Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics Procedia 33 (2012) 244 – 251

Physics Procedia

2012 International Conference on Medical Physics and Biomedical Engineering

Research on the comparability of multi-attribute evaluation methods for academic journals

Yu Liping

School of Business,Ningbo University,Ningbo, China
chinayangzhou@yahoo.com.cn

Abstract

This paper first constructs a classification framework for multi-attribute evaluation methods oriented to academic journals, and then discusses the comparability of the vast majority of non-linear evaluation methods and the majority of linear evaluation methods theoretically, taking the TOPSIS method as an example and the evaluation data on agricultural journals as an exercise of validation. The analysis result shows that we should attach enough importance to the comparability of evaluation methods for academic journals; the evaluation objectives are closely related to the choice of evaluation methods, and also relevant to the comparability of evaluation methods; the specialized organizations for journal evaluation had better release the evaluation data, evaluation methods and evaluation results to the best of their abilities; only purely subjective evaluation method is of broad comparability.

Keywords:academic journals; evaluation methods; comparability

1. Introduction

Academic journal is an important bridge for translating knowledge innovation and scientific and technological achievements into productive forces, and it plays an irreplaceable role in promoting scientific and technological progress in a society. Journal evaluation is an important component of bibliometric study. It tries to reveal the regularity in the distribution of papers among journals through the quantitative analysis of the development pattern and growth trend in journals, so as to provide insight for the optimization of academic journals. At the same time, it may help improve the quality of academic journals, and promote the healthy growth and development of academic journals. The journal evaluation theory originated from Dr. E. Garfield (1963).

The quantitative methods for journal evaluation include two broad categories, the single-index evaluation and the multi-index comprehensive evaluation. Various indicators have been designed for

 doi:10.1016/j.phpro.2012.05.058

journals evaluation. The traditional indicators are relatively simple, containing small amount of information, such as the impact factor, cited half-life, and the share of grant-supported papers in total papers. Later, the traditional indicators are integrated into a number of composite indicators, which include a larger amount of information, with the typical examples being FCSm (Moed et al., 1995) , H index (Hirsch, 2005) , and ACIF (Markpin, 2008) .

Since the single-index evaluation methods provide a limited amount of information, the multi-index comprehensive evaluation (also known as the multi-attribute evaluation,MAE) methods have been widely used. Weiping Yue and Concepcion S. Wilson (2004) established an analytical framework for journal influential force using the principle of the structural equation, but they did not conductan empirical tast. Xinning Su(2008) evaluated the Chinese journals in humanities and social sciences by use of the weighted indicators. Junping Qiu, Rong Zhang, et al (2004) proposed a three-dimensional hierarchy for journal evaluation indicator framework, and carried out an evaluation exercise through gray correlation method. Jingan Pang, Yuhua Zhang, et al (2000) and Kaiyang Li, Yuping Jiang(2005) used AHP to evaluate journals. Xiaowei Wang, Bo Yang et al (2003) took the evaluation results of journals' past performance as a measure for journals' basic condition, and then re-estimated their relative evaluation value by use of data envelopment analysis (DEA). Xiujie Li and Jingwu Chen(2006) established an index framework for journals evaluation using discriminant analysis. Jiu Wang and Tianhe Xu (2003) carried out a comprehensive evaluation of the academic quality of medical journals using RSR methods. Hanzhong Chen(2004) evaluated the academic journals by means of principal component analysis. Chunyan Ling and Lin Mo (2004) proposed a comprehensive attributes evaluation model for the quality of natural science journals and conducted an evaluation exercise.

Through the above literature review, we see that there have been dozens of multi-attribute evaluation methods to date, and many of them have already been applied to academic journals evaluation. It can be expected that more of new multi-attribute evaluation methods will appear. The academic community has formed a consensus on a shortcoming of journal evaluation, namely different evaluation methods towards the same objects would lead to different evaluation results. However, another issue in journal evaluation has gained little, if any, attention, namely the comparability of the evaluation results over academic journals in different disciplines based on the same evaluation methods, or the comparability of the evaluation results on the performance of the same disciplinary journals in different time periods (this type of comparability could also be known as "inheritance"). Since there are huge amount of academic journal titles in the world, evaluation methods must be chosen carefully to ensure fair evaluation.

Observing from the horizontal perspective, although the evaluation results for journals of different disciplines are not comparable to each other, we had better seek to adopt the same evaluation method, and we had better pay special attention to this issue in the evaluation of different sub-discipline journals in given discipline or field. For example, medicine is a big field. There are 54 clinical journals and 43 surgery journals in the 2007 edition of "China Scientific and Technical Journal Citation Report". If we use the principal component analysis to evaluate clinical medicine journals and use the gray correlation method to evaluate surgery journals in our medical journals evaluation, then this approach is clearly inappropriate. Preferably, we had better use the same evaluation method. A more important issue is that even if we use the principal component analysis to evaluate both clinical medicine journals and surgery journals, is the nature of two evaluation exercises same? As the specific data processing is concerned, there are three kinds of treatments: using the principal component analysis to evaluate 54 clinical journals alone; using principal component analysis to evaluate 43 surgery journals alone; using principal component analysis to evaluate 97 clinical medicine and surgery journals as a whole. Are the first and the second approaches comparable? Is it reasonable to adopt the third evaluation approach instead?

Observing from the vertical perspective, if we use principal component analysis to evaluate the 54 clinical journals in 2006 and 2007 separately, can we simply say that the quality of a journal is improving or shrinking based on the two years' evaluation results? Or, are the evaluation results of the two years comparable? Further, is the result of principal component analysis heritable? What kinds of evaluation methods are inheritable or comparable?

This paper first classifies the evaluation methods, and then builds an analytical framework, afterwards carries out an empirical research with the TOPSIS method for the purpose of exploring comparability of evaluation methods for academic journals in depth.

2. Research methodology

2.1Analytical Framework

Here we construct an analytical framework, as shown in Figure 1. The multi-attribute journal evaluation methods are divided into two categories, the linear and nonlinear evaluation methods. The so-called linear evaluation means that one adopts subjective or objective way to give weights to the evaluation indicators, and then carries out the linear summation of weighted indicators value to obtain the evaluation results of the journals. Such methods include the analytic hierarchy process (AHP), expert panel evaluation, entropy weight method, probability weight method and so on. The so-called non-linear evaluation means that one adopts fuzzy mathematics, operations research and other systematic approaches for evaluation, such as the principal component analysis, gray correlation analysis, data envelopment analysis, and TOPSIS, where non-linear relations exist between the indicators and the evaluation results. Further, the linear evaluation methods can also be divided into two groups, in one group one meets the data-dependent weights and in another group the weights are determined independent of the data involved. The former include the entropy weight method and the variation coefficient method; the latter include the analytic hierarchy process and the expert panel evaluation.

Then, what is the relationship between non-linear evaluation methods and evaluation data? So far, almost all evaluation results based on non-linear evaluation methods are heavily data-related, and they rely on the data.

In the above example, let's first consider the evaluation methods with data-dependent weights. If we like to evaluate medical journals, taking the different disciplines involved into account, it is generally not appropriate to evaluate hundreds of medical journals with the same method. Why? Take the entropy weight method as an example. If we only evaluate 54 clinical journals, the calculated weight of impact factor may be 0.234, but if we evaluate more than 500 medical journals together, the calculated weight for impact factor may be 0.195. In another word, the entropy weights for 54 o journals and for more than 500 journals are totally different. By the same token, the entropy weights for 54 clinical medicine journals and for 43 surgery journals would be also totally different even if we adopt the same evaluation indicators. That is to say, linear evaluation methods with data-dependent weights are not comparable horizontally.

On the vertical comparison, with the entropy weight method, the indicators weights for 54 clinical journals depending on the 2006 data and 2007 data respectively are not the same. That is to say, the linear evaluation methods with data-dependent weight are not comparable vertically.

For linear evaluation methods, since the weights are determined independent of evaluation data, so the weights for different indicators may be like this: impact factor (0.25), total cites (0.2), cited half-life (0.1), disciplinary impact (0.2) share of grant-supported papers in total papers (0.1), citing half-life (0.1), immediacy index (0.05). In this case, both the evaluation for 54 clinical medicine journals and the evaluation for 43 surgery journals are comparable horizontally. Similarly, for 54 clinical medicine journals, the evaluation results based on the data of different years are also comparable. In other words,

linear evaluation methods with the weights independently determined are not only horizontally comparable but also vertically comparable or inheritable.

Since all the non-linear evaluation methods are highly dependent on data, they are neither horizontally comparable nor vertically inheritable, unlike the cases for linear evaluation methods with data-dependent weight.

To date, there have been dozens of multi-attribute evaluation methods, which are mostly non-linear evaluation methods. Many of the remaining linear evaluation methods are the type of data-dependent weight. Only the expert panel evaluation, analytic hierarchy process and a few other methods are horizontally and vertically comparable, which is often overlooked by the academic community.

Then, how should one compare different evaluation methods? As the principles of different evaluation methods are different, it is only too natural to find that different evaluation methods themselves are not comparable, and it is difficult to conduct thorough and meticulous comparison horizontally. Taking into account the important status of weight in multi-attribute evaluation, this paper uses simulated weight through regression analysis to make the comparison.

Regression analysis is originally a calculation method for investigating the specific dependence of a variable on the other (s). It inspects the overall mean of the explained variables based on the known or given value of the explanatory variables, namely, when the explanatory variable takes a certain value, the average value of all possible corresponding values of the statistically associated explained variable.

For the majority of non-linear evaluation methods, we may suppose that all the individual indicators are the explanatory variables, and the total index value is the explained variable. So the following model may be established:

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3 \ldots + b_n x_n \quad (1)$$

Where, x1,x2,x3…xn are indicator values, and b1,b2,b3…bn actually become the weights after normalization treatment.

2.2Choice of non-linear evaluation methods

There are many non-linear evaluation methods. This paper takes TOPSIS as an example to analyze the comparability and inheritance of the evaluation methods. TOPSIS (Technique for Order Performance by Similarity to Ideal Solution) is a classical multiple attribute decision making method, which is developed by Hwang and Yoon (1981). It is based on the concept that the chosen alternative should have the shortest distance to the Positive Ideal Solution (the solution that maximizes the benefit criteria and minimizes the cost criteria) and the farthest distance to the Negative Ideal Solution (the solution that maximizes the cost criteria and minimizes the benefit criteria).

3. Data

The evaluation data in this paper come from the agricultural academic journals recorded by CSTPC database of the Institute of Scientific & Technical Information of China. The Institute of Scientific & Technical Information of China has carried on the statistical analysis over the publication amount of China's scientists and citations those publications gathered since 1987, established the Chinese scientific and technical papers and citation database, and also publishes "Chinese Academic Journal Citation Report" annually. In order to analyze the comparability and inheritance of journal evaluation, we select the 2005 and 2006 data on the agricultural academic journals, but delete some journals whose data are incomplete and a small number of new journals, and analyze the two years' panel data on 96 agricultural journals. For a comparison between the evaluation methods, for convenience, this paper only selects a total of five indicators, namely, total cites, impact factor, share of grant-supported papers in total papers, cited half-life, disciplinary impact.

Before the evaluation of journals, we must treat data through normalization. Set the maximum of each indicator to be 100, and then conduct pro-rata adjustment. In addition, since the cited half-life is a

negative indicator, it needs a proper treatment. The method here is to use 100 minus its normalization results and then make normalization again over the above difference, which turns it into a positive indicator.

4. Empirical results

4.1Comparison of simulation weights

First we evaluate 96 agricultural journals based on 2005 data through TOPSIS, and then make a regression with the evaluation results as dependent variables and the five evaluation indicators as independent variables, afterwards treat the data in 2006 following the same method. The results are shown in Table 1. As a weights simulation method for TOPSIS, regression analysis is quite effective here. The goodness of fit R2 for the data of both years is high, all above 0.98, and t-test values of all the evaluation indicators are significant, passing the statistical test on the level of 1%.

TABLE 1 Regression and simulation weights

	R^2	X_1	X_2	X_3	X_4	X_5
Regression(2005)	0.98	0.166^{***}	0.145^{***}	0.203^{***}	0.221^{***}	0.164^{***}
		（13.555）	（9.120）	（29.524）	（34.281）	（21.773）
Regression(2006)	0.99	0.179^{***}	0.162^{***}	0.197^{***}	0.203^{***}	0.173^{***}
		（18.672）	（14.705）	（31.695）	（43.506）	（26.864）
Simulation weights(2005)	--	0.185	0.161	0.226	0.246	0.182
Simulation weights(2006)	--	0.196	0.177	0.216	0.223	0.189
WeightsMOL(%)	--	5.95	9.94	-4.42	-9.35	3.85

Normalize the regression coefficients, and obtain the TOPSIS simulation weights for 2005 and 2006 data. Comparing the 2005 simulation weights with the 2006 weights, we find that the change range of absolute error is between 3.85% ~ 9.94%. Though it is within 10%, the error is still too large. It is somewhat similar to the case when the weights are given through expert panel discussion. The weights given by the experts based on the data of last year and those of this year are not the same. Different combinations of weights are equivalent to different evaluation methods, which are not comparable theoretically. However, if we evaluate the 2005 and 2006 journals with the 2005 weights, they are comparable.

4.2Comparison of evaluation results

In order to compare the evaluation results of different evaluation methods, this paper presents the evaluation results and journal ranks of four evaluation methods: 2005 TOPSIS evaluation results, 2006 TOPSIS evaluation results, 2005 data evaluation results with 2006 simulation weights, 2006 data evaluation results with 2005 simulation weights. Taking into account the large amount of data, this paper gives only the top 30 journals with the 2005 simulation weights evaluation, as shown in Table 2.

TABLE 2. Comparison of different evaluation results

	TOPIS evaluation and journal rank (2005)		TOPIS evaluation and sorting rank (2006)		Simulation weights evaluation and sorting rank (2005)		Simulation weights evaluation and sorting rank (2006)	
Scientia Agricultura Sinica	67.22	1	72.91	1	76.85	1	80.58	1
Acta Agronomica Sinica	66.41	3	65.73	4	73.99	2	72.12	4
Soil Sinica	67.00	2	69.20	2	71.14	3	73.08	2
PEDOSPHERE	59.71	5	61.09	6	69.76	4	72.25	3

Journal of Soil and Water onservation	61.98	4	65.99	3	69.51	5	71.36	5
Transactions of Agricultural Engineering	58.22	6	64.27	5	65.41	6	70.74	6
Journal of Northwest Sci-Tech University of Agriculture and Forestry	54.59	8	53.95	12	62.67	7	61.58	11
Journal of Hunan Agricultural University	53.27	10	51.93	16	61.72	8	59.11	16
Soil	56.45	7	54.54	11	60.85	9	59.38	15
Chinese Agricultural Science Bulletin	51.82	13	55.00	10	60.45	10	65.24	7
Plant Nutrition and Fertilizer Science	52.83	12	59.61	7	59.05	11	65.12	8
Acta Agriculturae Universitatis Jiangxiensis	51.14	15	53.40	13	59.02	12	62.32	10
Journal of Jilin Agricultural University	50.77	17	48.80	21	58.97	13	55.55	18
Acta Phytopathologica Sinica	54.36	9	55.73	9	58.86	14	60.73	13
Chinese Journal of Rice Science	53.14	11	58.56	8	58.31	15	64.62	9
Journal of Fujian Agriculture and Forestry University	51.09	16	44.93	44	58.22	16	49.42	44
Journal of Nanjing Agricultural University	51.32	14	49.59	19	57.80	17	55.11	21
Chinese Journal of Eco-Agriclture	49.72	20	52.24	15	56.90	18	61.18	12
Journal of Huazhong Agricultural University	49.97	19	48.28	23	56.83	19	53.98	24
Agricultural Research in the Arid Area	50.14	18	49.04	20	56.75	20	55.53	19
Journal of Zhejiang University(Agriculture and Life science)	49.09	21	46.31	30	55.41	21	51.25	32
System Science and Comprehensive Studies in Agriculture	47.40	25	47.53	24	54.28	22	54.32	23
Journal of Yangzhou University (Agricultural and life science edition)	47.81	23	45.27	38	53.98	23	50.36	37
Journal of Southwest Agricultural University	47.66	24	45.13	40	53.59	24	50.41	36
Journal of Agricultural Biotechnology	47.13	27	46.26	32	53.41	25	52.17	28
Xinjiang Agricultural Science	46.67	28	50.89	17	53.40	26	58.88	17
Journal of Henan Agricultural University	47.19	26	41.05	58	53.03	27	45.34	58
Chinese Journal of Plant Pathology	48.09	22	45.54	36	52.87	28	49.29	46
Jiangsu Agricultural Sciences	46.12	31	47.09	25	52.22	29	53.39	25
Journal of Agricultural University of Hebei	46.14	30	45.05	42	51.68	30	49.82	43

According to the above analysis, if you want to examine whether a journal's quality has raised or fallen in 2006, the 2006 TOPSIS evaluation rank cannot be compared with the 2005 TOPSIS evaluation rank, and the only comparison we can make is over the evaluation results of the 2005 and 2006 journals with the same simulation weights based on 2005 data (total cites 0.185, impact factor 0.161, share of grant-supported papers in total papers 0.226, cited half-life 0.246, and disciplinary impact 0.182) .

For example, let us evaluate the "Journal of Soil and Water Conservation" with TOPSIS in two years. Its rank rises from No. 4 in 2005 to No. 3 in 2006. However, this result is misleading, because the results of the two years are not comparable. According to the evaluation results by use of the simulation weights, the journal has been ranked No. 5 for both two years, which is the true comparable result.

As another example, let's evaluate the "Agricultural Research in the Arid Area" again, through adopting TOPSIS in two years. Its rank falls from No. 18 in 2005 to No. 20 in 2006. However, it gives a false impression again. According to the comparable ranking results by use of the simulation weights, the rank rises from No. 20 in 2005 to No. 19 in 2006, which is the true result.

It should be noted that, because the simulation weights is merely a simulation over TOPSIS results after all, there are some differences between the TOPSIS evaluation results and simulation weights evaluation results of 2005, which is normal .

5. Conclusion and discussion

5.1Comparability of journal evaluation methods should be taken seriously

We should ensure the fairness of journals evaluation as much as possible. Many factors affect the evaluation fairness, such as the indicators selection, evaluation methods choice, weights determination, and the design of evaluation task, etc. Because almost all the non-linear evaluation methods and most linear evaluation methods are not comparable, this hidden problem should arouse enough attention in the academic community.

5.2The choice of evaluation methods for academic journals is closely related to the evaluation purpose

If the main evaluation purpose is to sort out various journals , then the comparability and inheritance of the evaluation results should be noted closely; if the main evaluation purpose is to analyze the factors impinging on the journal quality, which is a macro-level application of journal evaluation, however, comparability will not be a big problem. Though the different evaluation results of different evaluation methods are not comparable, they are often of high correlation, so the incomparability will not affect the macro-analysis.

5.3It may be necessary for the organizations conducting journal evaluation to release the evaluation methods

When the professional organizations conduct annual journal evaluations, it is normal for them to change evaluation methods with the improvements of evaluation techniques, but once the evaluation method is changed, the evaluation results of different years will not be comparable. If such organizations do not release the evaluation method, they will give the public a false impression that the evaluation results of different years are comparable, and reach possibly wrong conclusion that the rank of a journal has raised or fallen. How should one balance the desire to try advanced evaluation technology with the comparability of the evaluation objects? The best way is to release the raw data, evaluation methods and evaluation results annually, and make renewed evaluations if necessary. That is, if we adopt A evaluation method last year and B evaluation method this year, it is necessary not only to give the evaluation result of this year by method B but also to give the evaluation result based on the last year's data by B method.

5.4 Only the purely subjective evaluation methods are comparable

Because of unavoidable subjective views of evaluation experts, the application of subjective evaluation methods has been disputed all along. According to this study, however, a very interesting point is found that only the purely subjective evaluation methods ,such as AHP and Expert Panels Evaluation, are

comparable because they do not rely on data, At least it is certain that a complete denial of subjective human judgments and the experts' advice is not desirable.

Acknowledgement

The authors acknowledge the support from the grants provided by The National Social Science Foundation of China(Grant No. 10FTQ003).

References

[1] Chunyan Lin、Lin mo. Attribute Mathematical Model for Comprehensive Index System' Quality Evaluation for Natural Science Academic Journals [J]. *Mathematics in Practics and Theory*(in chinese)，2004（5）:1-7

[2] E.Garfield, Citation Indexes in Sociological and Historical Research [J]. *American Documentation* 1963(14)：289-291.

[3] Hanzhong Chen. Application of Principal Component Analysis in Evaluation of Sci-tech Periodicals [J]. *Chinese Journal of Scientific and Technical Periodicals*(in chinese)，2004（6）：658-660

[4] Hirsch,J.E. An index to qualify an individual's scientific research output[M]. *Proceeding of the national academy of sciences USA*. 2005,102:16569-16572

[5] Jingan Pang, Yuhua Zhang etc. Study on comprehensive evaluation index system of Chinese academic journals [J]. *Chinese journal of scientific and technical periodicals*(in chinese)，2000（11）：217-219

[6] Junping Qiu, Rong Zhang etc. Study on academic journal evaluation index system and quantitative method [J]. *New Technology of Library and Information Service*(in chinese)，2004（7）：23-26

[7] Kaiyang Li, Yuping Jia. Fuzzy comprehensive evaluation for full-text periodical databasesbased on AHP[J]. *Information Science*(in chinese)，2005（11）：1688-1703

[8] Markpin, T., Boonradsamee, B., Ruksinsut, K., Yochai, W., Premkamolnetr, N., Ratchatahirun, P. and Sombatsompop, N. (2008). Article-count impact factor of materials science journals in SCI database. *Scientometrics*, 75(2), 251–261

[9] Moed. H. F.,R.E. De Bruin. New bibliometric tools for the assessment of national research performance[J]. *Scientometrics*,1995,33:381-422

[10] Weiping Yue、Concepcion S. Wilson. Measuring the citation impact of research journals in clinical neurology: a structural equation modeling analysis[J]. *Scientometrics*,2004（3）：317-334

[11] Wang Jiu、Tianhe Xu. Comprehensive Evaluation on Medical Periodicals Quality with Weighted RSR Method[J]. *Journal of Mathematical Medicine*, 2003（3）：266-267

[12] Xiaowei Wang, Bo Yang etc. Secondary relative quality evaluation for sci-tech periodical publications [J]. *Acta Editologica*(in chinese)，2003（6）：231-232

[14] Xinlin Su. Constructing the evaluation system for the academic periodicals of humanities & social sciences [J]. *Dongyue Tribune*(in Chinese)，2008（1）：35-42

[15] Xiujie Li, Jingwu Cheng. Applying discriminant analysis to build index system of periodicals estimate [J]. *The Journal of the Library Science in Jiangxi*(in chinese)，2006（3）：48-50

[16] Zifeng Li. A new method to evaluate Core Periodicals based on TOPSIS[J]. *Journal of Information*(in Chinese)，2003（1）：38-42