EVALUATING SCIENTIFIC JOURNALS: PROPERTIES, LIMITS AND CONDITIONS OF EFFECTIVENESS OF CLASSIFICATION METHODOLOGIES

Alessandro Carretta, University of Rome "Tor Vergata"
Gianluca Mattarocci, University of Rome "Tor Vergata"

ABSTRACT

During last years a large debate was developed among academics about research evaluation procedure. Approaches proposed in literature considers quantitative and/or qualitative aspects of each journal but each methodology has limits and characteristics high heterogeneous. The meaning and implication of results achieved with a ranking procedure is strictly influenced by the approach selected for the journal evaluation.

The paper presents a literature review of the main qualitative and quantitative approaches proposed for journal ranking focusing the attention on the main differences of approaches. The study is completed with an empirical analysis on the database Thompson Scientific, one of the main provider of quantitative rankings. The analysis considers the characteristics of quantitative rankings proposed, look at the qualitative characteristics of best and worst ranked journals and compares results obtained with those achieved by an international qualitative survey (Harzing database). Results obtained demonstrate the low degree of coherence of ranking based on different approaches and point out some risks related to the use of these approaches to evaluate research.

INTRODUCTION

The evaluation of the scientific production of an individual or an institution has always been an issue subject to a great deal of debate (Liner, 2002), and, in recent times, the lack of resources for research gives new impulse on the studies of the usefulness of these approaches (Addis et al., 2002).

First approaches used to evaluate the quality of the research make a survey among academics and/or readers in order to evaluate the reputation of the journal but all rankings based on these approaches are highly influenced by the criteria adopted in the selection of interviewers (Kostoff et al. 2001). On the second half of the Nineties, new approaches based on more objective data are proposed to substitute these subjective measures (Garfield, 1952) and during the next years a growing number of criteria available are defined for the journal ranking (OECD 1997).

This paper is meant to contribute to the debate on the reliability of the indicators used in classifying scientific journals, proposing an empirical control meant to gauge the stability over time and the consistency of the results obtained with the different methodologies, in addition to establishing a "classification capacity" of the various indicators. Results obtained demonstrate the lack of coherence among the ranking based on different bibliometric indicators, the high variability of results over time and the failure to define ranking that rewards some qualitative journal characteristics.

The paper is organized as follows: section 2 presents a literature review of qualitative criteria (section 2.1) and bibliometric approaches (section 2.2), section 3 attains the empirical analysis presenting the sample analysed (section 3.1), looking at the consistency and persistency of quantitative rankings (section 3.2), at the characteristics of best and worst journals (section 3.3) and at the coherence with qualitative rankings (sections 3.4). The last section (section 4) presents some brief conclusions and implications of results achieved.

LITERATURE REVIEW

Qualitative Evaluation of Journals

Qualitative evaluations assess the general opinion of the academic community with regard to the quality and the scientific value of the articles published in a given journal (Benjamin and Brenner, 1974).

The qualitative approach normally entails the distribution of questionnaires, in order to obtain a ranking for a group of journals held to be similar in terms of the topics addressed and/or the readership target (Webb and Albert, 1995).

In order to be able to use the data collected during the individual surveys not only for one-off, non-recurring, evaluations but also for system-wide analyses of research, databases holding the results of the surveys carried out by different

interviewers have been created[1]. Though the use of different sources of qualitative research does not guarantee that the results will be objective, the higher the number of surveys, the lower the probability that the results will be distorted by the specific interests of interviewed (Barman et al., 2001).

The first result that emerges from these surveys is the fact that the perceived value of English-language journals is generally higher respect to other publications for the high diffusion of this language in world countries (Garfield and Welljams-Dorof, 1990).

Furthermore, authors appear to prefer, with all other conditions being equal, journals that are not particularly specialised. The leaning towards publishing one's article in journals that address a wide variety of topics and that are not concerned only with certain specific themes, makes possible greater visibility with the scientific community because the number of potential purchasers of the journal is higher (Brauninger and Haucap, 2002).

The quality of the editorial staff, and especially the reputation of the editor, constitute another qualifying aspect considered in the journal selection process. The presence of a qualified editorial staff is viewed as a guarantee of the quality of the journal and of the scientific significance of the articles published therein (Smith, 2004).

Qualitative analyses of periodicals demonstrate the importance given by the academic community on the mechanism used in selecting the articles published. The factors taken into consideration are the number of articles sent to the journal, the acceptance rate, the average time from the date of acceptance and actual publication and the type of the referee process (if it is scheduled).

The presence and the characteristics of a referee process influence the evaluation of the journal, because articles published in refereed journals are selected through a procedure, structured to varying degrees, which involves outside parties, and not merely the author and the editor, meaning the so-called referees (Surinach et al., 2002). If the identity of the author of the article is not communicated to the referee (blind referee), then the process, as a rule, should reinforced the objectivity of the selection of the articles. Regardless of the specific features of the process for selecting the articles, the use of a refereeing system can be considered an indicator of the reliability of a journal, and empirical analysis proposed in the literature have demonstrated noteworthy differences in the quality of the service offered, depending on the experience, the age and the academic backgrounds of the referees (Nylenna et al., 1994).

Journals can choose between two blind-referee options: the single blind referee and the double blind referee. Under the single blind referee procedure the referee does not know the identity of the author, though the author, when results of the refereeing process are notified, will know the identity of the those who have evaluated his or her article. With a double blind referee, not only every possible reference to the identity of the author is removed from the article submitted to the referee, but the author is informed of the final judgment of the refereeing process without receiving any data about the referee's identity. Empirical analysis have shown that, on the average, the presence of a double blind referee is related to an higher frequency of citations for a given article (Laband and Piette, 1994a).

Bibliometric Indicators

The quantitative evaluation of a journal is released constructing measures that regards its circulation, measured from various perspectives (Beattie and Goodacre, 2004). The majority of the quantitative approaches are based on the assumption that the bibliographies cited in scientific articles constitute a key instrument for assessing the quality of journals, and that an analysis of the articles cited makes it possible to identify the highest quality articles in a given discipline (Wang and White, 1999). The idea behind these approaches is that, in terms of the citations included in their articles, authors tend to favour articles that analyse topics of particular relevance, propose innovative approaches, present useful points of inspiration for research or lay the groundwork for a certain discipline (Small, 2004). Seen in this perspective, attention is focussed on studying the references of published articles, with the aim of evaluating the impact on the scientific community of the publication of a particular article (Kostoff, 2002).

One of the first measures proposed was *total citations*, an indicator calculated as the sum total of the citations of the articles published in a given journal within a sample group of journals during a established time horizon ($\sum_j\sum_i$citations(i,j)). The formula is:

$$TC_j = (\sum_j\sum_i \text{citations}(i,j)) \qquad (1)$$

Its limit is the influence of the time framework, plus the greater the difference in the number of articles published by the individual journals during the reference period, the lower the reliability of the ranking (Garfield, 1971). Those who support this approach have proposed updates of the measure in order to consider the

natural and progressive growth over time of the number of citations of a given journal and the practice of *self citation* (Linton and Narongsak, 2004).

Article effectiveness is calculated as the ratio between the number of citations of articles published in a given journal ($\sum_j\sum_i$citations(i,j)) and the average number of articles published during the period under consideration by the k^{th} journal ($MeanPub_k$) (Arnold, et al., 2003). The formula is:

$$AE_k = (\sum_j\sum_i citations(i,j))/MeanPub_k \qquad (2)$$

The strength of this indicator is that it defines a ranking which takes into account the rate of production of scientific articles during the period observed, in addition to providing a useful instrument when a noteworthy change in the volume of scientific articles has occurred during the time period considered.

Impact efficiency represents the number of citations of articles published in a given journal ($\sum_j\sum_i$citations(i,j)) for each 10,000 words published in the journals taken into consideration ($\sum_i words_i$). The formula is:

$$IE_k = (\sum_j\sum_i citations(i,j))/(\sum_i words_i) \qquad (3)$$

The decision to normalise the number of citations on the basis of the number of words found in the journal makes it possible to obtain measurements that can be used to establish comparisons between journals of different sizes, though the approach is highly influenced by the style adopted by the authors in writing the articles, as well as by the topic addressed, which can call for a greater or lesser frequency of citations (Chan et al., 2004).

The most widely used indicator is the *impact factor*, an index that represents the ratio between the number of citations in a journal ($\sum_j\sum_i$citations(i,j)) and the total number of articles published over the last two years by different journals ($\sum TotArt_i$) (Garfield, 1955). The formula is:

$$IF_k = (\sum_j\sum_i citations(i,j))/ (\sum_i TotArt_i) \qquad (4)$$

The indicators presented do not make it possible to illustrate factors of noteworthy relevance, such as the immediate impact of the article published and the persistence of the citations over time. The supporting instruments proposed for analysing these factors are the *immediacy index* and the *cited half life*.

The *immediacy index* is calculated by taking the number of citations of a journal ($\sum_j\sum_i$citations(i,j)) and considers the ratio to the total number of articles published during the year of publication by the journal ($TotArt_k$), with the resulting index representing an estimate of the immediate reaction of the scientific community following the appearance of the article (Harter, 1998). Its formula is:

$$II_k = (\sum_j\sum_i \text{citations}(i,j))/(TotArt_k) \qquad (5)$$

A high index points to the most valuable articles, meaning those which are immediately considered to be relevant by authors who write on a given topic (Ahmed et al., 2004).

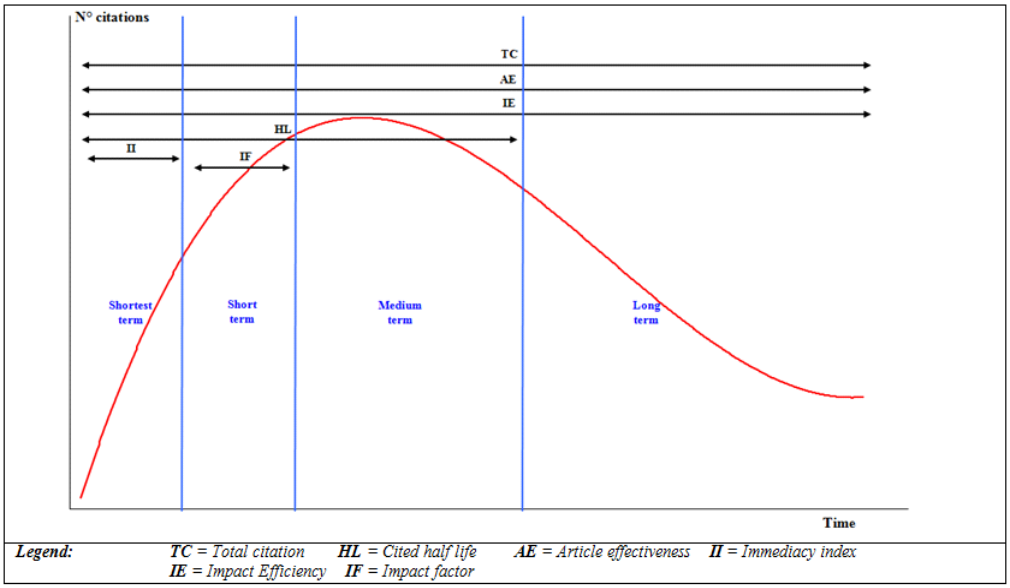
The *cited half life* represents the number of years needed for the number of citations of an article in a journal to decrease to a level that is half that of the maximum number of citations registered (Kademani and Kumar, 2002). Assessments of journals that fail to take into consideration the persistence of citations could lead to results that are not reliable, though the final result should always be examined with care: in certain cases, a high index value can simply mean that there has been a gradual deterioration in the average level of the publications on the subject, so that the citations always return to the consolidated sources in the literature (Diamond, 1989).

Based on the time frameworks normally used for application of the individual indexes, a relationship can be established between the type of index used and the curve estimating the general performance, in terms of citations, of a journal and/or article (Figure 1).

All indicators constructed through an analysis of bibliographies are affected by problems of self-citation and citations identity, which lead researchers to cite certain articles less because of their contents than on account of their authors (Fang and Rosseau, 2001).

The term self-citation refers to the natural tendency of authors to cite their own articles and/or the articles of colleagues who are part of the same sector of disciplines (Glanzel et al., 2004). The fact that such a practice is impossible to eliminate does not justify ignoring the impact it can have on estimates; in recent years there has also been a noteworthy increase in the number of articles by more than one author (Hudson, 1996), adding to the potential effect of self-citation on the evaluation of journals (Axarloglou and Theoharakis, 2003).

Figure 1: Citation results and the time framework of reference for the main quantitative indicators



Source: the authors' reformulation of the "generalized citation curve" proposed by Amin and Mabe (2000)

In writing articles, the tendency of authors, over time, is to consolidate their reference bibliography for a given topic, with the result that the citations of articles by a given author on a given subject shall not show noteworthy changes over time (citations identity). Analyses of journals that include authors who have published a number of articles on a given subject will thus be influenced by this circumstance, with the result that the greater the number of articles published by an author on a given argument, the greater the importance placed by the methodology of quantitative analysis on the reference articles used by that author (White, 2001).

The styles followed in drawing up articles differ from country to country, and there are even noteworthy differences in the approaches taken by authors to the existing literature. The differences between countries can be reflected in the use of bibliographies that are longer or shorter, or that go into greater or lesser historical depth. It follows that analyses which examine journals published in different countries can be influenced by the editorial styles followed by the authors of a given country (Bordons et al., 2002).

The number of citations registered for a journal is also tied to the characteristics of the authors whose articles it has published, meaning that, if the quality of the articles by those authors has been especially high, then the articles published in that journal will have greater visibility. It follows that the quantitative result is distorted by the importance given to results achieved in the past by subjects who shall not necessarily be presenting new contributions in the journal. Empirical analyses have shown that classifications drawn up with these indicators are relatively stable over time, making it difficult for new journals to enter the leading positions in a relatively short period of time (Laband and Piette, 1994b).

The results obtained using the quantitative approach are influenced by the time horizon considered, and an erroneous choice of valuator can render the estimate pointless for determining the current importance of the journal. Data collected at brief intervals exclude from the analysis citations in journals characterised by lengthy processes of refereeing, though such publications, thanks precisely to such procedures, could offer articles of greater scientific worth (Garfield, 2000). With the time framework defined in terms of the average article publication times, the approaches penalises the more innovative articles: an article that differs too greatly from the existing literature may not be accepted in the short run by the academic community, meaning that journals which choose to publish such articles would be penalised by the limits of the approach used (Hogson and Rothman, 1999).

Finally, in evaluating the results obtained, consideration must be given to a limitation that cannot be eliminated from the methodology, being tied to its inability to analyse the nature of the citation (Posner, 1999). In fact, an author can decide to cite another article as support of his or her own thesis (positive citation) or in order to criticise a particular approach or the results obtained (negative citation): using quantitative approaches, there is no way of distinguishing between the two types of references (MacRoberts and MacRoberts, 1989). The assessments obtained through these approaches must, therefore, take into consideration the possibility that negative citations were included in the calculation of the indicators, with the result that, the greater the number of negative citations, the less reliable the rankings obtained.

EMPIRICAL ANALYSIS

The Sample

The sample is constructed starting from the database of the journals constructed by one of the main data provider specialised in quantitative bibliometric evaluation, Thomson Scientific. The database consists of more than 1800 journals regarding 54 subjects, all falling under the category of the social sciences (Table 1)[2].

Table 1: The sample

Subject	Relevance in the sample	Subject	Relevance in the sample
Anthropology	1.96%	History of social sciences	0.21%
Applied linguistic	1.27%	Industrial relations & labor	0.48%
Applied Psychology	1.54%	Information science & library science	2.60%
Area studies	2.07%	Interdisciplinary social sciences	1.33%
Biological Psychology	0.37%	International relations	1.12%
Biomedical social sciences	0.05%	Law	3.45%
Business	1.38%	Management	1.49%
Business Finance	1.22%	Mathematical methods social sciences	0.21%
Clinical Psychology	2.66%	Mathematical psychology	0.05%
Communication	1.54%	Multi-disciplinary	30.38%
Criminology and penology	0.85%	Multi-disciplinary psychology	3.35%
Demography	0.42%	Nursing	1.81%
Developmental psychology	1.43%	Planning & development	0.69%
Economics	4.83%	Political science	2.50%
Education/educational research	3.66%	Psychiatry	2.71%
Education special	0.58%	Psychoanalysis psychology	0.53%
Educational psychology	1.12%	Public administration	0.74%
Enviromental studies	0.64%	Public environmental & education health	1.49%
Ergonomics	0.32%	Rehabilitation	0.85%
Ethics	0.42%	Social issues	0.64%

Table 1: The sample

Subject	Relevance in the sample	Subject	Relevance in the sample
Ethnic studies	0.27%	Social Psychology	1.75%
Experimental Psychology	2.02%	Social work	0.90%
Family studies	0.11%	Sociology	2.87%
Geography	1.06%	Substance abuse	0.74%
Gerontology	0.58%	Transportation	0.42%
Health policy & services	1.22%	Urban studies	0.42%
History	0.69%	Women's studies	0.64%
History & philosophy of science	1.33%		

* The category multi-disciplinary considers all journals classified by Thompson Scientific in two or more subjects

Source: the authors' processing of Thompson Scientific data

Data collected from Thompson Financial regard all the main quantitative indicators calculated for each journal during the time period 2000-2006[3]. These data are integrated with qualitative data collected from journal websites that, on the basis on a review of the literature previously presented, are relevant to distinguish among different academic journals and with results of qualitative surveys collected in the Harzing database[4].

Persistence and Consistency of the Indicators in Classifying Journals: an Empirical Control

The degree of consistency between the classifications is released by calculating the main quantitative indicators for the period considered and constructing the ranking for each one.

On the constructed ranking, journals are grouped in ten subclasses on the basis of percentiles distribution. The analysis studies the frequency with which the subclasses attributed on the basis of each quantitative indicator coincide with the subclasses determined on the basis of the other indicators (Table 2).

Table 2: Consistency between the classifications based on different approaches, year by year for the period 2000-2006 (frequency as a %)

		TC	IF	II	HL	AE
TC	Mean	100.00%	-	-	-	-
	Max					
	Min					
IF	Mean	39.20%	100.00%	-	-	-
	Max	44.62%				
	Min	28.59%				
II	Mean	28.28%	32.93%	100.00%	-	-
	Max	32.53%	37.48%			
	Min	23.04%	28.79%			
HL	Mean	30.92%	26.29%	28.23%	100.00%	-
	Max	38.26%	29.63%	39.60%		
	Min	24.26%	20.81%	23.01%		
AE	Mean	36.23%	59.46%	28.88%	25.05%	100.00%
	Max	38.26%	29.63%	39.60%	38.26%	
	Min	10.69%	10.69%	10.69%	11.65%	

Legend: ***TC*** *= Total Citations* ***IF*** *= Impact Factor* ***HL*** *= Half Life*
AE = Article effectiveness II = Immediacy Index

Source: the authors' processing of Thompson Scientific data

The empirical evidence points out a significantly low degree of coherence of the results obtained by using different indicators, seeing that the average ratio of correspondence for the classifications falls below 35% for the entire time interval considered.

The lack of correspondence between the classifications is extensive in all the disciplines considered[5], even though marked differences are more observable for certain disciplines because for certain discipline, on certain years, almost all the journals belonging to the category are affected[6].

For effective use of the results obtained from classifying journals, it is obviously best that the rankings proposed following application of the indicator

present a noteworthy level of persistence over time (Garfield, 1972), especially if the time intervals for collecting data are not too long.

The persistence over time of the different classifications was analysed by considering, for each journal, the ranking assigned for time period *t* by the various quantitative indicators and controlling whether or not there were differences, in terms of the positioning of the journals within the deciles, compared to the ranking assigned for time period *t+n*. The comparison was established between the situation for the previous year and that for the subsequent year, as well the situation at the start and the end of the overall period considered (Graph 1).

Graph 1: Persistence over time of the classifications established with the different approaches per year

Source: the authors' processing of Thompson Scientific data

The analysis of persistence for the entire period points to a scarce correspondence of the rankings in the different time intervals, especially when measures expressing the short-term impact of the publication of the articles in the journals are considered (the *impact factor* presents average levels of correspondence of less than 30%, while the *immediacy index* is equal to approximately 15%). Furthermore, a year-by-year comparison of the classifications points to noteworthy variability within the period, and especially in the case of certain sub-periods.

Quantitative Rankings and Qualitative Standards of Excellence

The formulation of quantitative rankings should make possible identification, above all else, of publications that guarantee published articles at high

standards of quality, meaning that they should represent the best sources for citations.

Looking at literature available one of the desired characteristics of a journal is represented by the international language and the Thompson scientific database is so characterized by an higher preponderance of English language journal (Table 3)

Table 3: Statistics of quantitative indicators by language for the period 2000-2006
(n° journal, mean, maximum and minimum ranking position)

Language	N° journals in the database	Percentile ranking on the basis of bibliometric indicator														
		TC			IF			II			AE			HL		
		Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min
Cezch	3	VIII	VIII	X	VIII	VIII	X	VIII	VIII	X	VIII	VIII	X	VIII	VIII	X
Dutch	1	X	X	X	IX	IX	IX	VII	VII	VII	IX	IX	IX	VII	VII	VII
English	1776	IV	I	X	IV	I	X	IV	I	X	IV	I	X	IV	I	X
French	8	VII	I	X	VII	I	X	VII	I	X	VI	I	X	VI	I	X
German	28	VI	I	X	VI	I	X	VI	I	X	VI	I	X	VI	I	X
Greek	1	VIII	VIII	VIII	VIII	VIII	VIII	VII	VII	VII	IV	IV	IV	VI	VI	VI
Japanese	3	VIII	VIII	X	VIII	VIII	X	VII	VII	X	VII	VII	X	VII	VII	X
Multi lang-uage	76	V	I	X	V	I	X	V	I	X	V	I	X	V	I	X
Portu-guese	1	VII	VII	VII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII
Spanish	6	VI	I	X	VII	II	X	VI	II	X	VII	II	X	VII	III	X
Turkish	1	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VII	VII	VII

Source: the authors' processing of Thompson Scientific data

Looking at the relationship between ranking position and the language it is possible to assume that normally journals written in English are, in mean, ranked in the better position of the ranking (IV percentile) even there is an high variability of ranking position among these journal (the range of variation considers all the possible position on the ranking). Considering journals written in other languages the analysis point out that in mean they are ranked in the second half of the

percentile distribution but the variability among different journals written in the same language do not allow to exclude the possibility to find some outliers classified in the best position of the ranking.

Many studies have highlighted, in terms of profiles of quality, the significant role of the reputation of the editorial staff in determining the success of a journal, though no established set of criteria has yet been identified to evaluate this profile.

Based on the information available, the present assessment considers, from among the qualitative profiles subject to analysis, only the characteristics of the refereeing process.

An analysis of the refereeing process makes possible the identification of three categories: journals with a system of refereeing that is not declared, or that is characterised by a noteworthy degree of arbitrary discretion on the part of the editor with respect to the procedures of the refereeing (approximately 58% of the sample); journals that have a blind referee (less than 39% of the sample); journals that utilise a double blind referee (roughly 3% of the sample).

Using the same sample presented earlier, the study of the relation between the refereeing process and the ranking shows that the quantitative indicators are capable only in part of valorising in decisive fashion the type of refereeing utilised by the journal (Table 4).

The journals that utilise a *blind* or *double blind* refereeing system obtain, on the average, higher rankings than the journals without refereeing, or than those whose refereeing is not declared, in more than 80% of the cases examined. Furthermore, the indicators for such journals register a variability that tends to be greater than that recorded for refereed journals. As a result, it is possible that the latter may obtain better positions than journals which utilise an explicit process for the selection of contributions.

Comparing Quantitative and Qualitative Rankings

The scarce relevance of certain qualitative factors in rankings based on bibliometric indicators can result in noteworthy discrepancies between qualitative and quantitative rankings.

Table 4: Descriptive statistics of quantitative indicators by type of refereeing
(mean and St. Dev.)

	Type of referee	Statistics	Type of bibliometric indicator				
			TC	IF	II	HL	AE
Time period	Not declared	Mean	746.6042	0.8568	0.1843	6.0542	0.0683
		St. Dev.	2078.754	1.0353	0.3254	3.5908	0.1912
		Mean worst year	612.5	0.7854	0.1645	36.1385	3.2803
		Mean best year	2005.783	1.008	0.3087	45.0254	6.0197
	Blind referee	Mean	649.2492	0.7954	0.1682	5.619	0.1539
		St. Dev.	1854.8876	0.9946	0.3037	3.7488	0.4401
		Mean worst year	683.1382	0.8607	0.1755	30.8666	2.7223
		Mean best year	1426.189	0.946	0.2739	38.9327	6.4956
	Double blind referee	Mean	819.6205	0.9029	0.1964	6.3805	0.0042
		St. Dev.	2246.6539	1.0659	0.3417	3.4723	0.0046
		Mean worst year	1264.69	0.7616	0.1983	58.2914	2.1091
		Mean best year	2296.601	1.099	0.2449	64.541	7.2689

Legend
TC = Total Citations HL = Half Life,,,,,
IF = Impact Factor AE = Article effectiveness,,,,,
II = Immediacy Index,,,,,,

Source: the authors' processing of Thompson Scientific data

In the interests of selecting an extensive database of qualitative surveys on international journals, consideration was given to the Journal Quality List (JQL), a database that is updated quarterly and holds the results of scientific surveys published in recent years[7]. The journals placed in the ranking and utilised in the survey regard the following disciplines: Economics (Eco), distributed among two subject headings; Finance & Accounting (F&A); Management (Mgmt), distributed among 7 subject headings; Marketing (Mkt), distributed among 5 subject headings.

Studies presented in the literature have shown that the rankings assigned to individual journals by the different surveys are, on the average, consistent with one another, meaning that the indications provided by this dataset can be considered reliable (Mingers and Harzig, 2007).

To verify the consistency with the data sample available for the bibliometric indicators, it was decided to use the data collected in the twenty-fifth edition of the JQL, published on 1 February 2007, which holds data for the surveys carried out through 2006[8].

The analysis took into consideration the relevance of the journals examined in Thompson Financial with respect both to the all the journals considered in the survey and to the journals belonging to the individual classes or disciplines (See Table 5).

The results obtained demonstrate that, on the average, only slightly less than half of the journals considered in the qualitative surveys constitute journals included in the Thompson Financial database, even though the percentage weight of the journals in the database varies significantly from survey to survey.

A detailed analysis of the role of journals for which bibliometric indicators are available within the qualitative rankings considered shows that their percentage weight in the higher quality rankings is greater. It would appear, therefore, that the presence of citation indexes ensures greater notoriety for the journal, thus increasing the probability that those interviewed will place it in the top positions of the classification (Clark, 1957).

After considering the disciplines of the journals found on both the JQL and the Thompson Financial database, a higher frequency can be noted for journals belonging to the disciplines of Economics and Management, while the bibliometric indicators prove less relevant for the disciplines of Marketing and Finance, and Accounting. Underlying the difference in relevance for certain disciplines there would appear to be a discrepancy in the extent to which the qualitative/quantitative rankings can be interchanged.

CONCLUSIONS

Classifications of journals can be drawn up using a variety of procedures designed either to make direct registration of the opinion of the scientific community or to measure the level of readership and distribution of the journal, primarily by recording the citations found in the literature of reference. The use of quantitative bibliometric indicators is the most widely used approach, given that the procedure can be replicated and verified by third parties other than the promoter of the analysis, though it presents a number of limitations that have given rise to proposals for revision of the formulation of the indicators (Hartes and Nisonger, 1997) and do not allow to achieve uniform results for the different choices taken by the different

subjects involved (authors, editors, evaluators) (Judge et al. 2007; Cacciafesta, 2007) and for the specific characteristics of the subject.

Table 5 : The relation between the rankings included on the Journal Quality List and the Thompson Financial Database

Survey data			Percentage of journals registered also in Thompson Financial database							
	Total		Break-down by class				Break-down by subject			
Survey code JQL database*	Journals considered		Class A	Class B	Class C	Class D	Eco	F&A	Mgmt	Mkt
Not95	208	53.37%	18.27%	26.92%	8.17%	0.00%	12.98%	27.20%	6.62%	14.52%
US98	125	72.00%	16.80%	28.80%	23.20%	3.20%	6.25%	12.00%	16.18%	1.61%
NL99	136	58.82%	15.44%	20.59%	22.79%	0.00%	25.48%	23.20%	26.47%	51.61%
SMJ99	62	91.94%	22.58%	22.58%	19.35%	27.42%	8.65%	9.60%	10.29%	24.19%
Wiei01	483	57.97%	44.72%	12.42%	0.62%	0.21%	12.98%	27.20%	6.62%	14.52%
UQ03	345	46.67%	8.12%	17.68%	16.52%	4.35%	6.25%	12.00%	16.18%	1.61%
VHB03	393	52.42%	28.75%	15.27%	6.11%	2.29%	25.48%	23.20%	26.47%	51.61%
BJM04	401	58.35%	27.68%	21.95%	8.48%	0.25%	8.65%	9.60%	10.29%	24.19%
CNRS04	290	71.03%	23.45%	20.00%	17.59%	10.00%	12.98%	27.20%	6.62%	14.52%
ESS05	257	60.31%	48.25%	11.67%	0.39%	0.00%	6.25%	12.00%	16.18%	1.61%
HKB05	322	55.28%	13.04%	15.22%	20.19%	6.83%	25.48%	23.20%	26.47%	51.61%
Theo05	257	64.59%	5.06%	3.50%	10.51%	45.53%	8.65%	9.60%	10.29%	24.19%
Ast06	598	56.02%	37.12%	12.37%	6.19%	0.33%	12.98%	27.20%	6.62%	14.52%
Cra06	361	62.60%	17.73%	28.53%	13.02%	3.32%	6.25%	12.00%	16.18%	1.61%
EJL06	268	57.09%	26.12%	0.37%	30.60%	0.00%	25.48%	23.20%	26.47%	51.61%
HMB06	474	54.22%	25.32%	22.15%	5.70%	1.05%	8.65%	9.60%	10.29%	24.19%
FT06	34	94.12%	94.12%	0.00%	0.00%	0.00%	12.98%	27.20%	6.62%	14.52%

* For further details on the meaning of the symbols proposed, see table A1 in the appendix, which presents the main data from the different surveys

Source: the authors' processing of data from Thompson Scientific and the Journal Quality List

The empirical assessment performed in the paper highlights certain instances of discontinuity in the classifications, as well as a low level of consistency in terms of the results. This last observation is not entirely justifiable, given that the different indicators all use the same source of information (the bibliographic

citations of the articles) (Block and Gary, 2001), in addition to which there exists a general relation between the indicators and the life cycle of the journal and its citations, which should produce essentially consistent results, apart from a few individual differences tied to the specific characteristics of each indicator.

The combined analysis of the qualitative and quantitative profiles shows that the quantitative indicators lack an adequate capacity to reward journals that present objectively better qualitative features, such as a rigorous process of refereeing. As a consequence, there are noteworthy differences in the results produced by the qualitative and quantitative evaluations that do not allow to consider these approaches as substitutes and in order to understand the journal raking is necessary to consider limits and characteristics of the approach used for the classification.

Furthermore, an evaluation of academic production carried out by analysing the articles published in journals provides only a partial vision of research activities (Moore et al., 2002), one judged a priori to be representative of the sum total of the most innovative contributions for a given discipline (Ding et al, 2000; Carretta, 2006), even though it should be supplemented with the study of other profiles. Along these lines, it is worth noting the profile of academic works produced in different formats, pointing to the advisability of valorising published monographs as well (Gray et al., 1997), in order to avoid unjustifiably penalising (apart from an arbitrary ex ante selection) those subjects who prefer to publish monographs or volumes (Seglen, 1997) and excluding all the citations of journals made by such authors in their books (Johnes and Johnes, 1993). The next step of the analysis will be to define criteria, methodologies and databases that allow to evaluate also these type of contribution in order to define more complete journal ranking.

ENDNOTES

* This article is taken from a wider-ranging work published in Italian as a supplement in Banking & Finance Lab, no. 1, 2008, and also available on the site of the Association of Professors in Economics of Financial Intermediaries and Markets (www.adeimf.it). Reference should be made to the larger work for an all-encompassing overview of the issue in both theoretical and empirical terms. The work is a joint effort by the two authors. Alessandro Carretta wrote sections 1 and 4, while Gianluca Mattarocci wrote sections 2 and 3. Authors are grateful to Fabrizio Cacciafesta, Roberto Cafferata, Giacomo de Laurentis, Franco Fiordelisi, Mario Masini, Paolo Mottura, Luciano Munari, Claudio Porzio, Daniele

Previati, Francesco Ranalli, Paola Schwizer and Giulio Tagliavini for all suggestions given.

[1] See, for example, Harzing, A.W. (various years), *Journal Quality List,* www.harzing.com.

[2] For further details about journals included in each category, see the internet website of Thompson Scientific (http://scientific.thomson.com/).

[3] In the Thomson Scientific database are available data about all bibliometric indicators previously presented except for the impact efficiency. In the analysis this measure had to be excluded.

[4] For further details about the Harzing database see www.harzing.com.

[5] The sample group of journals was segmented in accordance with the 54 sub-classes proposed by Thompson Scientific.

[6] For further details see Table A1 in the appendix.

[7] Cfr. Harzing (various years), *Journal quality list*, www.harzing.com.

[8] For more detailed information on the surveys considered as part of the analysis, as well as on the characteristics of the individual surveys and the criteria followed to ensure that the number of classes of journals considered in the different rankings is uniform, see Table A2 in the appendix.

REFERENCES

Addis M., M. Costabile & K. Matzler (2002). La valutazione delle pubblicazioni italiane per gli studiosi di marketing. *Finanza, Marketing e Produzione,* 20 (3) 70-95

Ahmed T., B. Johnson, C. Oppenheim & C. Peck (2004). Highly cited old papers and the reasons why they continue to be cited, Part II. *Scientometrics,* 61 (2)147-156

Amin M. & M. Mabe (2000), Impact Factors: use and abuse. *Perspectives in Publishing,* 1(10) 1-5

Arnold T., A.W. Butler, T.F. Crack & A. Altintig (2003). Impact: what influences finance research?. *Journal of Business,* 72 (3) 343-361

Axarloglou K. & V. Theoharakis (2003). Diversity in economics: an analysis of journal quality perceptions. *Journal of the European Economic Association,* 1(6), 1402-1423

Barman S., M.D. Hanna & R.L. Laforge (2001). Perceived relevance and quality of POM journals: a decade later. *Journal of Operations Management,* 19(3), 367-385

Beattie V. & A. Goodacre (2004). Publishing patterns within the UK accounting and finance academic community. *British Accounting Review,* 36 (1) 7-44

Benjamin J.J. & V. Brenner (1974). Perceptions of journal quality. *Accounting Review*, 49(2) 360-362

Block S. & W. Gary (2001). The Impact Factor: time for change. Australian and New Zealand *Journal of Psychiatry*, 35 (5) 563-568

Bordons M., M.T. Fernandez & I. Gomez (2002). Advantages and limitations in the use of the impact factor measures for the assessment of research performance in peripheral countries. *Scientometrics,* 53 (2) 195-206

Brauninger M. & J. Haucap (2002), What economists think of their journals and how they use them: reputation and relevance of economic journals, University of the Federal Armed Forces discussion papers in economic policy n° 111, Hamburg

Cacciafesta F., (2007). Sulla valutazione (dei ricercatori e degli articoli di ricerca). *Notiziario della unione matematica italiana,* 7 (3) 31-33

Carretta A. (2006). Gli studi di economia degli intermediari finanziari: una prospettiva di analisi basata sulla letteratura citata. in L. MUNARI (eds), *Attualità e prospettive negli studi di economia dei mercati e degli intermediari finanziari* (pp. 1103-1109). Parma: Monte Università Parma Editore

Chan K.C., C.R. Chen & T.L. Steiner (2004). Who is publishing? An analysis of finance research productivity in the European region. *Journal of Business, Finance and Accounting,* 31(3-4), 401-437

Clark K.E. (1957), *America's Psychologists: a survey of growing profession,* Washington: American Psychological Association

Diamond A.M. (1989). The core journals of economics, *Current Contents,* 1 (1) 4-11

Ding Y., G.G. Chowdhury & S. Foo (2000). Journals as makers of intellectual space: journal co-citation analysis of information retrieval area, 1987-1997, *Scientometrics*, 47(1), 53-73

Fang Y. & R. Rosseau (2001), Lattices in citations networks: an investigation into the structure of citation graphs. *Scientometrics,* 50(2), 273-287

Garfield E. (1955). Citation indexes for science: a new dimension in documentation through association of ideas *Science,* 122(3159), 108-111

Garfield E. (1971). Citation indexing, historio-bibliography and the sociology of science. *Essays of an Information Scientist,* 1(1), 158-174

Garfield E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(11), 471-479

Garfield E. (2000). Use of journal citation reports and journal performance indicators in measuring short and long term journal impact. *Croatian Medical Journal,* 41(4), 368-374

Garfield E. & A. Welljams-Dorof (1990), Language use in international research: a citation analysis. *Annals of the American Academy of Political and Social Science,* 511(1), 10-24

Glanzel W., B. Thijs & B. Schlemmer (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics,* 59(1), 63-77

Gray R.H., J. Guthrie & L.D. Parker (1997), Publish or perish: the social construction of quality in accounting and management research, *Fifth Interdisciplinary Perspectives on Accounting Conference* 4, 1111 - 1123

Hartes S.P. & T.E. Nisonger (1997). ISI's impact factor as misnomer: a proposed new measure to assess journal impact. *Journal of American Society for Information Science*, 48 (12) 1146-1148

Harzing A.W. (different years), *Journal Quality List,* Retrieved December, 2008, from http://www.harzig.com

Hogson G.M. & H. Rothman (1999). The editors and authors of economics journals: a case of institutional oligopoly. *Economic Journal,* 109 (453) 165-186

Hudson J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives,* 10 (3) 153-158

Johnes G. & J. Johnes (1993). Measuring the research performance of UK economics departments: an application of data envelopment analysis. *Oxford Economic Papers*, 45 (2) 332-347

Judge T.A., D.M. Cable, A.E. Colbert & S.L. Rynes (2007). What causes a management article to be cited - article, author or journal?. *Academy of Management Journal,* 50(3) 491-506

Kademani B.S. & V. Kumar (2002), Citation index: an indispensable retrieval tool for research and evaluation, presented to the ICSSR Sponsored Training Workshop on Exploring Social Science Information in a Digital Environment, Mumbai, India

Kostoff R.N. (2002). Citation analysis of research performer quality. *Scientometrics*, 53(1) 49-71

Kostoff R.N., J.A. De Rio, J.A. Hurnenik, E.O. Garzia & A.M. Ramirez (2001). Citation mining: integrating text mining and bibliometrics for research user profiling. *Journal of the American Society for Information Science and Technology*, 52(13), 1148-1156.

Laband D.N. & M.J. Piette (1994a). A citation analysis of the impact of the blinded peer review. *Journal of the American Medical Association,* 272 (2)147-149

Laband D.N. & M.J. Piette (1994b). The relative impacts of economics journals: 1970-1990. *Journal of Economic Literature*, 32 (2) 640-666

Liner G.H. (2002). Core journals in economics. *Economic Inquiry*, 40 (1) 138-145

Linton J.D. & T.T. Narongsak (2004). Perspective: ranking the technology innovation management journals. *Journal of Product Innovation Management*, 21 (2) 123-139

Macroberts M.H. & B. Macroberts (1989). Problems of citation analysis: a critical review. *Journal of American Society for Information Science,* 40 (5) 342-349

Mingers J. & A.W. Harzing (2007). Ranking journals in business and management: a statistical analysis of the Harzing dataset, *European Journal of Information Science,* 16 (4) 303-316

Moore W.J., R.J. Newman, P.J. Sloane & J.D. Steely (2002), Productivity Effects of Research Assessment Exercises, Louisiana State University Departmental Working Papers n˚ 2002-15, Baton Rouge

Nylenna M., P. Riis & Y. Karlsson (1994). Multiple blinded reviews of the same two manuscripts. *Journal of the American Medical Association,* 272 (2) 149-151

OECD (1997), The evaluation of scientific research: selected experience, presented to the OECD Workshop on the Evaluation of Basic Research, Paris, France

Posner R.A. (1999), The theory and practice of citations analysis with special reference to law and economics, University of Chicago Law School, John M. Olin Law & Economics Working Paper No. 83 Chicago

Seglen P.O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314 (7079), 498-502

Small H. (2004). Why authors think their papers are highly cited. *Scientometrics,* 60 (3), 305-316

Smith S.D. (2004). Is an article in a top journal a top article. *Financial Management*, 33 (4)133-149

Surinach J., J.C. Duque, R. Ramos & V Royuela (2002), Authors', institutions' and countries' ranking in regional and urban science: an analysis for nine top international journals from 1991 to 2000, University of Barcelona working paper n˚ 92, Barcelona

Wang P. & M.D. White (1999). A cognitive model of document use during a research project, study ii. decision at the reading and citing stage. *Journal of American Society for Information Science*, 50(2) 98-114

Webb J.R. & J.D. Albert (1995). Evaluating the real estate journals: the mainstream finance perspective. *Journal of Real Estate Research,* 10 (2) 217-226

White H.D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology,* 52 (2) 87-108

APPENDIX

Table A1: Consistency between the classifications based on different approaches for subject (frequency as a %)

	2000	2001	2002	2003	2004	2005	2006		2000	2001	2002	2003	2004	2005	2006
Anthropology	24.59%	29.73%	27.30%	24.59%	31.08%	29.46%	12.43%	History of social sciences	10.00%	17.50%	20.00%	25.00%	37.50%	67.50%	60.00%
Applied linguistic	38.33%	56.67%	48.75%	47.08%	47.50%	38.75%	29.58%	Industrial relations & labor	22.22%	21.11%	18.89%	22.22%	24.44%	18.89%	10.00%
Applied psychology	35.33%	34.00%	29.00%	28.67%	29.67%	24.67%	12.67%	Information science & library science	22.08%	30.42%	30.42%	29.79%	26.88%	35.42%	22.71%
Area studies	26.41%	37.95%	39.74%	44.36%	40.00%	41.79%	33.08%	Interdisciplinary social sciences	25.20%	26.00%	26.80%	22.40%	20.00%	20.80%	12.80%
Biological psychology	42.86%	40.00%	37.14%	31.43%	41.43%	41.43%	25.71%	Intern-ational relations	14.29%	20.00%	26.19%	26.67%	25.71%	32.38%	23.33%
Biomedical social sciences	100.00%	100.00%	20.00%	30.00%	30.00%	40.00%	100.00%	Law	29.54%	32.31%	28.77%	29.85%	30.77%	28.77%	17.38%
Business	39.63%	39.26%	45.19%	42.59%	46.67%	32.59%	26.67%	Manage-ment	35.00%	40.71%	36.07%	33.21%	25.36%	22.86%	13.93%
Business finance	45.22%	54.35%	53.04%	47.39%	40.00%	29.13%	17.39%	Mathematical methods social sciences	35.00%	52.50%	42.50%	35.00%	7.50%	10.00%	30.00%
Clinical psychology	26.47%	33.53%	36.08%	34.90%	27.45%	28.04%	12.75%	Mathematical psychology	0.00%	10.00%	10.00%	10.00%	0.00%	30.00%	10.00%
Communication	26.21%	43.10%	42.41%	39.31%	45.86%	36.55%	31.03%	Multidisciplinary	24.34%	27.41%	25.71%	25.50%	24.97%	25.42%	14.67%
Criminology and penology	48.82%	47.06%	44.12%	44.71%	34.71%	36.47%	16.47%	Multidiscip-linary psychology	21.75%	26.03%	24.92%	29.21%	28.73%	30.32%	17.30%
Demography	16.25%	15.00%	22.50%	18.75%	33.75%	18.75%	8.75%	Nursing	28.33%	34.72%	39.17%	35.83%	37.22%	36.11%	22.78%
Developmental psychology	30.00%	29.64%	30.00%	23.93%	29.64%	24.29%	16.43%	Planning & development	54.62%	53.08%	36.15%	21.54%	20.77%	30.77%	11.54%
Economics	26.81%	32.75%	29.34%	28.68%	28.90%	29.45%	16.92%	Political science	25.96%	33.40%	32.34%	34.04%	31.28%	29.57%	18.51%
Education and educational research	35.14%	36.11%	35.00%	35.14%	34.58%	30.97%	19.58%	Psychiatry	32.55%	29.22%	32.75%	33.14%	35.29%	37.25%	25.10%
Education special	26.36%	42.73%	31.82%	20.00%	30.00%	49.09%	44.55%	Psychoanalysis psychology	28.00%	51.00%	42.00%	38.00%	45.00%	44.00%	28.00%

Table A1: Consistency between the classifications based on different approaches for subject (frequency as a %)

	2000	2001	2002	2003	2004	2005	2006		2000	2001	2002	2003	2004	2005	2006
Educational psychology	16.67%	16.19%	29.05%	31.90%	17.14%	18.57%	11.43%	Public administration	21.43%	37.86%	41.43%	25.00%	26.43%	20.00%	19.29%
Environmental studies	53.85%	48.46%	49.23%	50.00%	27.69%	27.69%	8.46%	Public environmental & education health	46.77%	46.45%	37.10%	40.65%	33.55%	35.48%	16.13%
Ergonomics	26.67%	40.00%	18.33%	20.00%	15.00%	40.00%	13.33%	Rehabilitation	14.38%	26.88%	22.50%	20.00%	29.38%	21.88%	10.63%
Ethics	45.00%	42.50%	35.00%	30.00%	25.00%	26.25%	12.50%	Social issues	30.83%	49.17%	51.67%	50.83%	49.17%	35.00%	29.17%
Ethnic studies	34.00%	56.00%	48.00%	40.00%	48.00%	44.00%	34.00%	Social psychology	40.30%	39.70%	34.24%	29.70%	26.97%	30.91%	11.21%
Experimental psychology	39.74%	43.08%	40.51%	41.03%	39.49%	37.95%	17.44%	Social work	15.29%	24.71%	25.29%	27.65%	31.18%	32.94%	22.35%
Family studies	10.00%	20.00%	10.00%	15.00%	5.00%	5.00%	5.00%	Sociology	23.70%	30.74%	30.93%	33.70%	37.78%	28.89%	20.74%
Geography	20.50%	20.50%	23.50%	23.50%	22.00%	21.00%	10.50%	Substance abuse	17.69%	20.77%	19.23%	19.23%	26.15%	25.38%	16.92%
Gerontology	21.82%	22.73%	25.45%	35.45%	29.09%	24.55%	20.91%	Transportation	22.22%	31.11%	36.67%	25.56%	42.22%	38.89%	27.78%
Health policy & services	29.13%	34.35%	40.00%	38.26%	34.78%	28.26%	23.48%	Urban studies	36.25%	42.50%	30.00%	31.25%	40.00%	51.25%	30.00%
History	16.15%	34.62%	30.00%	30.00%	32.31%	40.00%	24.62%	Women's studies	20.83%	27.50%	34.17%	34.17%	23.33%	27.50%	21.67%
History & philosophy of science	16.80%	26.80%	25.60%	27.60%	18.80%	21.20%	18.00%								

Source: the authors' processing of data on Thompson Scientific

Table A2: The Qualitative Rankings considered on the Journal Quality List

JQL symbol	Promoting organisation	Nation and Year	Distinctive features of the survey	Criteria for ensuring uniformity of the number of classes			
				Class A	Class B	Class C	Class D
Not95	Nottingham University	UK -1995	Scholars from 27 different institutions	4<Rank≤5	3<Rank≤4	3≤Rank<4	1≤Rank≤2
US98	Virginia Commonwealth University	USA-1998	Scholars from American universities	0.75<Rank≤1	0.5<Rank≤0.75	0.25<Rank≤0.5	0≤Rank≤0.25
NL99	Netherlands Academics in Business Administration	Holland-1999	Dutch scholars of Business Administration	Rank=A or A(P)	Rank=B or BP	Rank=C or CP	-
SMJ99	Strategic Management Journal	1999	Frequencies of citations in a sample group of 17 journals	0<Rank≤15	15<Rank≤30	30<Rank≤45	45<Rank≤65
Wiei01	Wirtschaftsuniv erisitat Wien	Austria-2001	In-house survey	Rank A or A+	Rank B	Rank C	Rank D
UQ03	University of Queensland	USA-2003	In-house survey	Rank 1	Rank 2	Rank 3	Rank 4 or 5
VHB03	Association of Professors of Management in German speaking countries	Germany-2003	Scholars ad researchers in Germany	Rank=A or A+	Rank=B	Rank=C	Rank=D or E
BJM04	British Journal of Management	UK-2004	Ranking of scientific and academic institutions in the UK	5.5<Rank≤7	4<Rank≤5.5	2.5<Rank≤4	1<Rank≤2.5
CNRS04	Centre de la Recherche Scientifique	France-2004	Opinion of select experts	Rank=5 or 4	Rank=3	Rank=2	Rank=1
ESS05	ESSEC Business School Paris	France-2005	Opinion of 7 scholars of the ESSEC	Rank=0 or 1	Rank=2	Rank=3	Rank=4
HKB05	Hong Kong Baptist University School of Business	Hong Kong-2005	List approved by the HBKU Executive Committee	Rank=A	Rank=B+	Rank=B	Rank=B-
Theo05	Survey performed by Theoharakis et al.	2005	Opinions of scholars and doctoral candidates in 7 different disciplines	71<Rank≤95	47<Rank≤71	23<Rank≤47	0≤Rank≤23
Ast06	Aston Business School	UK-2006	Opinions of scholars of the University of the Midlands	Rank=3	Rank=2	Rank=1	Rank=0
Cra06	Cranfield University School of Management	UK-2006	In-house survey	Rank=4	Rank=3	Rank=2	Rank=1
EJL06	Erasmus Research Institute of Management	Holland -2006	In-house survey	Rank=STAR or P	Rank=PA	Rank=S	Rank=SD

Table A2: The Qualitative Rankings considered on the Journal Quality List

JQL symbol	Promoting organisation	Nation and Year	Distinctive features of the survey	Criteria for ensuring uniformity of the number of classes			
				Class A	Class B	Class C	Class D
HMB06	Harvey Morris Business Journal Listing	UK-2006	Opinions of university rectors and directors of research centres in the UK	Rank=4 or 4*	Rank=3	Rank=2	Rank=1
FT06	Financial Times Survey	2006	Research objective is a ranking of top business schools	Classified	-	-	-

Source: the authors' processing of data on the Journal Quality List

www.alliedacademies.org