financial backing and by fashions that dictate the type of thematic areas published. Garfield (1996) in fact acknowledges that the impact factor is calculated without including hundreds of journals normally published in

developing countries, and which therefore have fewer financial and economic resources. Thus, it is not appropriate to make comparisons in relation to scientific productivity and impact factor between countries lacking equivalence in terms of language and journals included in JCR. For comparisons of this type it is advisable to use the “activity index” and the “relative impact factor” (Bordons, Fernández & Gómez, 2002). Buela-Casal (2001) argues that, among other factors, the language of publication affects impact, and quality is obviously not affected by the language used. Other crucial factors are formal aspects of the journals (such as regularity and punctuality of issue) that have nothing to do with the quality of the work; or others even more irrelevant to quality, such as the failure of those running the journal to fulfil the procedures for its inclusion in the different bibliographic indices. Likewise, Sternberg (2001) concludes that place of publication is not a valid indicator of quality and impact, so that only in exceptional cases can more importance be given to the journal than to the actual content of the article published in it.

Another significant problem, habitually overlooked, concerns the limitations of the review process based on assessment “by peers”. Detailed analysis has cast serious doubts, given the biases on the part of both reviewers and editors, as to the validity and reliability of this system. Indeed, Sokal (1996a, 1996b) demonstrated, by achieving the publication of a totally nonsensical article, that the reviewer filter does not always work. In recent years, more and more authors have questioned both the quality and the veracity of publications as a direct equivalent of the prestige of the scientific journals in which they were published (Benach de Rovira & Tapia, 1995; Bobenrieth, 2000; Buela-Casal, 2001, 2002; Colom, 2000; Di Trocchio, 1993; Pelechano, 2000; Seglen, 1993, 1997; Sokal & Bricmont, 1999; Sternberg, 1999, 2002, 2003).

The above comments should not be understood as unmitigated criticism, implying that this system of assessment has no value whatsoever. Criticism should be kept in perspective: clearly, if this type of review were of no use at all, we would be forced to conclude that the research reports published to date had few guarantees – a view nobody would defend. What is most important is to consider that the review system is not perfect, and that within the possibilities currently available it can be improved, as proposed in this work. However, we must remember that the work of both journal directors and referees is based, in the majority of

cases, on altruism: both devote time and effort to tasks with no financial reward, which can basically be classed as altruistic. Therefore, directors can hardly demand more dedication and credentials of reviewers, even if, as we insist, the assessment of articles could be improved. In the meantime, we must take into account that the system functions with significant shortcomings, which have been highlighted in several studies (Shashok, 1997). For example, Howard and Wilkinson (1998) showed that referees and directors are more likely to coincide on the articles they reject than on those they consider suitable for publication. Thus, the role of directors is crucial, since it is they who make the final decision, and it is far from clear how they set the criteria of suitability and quality of an article (Crammer, 1998).

In the present work we propose an index of quality as an alternative to the assessment process for scientific publication, for both articles and journals. The quality index has nothing to do with citations, but is related rather to the “scientific validity” of the work.

There is no reason to stop using the *impact factor*, *prestige factor* and other bibliometric indices, as long as they are considered for what they really are: indices based on the number of citations (see, for example, Sierra & Buela-Casal, 2001). These indices, although apparently more objective than an assessment of quality, do not actually assess quality. The basic problem concerns the use of an apparently logical argument with no foundation: reviewers and journal directors try to assess quality, but once the article has been published this is forgotten, and quality is equated with number of citations, a criterion with no demonstrated justification. Thus, it is necessary to create procedures for assessing the quality of journals.

The proposal discussed in this article in relation to the creation of a possible quality index, though far from being the definitive solution, would at least serve to initiate an improvement in the process of assessing the quality of articles and scientific journals, which is becoming more and more discredited. Clearly, such a system would be much more expensive, since it would involve specialist reviewers and professionals, but it would be a way of truly assessing quality, and though it would never be perfect, it is undoubtedly a better option than the indices that restrict themselves to counting citations. If the parameters and application manuals proposed for assessing journals were also used by the reviewers of articles, there would certainly be an increase in the validity and reliability of reviews.

Finally, a very important aspect to consider in relation to the future of journals in Spanish is that they should not be assessed by means of the impact factor, at least as is currently the case in the Institute for Scientific Information. This system for assessing journals is quite biased with respect to the language of publication. An illustrative example of this bias is the fact that it is obligatory for titles, keywords and abstracts to be in English, whatever the journal's language of publication; this is one of the requirements for inclusion in JCR. Another example is the bias involved when referees reviewing an article written in English criticize the fact that some of the references are in another language (and this is not an isolated phenomenon).

The pressure to publish in English is placing the use of Spanish in a worrying situation. On the one hand, more and more researchers in Spain and Latin America are publishing their work in English-language journals, given that they have greater impact factors; on the other, every year more Spanish journals are published in English. Currently, of the Spanish journals included in JCR, 45% are published only in English, with the rest appearing in both languages or only in Spanish, and this out of reasons of pure survival, since those published only in Spanish are finding it difficult to continue (Díaz et al, 2001; Bordons, Felipe & Gómez, 2002; Jiménez-Contreras, Faba & Moya, 2001). Some serious reflection on the current situation is clearly essential; moreover, there is an urgent need for both institutional support for scientific journals published in Spanish and the creation of a system of quality assessment for journals, such as the quality index proposed in this article, which it is hoped would help to change the attitude of Spanish researchers so that they also sent their best work to journals published in their own language.

ACKNOWLEDGEMENTS

The author would like to thank Robert Sternberg for his comments, ideas and material in connection with this work, and José Ramón Fernández Hermida and José Muñiz for their advice, which helped to improve the final version of the article.

REFERENCES

Alcain, M.D. & 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. & 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. & Buela-Casal, G. (1997). Deontología y evaluación psicológica. In G. Buela-Casal, G. & Sierra, J.C. (Dirs.), *Manual de Evaluación Psicológica. Fundamentos, Técnicas y Aplicaciones* (pp. 239-247). Madrid: Siglo XXI.
- Amin, M. & Mabe, M. (2000). Impact factor: use and abuse. *Perspectives in Publishing*, 1, 1-6.
- Bados, A.; García, E. & 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. & 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. & 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. & Cañadas, I. (2001). El juicio contra la hipótesis nula. Muchos testigos y una sentencia virtuosa. *Psicothema*, 13, 173-178.
- Bornas, X. & 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. & 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. & 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. & 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 gliscienziati imbrogliono*. Milán: Arnoldo Moldadori Editore.
- Díaz, M.; Asensio, B.; Llorente, G.; Moreno, E.; Montori, A.; Palomares, F.; Palomo, J.; Pulido, F.; Senar, J.C. & 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. & 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. & 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. & 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. & Bordons, M. (1996). Limitaciones en el uso de los indicadores bibliométricos para la evaluación científica. *Política Científica*, 46, 21-26.
- Gross, P.L. & Gross, E.M. (1927). College libraries and chemical education. *Science*, 66, 385-389.
- Hamilton, K.E. & 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. & 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. & 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. & Montero, I. (1997). *Diseño de investigaciones*. Madrid: McGraw Hill.
- Méndez, F.X. (2001). Prólogo. In G. Buela-Casal & J.C. Sierra, *Manual de evaluación y tratamientos psicológicos* (pp. 21-24). Madrid: Biblioteca Nueva.
- Montero, I. & 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. & 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. In A. Cordero (Dir.), *La Evaluación Psicológica en el año 2000* (pp. 307-345). Madrid: TEA Ediciones.
- Muñiz, J. & Fernández-Hermida, J.R. (2000). La utilización de los tests en España. *Papeles del Psicólogo*, 76, 41-49.
- Muñiz, J. & 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. & 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. & 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. & 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. & Bricmont, J. (1999). *Imposturas intelectuales*. Barcelona: Paidós (Orig. 1998).
- Sternberg, R.J. (1988). *The psychologist's. A guide to scientific writing for students and researchers*. New 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, 159-161.
- Sternberg, R. & 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. & 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. & Chiattonne, 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).