Research on Error Issues in Indicator Selection in Multi-Attribute Academic Evaluation*

Yu Liping^{1,2}

2026-07-24

ChinaRxiv

Abstract

Multi-attribute academic evaluation is a type of method for comprehensive evaluation based on an indicator system, but the issue of errors in indicator selection has long lacked systematic research. On the basis of theoretical analysis, this paper takes the evaluation of academic journals as an example to analyze the mechanisms by which systematic errors arise, finding that the selection of a single indicator involves both systematic errors and random errors; it also analyzes the systematic errors existing in academic evaluation from the perspective of the indicator system. Overall, errors in the selection of academic evaluation indicators are dominated by systematic errors, while random errors account for only a small proportion. On this basis, methods for reducing systematic errors in indicator selection are proposed. The study concludes that the mechanisms generating errors in the selection of academic evaluation indicators are diverse; attention should be paid to the issue of systematic errors in indicator selection; the main problem with academic evaluation indicators is not an excessive number of indicators, but the consistency of the time spans of evaluation indicators; and eliminating systematic errors in the selection of academic evaluation indicators is a systematic undertaking.

Full Text

【Theory ·Exploration】

Research on Error Issues in Indicator Selection in Multi-Attribute Academic Evaluation*

Yu Liping1,2

1. School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, 310018

2. Collaborative Innovation Center for Statistical Data Engineering Technology and Application, Zhejiang Gongshang University, Hangzhou, 310018

Abstract Multi-attribute academic evaluation is a type of comprehensive evaluation method based on an indicator system, yet the issue of errors in indicator selection has long lacked systematic study. On the basis of theoretical analysis, this paper takes academic journal evaluation as an example to analyze the mechanism by which systematic errors arise. It finds that the selection of a single indicator involves both systematic error and random error; it also analyzes the systematic errors existing in academic evaluation from the perspective of the indicator system. Overall, errors in the selection of academic evaluation indicators are dominated by systematic errors, while random errors account for only a small proportion. On this basis, methods for reducing systematic errors in indicator selection are proposed. The study holds that the mechanism pro-

ducing errors in the selection of academic evaluation indicators is pluralistic; attention should be paid to the systematic-error problem in indicator selection. The main problem with academic evaluation indicators is not an excessive number of indicators, but the consistency of the time span of evaluation indicators. Eliminating systematic errors in the selection of academic evaluation indicators is a systems-engineering task.

[**Keywords**] multi-attribute evaluation; academic evaluation; indicator system; systematic error; random error

[**CLC Classification No.**] G302 [**Document Code**] A [**Article No.**] 1003-7845(2026)01-0038-07

[**Citation Format**] Yu Liping. Research on Error Issues in Indicator Selection in Multi-Attribute Academic Evaluation[J]. College & University Library Work, 2026(1):38-44.

Introduction

In academic evaluation, multi-attribute evaluation methods have become a mainstream application. Multi-attribute academic evaluation has gradually come into use with the development of bibliometrics and multivariate statistics. Early evaluation mainly relied on single-indicator evaluation, a typical example being the impact factor; later it gradually transitioned to using an indicator system for multi-attribute evaluation, namely multi-attribute academic evaluation. Its advantage lies in overcoming the insufficiency of information contained in a single evaluation indicator; different evaluation indicators can be selected according to the evaluation purpose, thereby providing richer information. At present, many mainstream evaluations adopt multi-attribute evaluation methods, such as the World Intellectual Property Organization' s Global Innovation Index, the United Kingdom' s QS World University Rankings, and Peking University' s *A Guide to the Core Journals of China*, among others.

The selection of evaluation indicators is an important foundation of multi-attribute academic evaluation. Multi-attribute academic evaluation is a systems-engineering process: beginning with the evaluation purpose, it proceeds through indicator selection, weight setting, data standardization, selection and application of evaluation methods, and finally calculation of evaluation results. It involves many links that are mutually connected. Indicator selection is the first step and bears on the quality of subsequent evaluation. In practical application, disputes over improper selection of evaluation indicators have never ceased, such as the long-standing debate over the use of the impact factor. Therefore, attention must be paid to the selection of evaluation indicators in order to improve the scientificity and credibility of evaluation.

Improper indicator selection inevitably leads to evaluation errors. Because academic evaluation objects are numerous and the issues involved are complex, this paper takes the selection of indicators for academic journal evaluation as an ex-

ample for discussion. Under normal circumstances, errors include systematic errors and random errors. Drawing on error theory in the natural sciences to analyze the selection of multi-attribute academic evaluation indicators in the social sciences is appropriate because the essential logic of the two is the same: both are measurements. This helps in reflecting on existing problems and improving evaluation quality. Here, the following questions need to be answered: What kinds of errors are mainly involved in indicator selection? What is the mechanism by which errors arise? How can errors in indicator selection be effectively reduced? Answers to these questions are of great value: at the theoretical level, they help advance the development of evaluation theory in science of science and information science and promote the application of multi-attribute evaluation theory; in practice, they help clarify issues related to multi-attribute academic evaluation, thereby improving the quality of academic evaluation.

1 Literature Review

Regarding problems existing in journal evaluation indicators and matters related to their selection, Xu Xinjun[1] proposed a five-dimensional path for optimizing journal evaluation indicators, including deepening holistic evaluation indicators, considering the quantity and intensity of important papers, introducing reverse indicators, supplementing timeliness indicators, and enriching webometric indicators. Chen Hongxia[2] pointed out that university performance evaluation indicator systems have problems such as unclear input–output meaning, an emphasis on the quantity of indicators over efficiency, and weak purposiveness of evaluation indicators. Yu Liping et al.[3], starting from the relationships among evaluation agencies, journal publishers, authors, and readers, and taking into account the human manipulation of indicators—

---

* This paper is one of the research outcomes of the National Social Science Fund later-stage funded project "Academic Journal Evaluation—Research on Indicator Innovation and Methods" (Project No.: 21FTQB016) and the Zhejiang Provincial Key-Construction University Advantageous and Characteristic Discipline (Statistics, Zhejiang Gongshang University) funded project "Research on Academic Evaluation Theory and Methods" (Project No.: 2024ZDA01).

Author profile: Yu Liping, Ph.D., professor, doctoral supervisor; research interests: technological economics and science-and-technology evaluation.

Date received: 2025-08-16 **Responsible editor:** Xie Weizhen

conditions, relative independence, and macroscopic and microscopic characteristics.

When the selection of evaluation indicators is discussed from a broader perspective, research has mainly focused on indicator elimination, including two types of studies: the first eliminates indicators of low importance, and the second

eliminates relevant indicators with overlapping information. Studies on eliminating indicators of low importance have produced abundant results. Hammer P. L. et al.[4] screened evaluation indicators using logistic regression equations. Sohn S. Y. et al.[5] used a Weibull distribution model to screen indicators in order to establish a credit evaluation model. Liu Limeng et al.[6] adopted a non-parametric kernel-density Bayesian discriminant method to remove unnecessary indicators. Hua Mingyan et al.[7] used grey theory to conduct statistical analysis of university digital library data and comprehensively screened out key evaluation indicators. Zhao Jinlou et al.[8] first used an expert elimination method to preliminarily screen indicators and reduce the number of indicators, and then used the minimum mean square deviation method to eliminate indicators with extremely small influence on the evaluation. Ni Zhiqing et al.[9] first classified similar indicators and then selected representative indicators. Yang Jianzheng et al.[10], through a comprehensive analysis of reasonable assumptions, statistical regularities, and data reliability, used structural equations to select key indicators.

There have also been abundant practical studies on screening indicators to reduce information overlap among evaluation indicators. To prevent overlap in evaluation-indicator information from distorting comprehensive evaluation results, Chen Honghai[11] proposed an ill-conditioned index cyclic analysis method. Addressing the problem of information overlap among evaluation indicators, Wang Yongzhong et al.[12] conducted Lasso analysis on 24 academic-journal evaluation indicators, used the AIC/BIC criteria to determine model size, and then selected key indicators. Zhang Lijun et al.[13] comprehensively adopted methods such as spatial distribution density, a mean algorithm optimized by clustering centers, and rough sets to reduce evaluation indicators.

Regarding theoretical research on evaluation error, Schwab D. P.[14] pointed out that, in human-resource evaluation, an important criterion of evaluation quality is to minimize various errors as much as possible. Yang Shengwei[15] divided the systematic errors of sample surveys into four major categories, including design, survey, response, and aggregation errors. Fei Yetai[16] defined systematic error relatively early, noting that it is caused by inherent factors, has the characteristic of repeated occurrence, and displays a certain regularity. Zheng Kai[17] analyzed the relationship between measurement precision and error, pointing out that measurement precision is the result of the superposition of random error and systematic error. Wright P. M. et al.[18] empirically found that objectifying human-resource evaluation cannot reduce or eliminate measurement error. Zhang Hua[19] analyzed the sample-selection mechanism of open online surveys, classified non-sampling errors, and proposed improvement suggestions. He Jianfeng et al.[20] proposed a big-data leverage sampling method based on outlier diagnosis, thereby reducing systematic error.

Judging from existing research, studies on the selection of academic-evaluation indicators are widely distributed across related papers; almost all papers involving multi-attribute academic evaluation include a step for indicator selec-

tion. Although some achievements have been made in research on problems existing in evaluation indicators, the overall quantity is small and the research lacks systematicity. There are many studies solely on indicator selection; as a methodological issue, they are distributed across various disciplines and mainly discuss the elimination of unimportant indicators and overlapping indicators. However, relatively few studies have introduced error theory into the field of multi-attribute evaluation. Overall, the following aspects still require further exploration.

First, the importance of selecting academic-evaluation indicators far exceeds that of eliminating indicators. Existing research ideas and frameworks for indicator selection have difficulty solving practical problems in academic evaluation.

Second, improper indicator selection will inevitably generate errors; it is necessary to conduct comprehensive research on the mechanisms by which errors arise, their influence, and measures for reducing them.

Third, it is necessary to rethink the traditional selection principles of reducing computational complexity and indicator redundancy, and to analyze the selection of evaluation indicators from the perspective of technical evaluation and the rapid development of modern information technology.

Because multi-attribute academic evaluation is complex and evaluation methods are numerous, in order to simplify the research, this paper discusses only the issue of errors in indicator selection. Meanwhile, for ease of explanation, academic-journal evaluation is temporarily used as an example.

2 Mechanisms by Which Indicator-Selection Errors Arise

2.1 Classification of Errors in Indicator Selection

Systematic error is error caused by non-random factors inherent in the measurement system. This type of error causes measurement results to deviate consistently in a certain direction, and it cannot be offset through repeated measurements.

Random error is error caused by random, unpredictable accidental factors. This type of error causes measurement results to fluctuate randomly around the true value; in repeated measurements, its influence can be reduced or offset by calculating the mean.

In the selection of academic-evaluation indicators, from the perspective of data collection for a single indicator, both random error and systematic error exist. From the perspective of macroscopic analysis of the indicator system, however, only systematic error exists. In fact, in terms of academic-journal evaluation, many current studies focus on the design, problem analysis, and optimization of single indicators, while research that conducts comprehensive analysis from the perspective of the indicator system remains generally insufficient (see Figure 1).

[Evaluation objective]
↓
[Indicator selection] $\rightarrow$ [Single indicator] $\rightarrow$ [Random error]
$\rightarrow$ [Systematic error]
$\rightarrow$ [Indicator system] $\rightarrow$ [Systematic error]

Figure 1 Classification of errors in indicator selection

2.2 Mechanisms by Which Single-Indicator Errors Arise

A single academic-evaluation indicator first involves random errors in the process of data collection (see Figure 2), including three major aspects. First, human mistakes: for example, when counting the number of papers published in journals, book reviews may be mistakenly classified as academic papers; when stat
–

When calculating citations to core journals, cited journals that have been removed from the catalogue of core journals may be mistakenly included. When calculating the immediacy index, because article data from the end of December of the preceding year are delayed and only posted online in the current year, they may be classified within the calculation scope of the current year' s immediacy index. Second, there may be deviations in understanding—for example, when counting the number of journal editorial staff, whether personnel on long-term leave or part-time editors from outside institutions should be included. Third, questionnaires may be inappropriate—for example, asking experts to score the quality of dozens of journals in a given discipline exceeds the range of experts' manual discriminatory capacity; the questionnaire questions are difficult to answer.

- Single indicator
  - Random error
    - Human mistakes
    - Deviations in understanding
    - Inappropriate questionnaire
  - Systematic error
    - Design defects
    - Sampling error
    - Human manipulation

Figure 2. Mechanism by which errors arise in a single indicator

The systematic errors of a single evaluation indicator mainly involve design defects, sampling error, and human manipulation. First are design defects. At present, there are numerous papers on optimizing journal bibliometric indicators, and their purpose is to address design-defect problems in evaluation indicators. Second is sampling error. For example, when specifically investigating the working status of journal editorial staff, if strict sampling is not carried

out, the representativeness of the sample will be insufficient. Third is human manipulation. For example, the impact factor can be artificially manipulated through self-citation, mutual citation, and other means, thereby distorting the true evaluative value of a journal' s impact factor and introducing systematic error.

2.3 Mechanism by which systematic errors arise in an indicator system

Examining errors in academic evaluation from the perspective of an indicator system, only systematic errors exist. Their mechanisms include a lack of monotonicity in value judgments, inconsistency in the time span of evaluation objects, insufficient systematicity in realizing evaluation purposes, information loss caused by indicator screening, insufficient mutual coverage among similar types of information, objective reasons that restrict data acquisition, and attempts to solve later-stage technical problems at an early stage.

2.3.1 Lack of monotonicity in value judgments

The so-called monotonicity of value judgments for an evaluation indicator means that the larger the indicator, the better it is—or, for a reverse indicator, the smaller the better—or that value-judgment monotonicity exists within a certain range of indicator values, as in an interval indicator. Taking the average number of authors per journal as an example, if it can be verified that the greater the average number of authors, the higher the journal' s influence or quality, then this indicator can of course be selected. In reality, however, because of disciplinary differences as well as individualized differences in research content, research teams, and so forth, a larger average number of authors sometimes does not mean that the journal' s influence and quality are also higher. In such a situation, this indicator cannot be selected. Because this problem is not a problem at the level of collecting data on the average number of authors, but rather a problem of indicator selection, it falls within the category of systematic error.

In addition, the situation is similar for indicators of regional-distribution counts, such as the number of countries in which authors involved in English-language journals are located and the number of provincial-level administrative regions involved in Chinese-language journals. For a certain field within a certain discipline, some research resources are relatively concentrated and distributed across only a small number of provinces nationwide, while others are relatively dispersed and distributed across many provinces nationwide; this produces differences in the number of regions represented by article authors. Because journal articles as a whole tend to emphasize basic research, the breadth of their regional distribution is actually not a sensitive factor: that is, a wider distribution of basic research is not necessarily an advantage, unlike applied research, which is closely linked to economic development. Therefore, the selection of this indica-

tor requires further testing.

2.3.2 Inconsistency in the time span of evaluation objects

In academic evaluation, an implicit premise is that a certain time span is used as the scope for data collection and evaluation of the evaluation object. Because of the inherent characteristics of bibliometrics, the time spans of evaluation objects reported in citation reports are inconsistent. For example, in the *Annual Report on Impact Factors of Chinese Academic Journals (Humanities and Social Sciences, 2024 Edition)*, the citation report year is 2024 and the statistical year is 2023; the evaluation objects of the immediacy index are journals in 2023; the evaluation objects of the impact factor are journals in 2021–2022; the evaluation objects of the Eigenfactor and the five-year impact factor are journals in 2018–2022; and total citation frequency and the h-index evaluate journals from their founding up to 2023 (see Figure 3). From the perspective of time span, these evaluation indicators differ significantly. This is often not easy to perceive when analyzing a single indicator alone, but it becomes immediately clear when analyzed from the perspective of an indicator system.

In practical evaluation applications, the mixed use of evaluation indicators with different time spans is a common phenomenon, such as the mixed use of total citation frequency, impact factor, and immediacy index. Because evaluation is often periodic, this kind of mixing can blur the actual level of academic journals, bias evaluation toward medium- and long-term analysis, and provide insufficient assessment of the latest development of journals, leading to the implicit problem that actual evaluation lacks timeliness.

2.3.3 Insufficient systematicity in realizing evaluation purposes

Every evaluation has a certain evaluation purpose. The evaluation purpose is the guiding idea of the evaluation; it is often relatively concise, while also carrying a certain degree of ambiguity. Sometimes there are also inherent deficiencies. For example, when evaluating the academic quality of journals, the actual situation is that there simply are no indicators for evaluating the quality of academic journals; instead, evaluation relies more on indicators for evaluating the influence of academic journals, even though the influence of academic journals is closely related to the quality of academic journals. As far as evaluation

For the purpose at hand, the insufficient systematic nature of the evaluation-indicator system is mainly reflected in two aspects. First, indicators directly related to the evaluation objective are lacking. For various reasons, evaluation objectives are often difficult to revise. Under such circumstances, although evaluation indicators may not correspond well to the evaluation objective—and may sometimes be related only by correlation—some indicator is still used to represent the evaluation objective; to a certain extent, this is a choice made out of necessity. Second, systematicity is inadequate. For example, quantitative indicators are not closely integrated with quality indicators[21-22]; bibliometric

indicators are not sufficiently integrated with webometric indicators[23-24]; indicators of influence, timeliness, and quality are not closely combined; and the effort to prevent human manipulation of indicators is insufficient. When systematicity is inadequate, if the evaluation indicators themselves contain insufficient information or have defects, the evaluation results will be significantly affected. For instance, in evaluating academic journals in the network era, bibliometric indicators alone should not be used; some webometric indicators should also be selected, such as download volume and the Web immediacy download rate, otherwise it will be difficult to reflect the journal' s level of network dissemination. In addition to normal evaluation, evaluation indicators designed to prevent human manipulation should also be introduced.

Figure 3. Temporal span of the evaluation object

Schematic labels in the figure: vertical axis, "statistical indicators" ; horizontal axis, "year" ; years 2018, 2019, 2020, 2021, 2022, 2023, 2024; "total cited frequency, h-index" ; "immediacy indicators" ; "impact factor" ; "Eigenfactor, 5-year impact factor" ; "statistical year" ; "citation report year."

2.3.4 Indicator screening leads to information loss

At present, many studies focus on the screening of indicators for academic evaluation. Their main purpose is to remove indicators of low importance or overlapping information, and many methods for indicator elimination have therefore been developed, such as factor analysis and rough sets. However, as long as evaluation indicators are removed, information loss will occur, and this can have a considerable impact on the ranking of evaluation results. Academic journal evaluation has functions such as value orientation and journal performance assessment, and many journals pay close attention to evaluation scores and rankings. If the same evaluation method is adopted, journal rankings may change substantially before and after indicator removal. In particular, when rankings fall, this will inevitably cause dissatisfaction among some journals, since it is, after all, caused by human factors. Of course, this is also related to the number of journals in the discipline. When there are relatively few journals in a discipline, removing indicators may have little impact on rankings, because the evaluation indicators have good discriminative power; when there are many journals in a discipline, however, the evaluation scores of adjacent journals often differ only slightly—that is, the discriminative power of the evaluation indicators is poor—so removing indicators has a relatively significant impact on the ranking of evaluation results.

2.3.5 Insufficient mutual coverage of similar information

A characteristic of multi-attribute evaluation is that academic journals are evaluated from multiple perspectives, such as influence, timeliness, and journal quality. For attributes of the same type, as many evaluation indicators as possible should be selected, because each indicator has its own specific connotation and

characteristics. If too few indicators are selected, the information for that attribute or that class of indicators will be incomplete, affecting the robustness of the evaluation information. For example, there are only a limited number of indicators for journal timeliness, and even fewer usable indicators. Taking journal evaluation over the past two years as an example, immediacy indicators cannot be selected because they count data from the current year; cited half-life can be selected, but cited half-life covers data from all years since the journal was founded, and therefore also cannot be selected. Under such circumstances, only one evaluation indicator—citing half-life—can be selected, which is relatively thin; efforts should therefore be made to find more indicators.

2.3.6 Objective reasons restrict data acquisition

For objective reasons, there are certain limitations in acquiring data for academic evaluation indicators, which in turn affects data acquisition and introduces systematic error. First, there is the issue of time constraints in data collection. For certain reasons, some data require a relatively long time to collect, whereas academic evaluation has time requirements; such indicators therefore have to be abandoned, which produces systematic error. Second, there is the issue of data-collection cost. Evaluation projects or evaluation work are subject to resource constraints. If the collection cost of certain indicators is too high and exceeds the limit, they can only be abandoned. Third, there is the issue of evaluation precision. Excessively high evaluation precision also imposes high requirements on evaluation indicators. The relationship between evaluation precision and evaluation indicators must be properly coordinated; when evaluation indicators cannot meet the requirements of evaluation precision, systematic error will also arise.

2.3.7 Attempting to solve later-stage technical problems at an earlier stage

When there are many evaluation indicators, they will affect subsequent weight assignment, data calculation, overlap in indicator information, and other matters; however, all these problems belong

are within the technical category of evaluation. In a period when evaluation technologies were not yet developed, evaluation methods were still at an initial stage, and computer technology was not sufficiently advanced, these issues had to be fully considered, while indicators also had to be appropriately streamlined. Today, however, when information technology and statistical methods are highly developed, there is no need to consider these issues from the perspective of evaluation technology; the primary task is to ensure the completeness and comprehensiveness of the basic evaluation data and information.

3 Paths and Methods for Reducing Errors in Academic Evaluation

3.1 Reducing Random Errors and Systematic Errors in Single Evaluation Indicators

3.1.1 Reducing Random Errors in Single Evaluation Indicators First, strive to reduce human errors. Data statistics and cleaning should be done well; data inspection, verification, and checking should be strengthened; problems should be identified and rectified in a timely manner. The application of big data and artificial intelligence should be strengthened, and modern information technologies should be used to assist in data quality inspection, so as to improve work efficiency.

Second, clarify the connotations of indicators and reduce deviations in understanding. Attention should be paid to the collection of data for evaluation indicators; when necessary, standards and operating manuals for indicator data collection should be issued. At the same time, professional training for personnel collecting evaluation indicators should be strengthened, and spot checks should be conducted on the quality of indicator data collection.

Third, strengthen the scientific design of questionnaires. On the premise of clearly defining the purpose of the survey, basic theoretical research should be strengthened and the questionnaire should be well designed. Experts should be invited to hold demonstration meetings to analyze possible problems and continuously optimize and adjust the questionnaire. Where conditions permit, a round of preliminary investigation may be conducted first, in order to identify possible problems and make corrections.

3.1.2 Reducing Systematic Errors in Single Evaluation Indicators First, strengthen research on evaluation indicators and optimize them. In recent years, bibliometric indicators have been continuously improved and have made substantial progress; many related research publications have emerged, and this momentum deserves affirmation. For every evaluation indicator selected, its connotation should be analyzed in depth, possible problems and deficiencies should be examined, and the latest research findings should be drawn upon to optimize it.

Second, improve the scientific nature of sampling surveys. Sampling survey standards should be strictly followed, and the representativeness of samples should be improved in actual practice.[1] According to different evaluation purposes, simple sampling, stratified sampling, cluster sampling, systematic sampling, and other methods may be adopted, while attention should also be paid to selecting an appropriate sample size.

Third, take measures to prevent the manipulation of indicators. The level of manipulation may be reduced by optimizing relevant indicators. For example,

[1] 25

Yu Liping et al.[2] proposed that, in response to the two important issues of excessive self-citation and artificial manipulation of article volume, evaluation indicators can be optimized through a variable-weight function.

3.2 Paths and Methods for Reducing Systematic Errors in Indicator Systems

3.2.1 Selecting Evaluation Indicators with Monotonicity in Value Judgment The same evaluation indicator has different meanings for different evaluation objects. It is necessary to make a comprehensive judgment on the monotonicity of value judgment of relevant indicators according to the evaluation purpose, and a comprehensive analysis may be conducted from the perspective of comparing micro- and macro-evaluation. For example, when the number of authors is used in the evaluation of papers, it belongs to micro-evaluation; there is no evidence showing that the more authors a single paper has, the better its quality or the higher its impact. However, for the evaluation of academic journals, this belongs to macro-level evaluation. If the greater the average number of authors is, the higher the quality or impact of an academic journal is, and if the results are statistically significant, then this indicator can be used for academic journal evaluation. Specific statistical testing methods include independent-samples t tests, regression analysis, and so on.

3.2.2 Paying Attention to Consistency in the Time Span of Evaluation Objects To address this hidden issue, the best method is to select evaluation indicators with the same time span for evaluation. From a long-term perspective, the h-index, total citation frequency, cited half-life, and so forth may be evaluated together; from a medium-term perspective, indicators such as the 5-year impact factor, Eigenfactor, and article influence score may be evaluated together; from a short-term perspective, indicators such as the impact factor and non-self-citation impact factor may be evaluated together. If the time spans of the indicators are inconsistent, then they need to be modified. For example, when the ratio of funded papers is used together with the impact factor to evaluate the past two years, the ratio of funded papers over the past two years should be calculated, rather than the ratio for the current year; when it is used together with the 5-year impact factor to evaluate the past five years, the ratio of funded papers over the past five years should be recalculated, rather than counting the ratio for the current year; when it is used together with total citation frequency over a long time span, the ratio of funded papers since the founding of the journal should be calculated.

Because of the inherent limitations of bibliometrics, it is very difficult to ensure that all evaluation indicators have consistent time spans. In this situation, a flexible approach is as follows: for long-term evaluation, relevant short-term indicators within five years may be appropriately adopted; when conducting

[2]26

short-term evaluation of journals over the past two years, current-year indicators may be appropriately adopted, but an explanation must be provided, as this is not an ideal state.

3.2.3 Emphasizing First-Level Indicators and Striving for a Comprehensive Evaluation Perspective With the evaluation purpose as the focus, attention should be paid to the comprehensiveness and systematicity of first-level indicators, and full consideration should be given to the balance among factors such as quantity and quality, subjectivity and objectivity, scale and efficiency, traditional indicators and web-based indicators, and evaluation orientation and management tools. These factors are mainly reflected in different first-level indicators; sometimes they are also reflected in lower-level indicators, but they are primarily reflected in higher-level indicators.

3.2.4 Appropriately Using Relevant Indicators in Evaluation In academic evaluation, the issue of the quality of outputs is of vital importance and has not yet been effectively resolved. From the micro perspective, peer review is generally considered the most effective method for evaluating the quality of papers, but it can sometimes fail. From the macro perspective, direct methods for evaluating the academic quality of journals are essentially lacking. At present, indicators related to the academic quality of journals include the h-index, Eigenfactor, impact factor, and so on. A large body of research has confirmed that the impact indicators of some journals are highly correlated with academic quality. Therefore, in order to make up for the inadequacy of macro-level quality evaluation indicators for journals, impact indicators may be appropriately used as substitute indicators for quality evaluation; this is

Measures that are adopted out of necessity in response to quality requirements for evaluation must nevertheless be fully explained in the evaluation system.

3.2.5 Prudently Remove Evaluation Indicators

Evaluation indicators may be removed according to the purpose of the evaluation. If a given evaluation indicator is unrelated to the evaluation purpose, it may be deleted after careful selection and in-depth analysis. However, indicators must not be deleted merely to reduce computational complexity or reduce overlap in indicator information, because doing so will lead to the loss of basic information contained in the evaluation indicators. Moreover, current evaluation methods and information technologies are already capable of overcoming such problems.

3.2.6 Appropriately Select More Indicators Carrying the Same Type of Information

So-called "information robustness" addresses precisely the problem of poor information reliability caused by a single source channel of information. For subdivided indicators such as secondary and tertiary indicators, if possible, more

such indicators should be selected to compensate for the insufficiency of "information robustness" caused by too few indicators of the same type. Selecting more indicators of the same type can better solve this problem.

Selecting more indicators of the same type can also solve the problem of incomplete information in individual indicators and realize mutual supplementation of information among indicators. For example, the impact factor can generally reflect the influence of a journal, but there may be excessive self-citation. If the non-self-citation impact factor is further introduced, it can compensate for the insufficient amount of information contained in the impact factor.

3.2.7 Improve Evaluation Management to Enhance the Quality of Evaluation Indicators

The preparation and reserve of evaluation data constitute an important foundation for ensuring the quality of evaluation indicators. Evaluation is a systems engineering undertaking, and indicator selection is only one aspect of it. Evaluation management should be performed well so as to improve the quality of evaluation-indicator selection and data collection. For example, planning and preparation should be carried out in advance; sufficient time should be reserved for indicator-data collection; comprehensive balance should be maintained; financial support for data collection should be strengthened; and the collection accuracy of necessary indicator data should be improved as much as possible.

3.2.8 After Indicator Selection, Attach Importance to Subsequent Technical Processing

The selection of evaluation indicators and the collection of data are prerequisites for academic evaluation. The number of indicators, data types, data distributions, and other factors will have a substantial impact on subsequent weight assignment, method selection, evaluation calculation, and so forth. At present, with the high development of modern information technology and multivariate statistical methods, relatively good solutions already exist for these problems. Therefore, it is essential to attach importance to the subsequent processing of data and to improve the quality of basic data.

4 Research Conclusions and Reflections

(1) The mechanism by which errors arise in the selection of academic evaluation indicators is multifaceted. In the multi-attribute evaluation of an indicator system, the selection of a single indicator involves both systematic error and random error; from the perspective of the indicator system, systematic error exists. Overall, errors in the selection of academic evaluation indicators are dominated by systematic error, while random error accounts for only a relatively small proportion. Random errors in single indicators include human mistakes, deviations in understanding, inappropriate questionnaires, and so on. Systematic errors in single indicators in-

volve such problems as design defects, sampling errors, and human manipulation. The mechanisms by which systematic errors arise in an indicator system include the absence of monotonicity in value judgment, inconsistency in the time spans of evaluation objects, insufficient systematization in achieving the evaluation purpose, information loss caused by indicator screening, inadequate mutual coverage of information of the same type, restrictions on data acquisition due to objective reasons, and attempts to solve later-stage technical problems at an earlier stage.

(2) Attention should be paid to reducing systematic errors in indicator selection. A large number of practices and cases show that systematic errors in evaluation indicators widely exist in the collection of basic data for academic evaluation and in the process of selecting evaluation indicators. Examples include logical confusion caused by inconsistent time spans among different evaluation indicators, and poor operation of the indicator-system design. Moreover, some practices that remove evaluation indicators or reduce the number of evaluation indicators instead increase the systematic error of academic evaluation. These problems should all be given sufficient attention. On the basis of the framework of this paper, they may be examined and analyzed one by one so as to reduce error problems in the selection of evaluation indicators.

(3) The main problem with academic evaluation indicators is not an excessive number of indicators. The time spans of the evaluation objects of different academic evaluation indicators are inconsistent; in order to solve this important problem, another new problem has arisen, namely that existing bibliometric indicators are, in essence, seriously insufficient. Under such circumstances, the necessity of selecting and removing indicators almost does not exist, and this will have a profound impact on the selection of evaluation indicators. It is recommended that useful indicators not be deleted lightly in academic evaluation.

(4) Eliminating systematic errors in the selection of academic evaluation indicators is a systems engineering task. Comprehensive measures must be adopted, and all-round coordination must be carried out, in order to reduce systematic errors in the selection of academic evaluation indicators. Such measures include selecting evaluation indicators that possess monotonicity in value judgment; paying attention to consistency in the time spans of evaluation objects; emphasizing first-level indicators and striving for comprehensive evaluation perspectives; appropriately adopting relevant indicators to participate in evaluation; prudently removing evaluation indicators; appropriately selecting more indicators carrying the same type of information; improving evaluation management to enhance the quality of evaluation indicators; and attaching importance to subsequent technical processing after indicator selection.

References

[1] Xu Xinjun. A five-dimensional path for optimizing the journal evaluation indicator system—Suggestions for *A Guide to the Core Chinese Journals* [J]. *Chinese Journal of Scientific and Technical Periodicals*, 2018(2): 118-124.

[2] Chen Hongxia. Rethinking the screening of indicators for university performance evaluation [J]. *Technology and Innovation Management*, 2008(3): 255-257.

[3] Yu Liping, Pan Yuntao, Wu Yishan. Research on the selection of academic-journal evaluation indicators based on stakeholder analysis [J]. *Forum on Science and Technology in China*, 2009(8): 20-23.

[4] Hammer P L, Kogan A, Lejeune M A. A logical analysis of banks financial strength ratings [J]. *Expert Systems with Applications*, 2012(9): 7808-7821.

[5] Sohn S Y, Jeon H. Competing risk model for technology credit fund for small and medium-sized enterprises [J]. *Journal of Small Business Management*, 2010(3): 378-394.

[6] Liu Yumeng, Li Zhanjiang, Yin Wei. A credit-indicator screening model based on Bayesian discrimination and cluster analysis under nonparametric conditions [J]. *Statistics & Decision*, 2018(22): 5-10.

[7] Hua Mingyan, Zheng Dejun. Screening of key evaluation indicators for information service capability in university digital libraries [J]. *New Century Library*, 2011(9): 32-35, 70.

[8] Zhao Jinlou, Liu Guoyan. Screening of evaluation indicators in the evaluation system for innovative large enterprises [[unclear: reference continues beyond the visible page]].

···segmented comprehensive algorithm [J]. *Science & Technology Progress and Policy*, 2009(1):124-126.

[9] Ni Yiqing, Lin Ye, Shen Yuelin, et al. A study on indicator selection for the urban innovation index: A case study of Hangzhou [J]. *Science & Technology Progress and Policy*, 2011(6):123-126.

[10] Yang Jianzheng, Xu Jin, Yang Lifan. Research on the screening of key statistical indicators for e-commerce [J]. *Journal of Zhengzhou University (Philosophy and Social Sciences Edition)*, 2009(2):74-76.

[11] Chen Honghai. Research on the screening of evaluation indicators based on cyclic analysis of ill-conditioned indices [J]. *Chinese Journal of Management Science*, 2019(1):184-193.

[12] Jiang Yongzhong, Tao Honglin, Du Yanpu. Application of the LASSO method in the selection of journal evaluation indicators [J]. *Statistics & Decision*, 2018(17):86-88.

[13] Zhang Lijun, Gao Chunxiao. Research on an indicator-screening method based on k-means clustering and rough-set algorithms [J]. *Operations Research and Management Science*, 2020(12):8-12.

[14] Schwab D P. Construct validity in organizational behavior [J]. *Research in Organizational Behavior*, 1980:3-43.

[15] Yang Shengwei. Systematic errors in sample surveys and their causes [J]. *Statistics & Decision*, 2017(20):189.

[16] Fei Yetai. *Error Theory and Data Processing: 7th ed.* [M]. Beijing: China Machine Press, 2017.

[17] Zheng Kai. Analysis of measurement reliability and its evaluation methods [J]. *China Sport Science*, 2007(2):90-93.

[18] Wright P M, Gardner T M, Moynihan L M, et al. Measurement error in research on human resources and firm performance: additional data and suggestions for future research [J]. *Personnel Psychology*, 2001(4):875-901.

[19] Zhang Hua. Non-sampling errors in online surveys: Taking open Web surveys as an example [J]. *Journal of Statistics*, 2022(5):70-82.

[20] He Jianfeng, Zhang Liwei. Big-data leverage sampling methods from the perspective of outlier diagnosis [J]. *Journal of Statistics*, 2022(3):70-85.

[21] Hu Bentian, Xie Kexin. The impact of economic resilience on the high-quality economic development of the Yangtze River Delta and its spatial effects [J]. *Journal of Changzhou University (Social Science Edition)*, 2022(2):54-64.

[22] Yao Zhenyu, Liu Zhiyuan. The impact of modern transportation infrastructure on the quality of economic development in metropolitan areas [J]. *Journal of Changzhou University (Social Science Edition)*, 2022(6):60-69.

[23] Li Shuqing, Yan Zihao, Zhang Zhiwang, et al. Research on the efficiency of knowledge exchange in academic virtual communities based on extended keyword-index analysis [J]. *Scientific Information Research*, 2022(4):12-25.

[24] Wang Lüfen, Dong Xiaoyi, He Jin. Research on the generation and evaluation of scenario-based information service solutions based on multi-source data aggregation [J]. *Scientific Information Research*, 2024(3):56-68.

[25] Li Zhiwei, Mi Zichuan, Li Yi. Adaptive design of dynamic surveys and assessment of data quality [J]. *Journal of Statistics*, 2024(4):67-79.

[26] Yu Liping, Huang Meihua. Network downloads: An alternative solution to the lag of citation-count indicators [J]. *Scientific Information Research*, 2024(1):10-21.

Research on the Error Problem of Indicator Selectionin Multi-attribute Academic Evaluation

Yu Liping[1,2]

1. School of Statistics and Mathematics of Zhejiang Gongshang University, Hangzhou, 310018

2. Collaborative Innovation Center of Statistical Data Engineering Technology & Application of Zhejiang Gongshang University, Hangzhou, 310018

Abstract Multi-attribute academic evaluation is a category of comprehensive assessment methods based on indicator systems, yet the problem of errors in indicator selection has long lacked systematic investigation. This study, based on theoretical analysis and taking academic journal evaluation as an example, analyzes the generation mechanism of systematic errors. It finds that the selection of a single indicator involves both systematic and random errors. From the perspective of the indicator system, academic evaluation also exhibits systematic errors. Overall, errors in the selection of academic evaluation indicators are predominantly systematic, with random errors accounting for only a small proportion. Building on this, methods to reduce systematic errors in indicator selection are proposed. The study concludes that the mechanisms generating errors in academic evaluation indicator selection are multifaceted, and greater emphasis should be placed on addressing systematic errors. The primary issue is not an excess of indicators but rather the inconsistency in their time spans. Eliminating systematic errors in the selection of academic evaluation indicators constitutes a systematic engineering task.

Keywords Multi-attribute evaluation; Academic evaluation; Indicator system; Systematic error; Random error

Note: Figure translations are in progress. See original paper for figures.