



**JME**

*Journal of Measurement and Evaluation*

*JME*, Vol. 1, No. 1, 2026, pp.21-35.

Print ISSN: 3135-4661; Online ISSN: 3135-467X

Journal homepage: <https://www.jmeacta.com>

DOI: <https://doi.org/10.64058/JME.26.1.02>



---

## Two Dimensions of Academic Evaluation: Evaluating Research Quality vs. the Quality of Evaluation Itself

**Liping Yu**

**Abstract:** Academic quality evaluation and academic evaluation quality are two distinct yet interrelated dimensions, and further research is needed on their connotations and relationships. On the basis of defining the relationship between quality evaluation and evaluation quality, this article analyzes their connotations and impact mechanisms, and examines misconceptions in academic quality evaluation from the perspectives of evaluation objects and scales, focusing mainly on representative work evaluation. This paper also examines common pitfalls in how evaluations are managed and executed, including issues in indicator design, weighting, and data handling, as well as in evaluation methods and result processing. Research has found that while academic quality evaluation may be context-dependent, the quality of the evaluation process itself is non-negotiable. Academic quality evaluation and the quality of that evaluation must be organically combined. Small-scale academic quality evaluations often suffer from blind spots. We need to strengthen theoretical research on the quality of academic evaluation.

**Keywords:** Academic Quality Evaluation, Academic Evaluation Quality, Representative Work Evaluation, Multi-attribute Evaluation, Weight Assignment

**Author Biographies:** Liping Yu holds a Ph.D. and is a Professor and Ph.D. Supervisor. His research primarily focuses on technology economics and science and technology evaluation. He is the corresponding author and can be contacted at [yvliping@126.com](mailto:yvliping@126.com).

## **0.Introduction**

Academic evaluation serves as a crucial safeguard for building an innovation-driven nation. Against the backdrop of high-quality economic development, scientific and technological innovation undoubtedly acts as a vital driving force. Whether it involves public science and technology resources or those allocated through market mechanisms, resources must be distributed according to established rules, with academic evaluation functioning as a vital tool and means in this process. Conducting academic evaluations effectively, ensuring fairness and impartiality in outcomes, and enhancing the credibility of evaluation results constitute fundamental principles for academic evaluation work.

From the perspective of evaluation objectives, academic evaluation encompasses two major categories: quantity-based evaluation and quality-based evaluation. Currently, most academic evaluations incorporate both quantitative and qualitative assessments. Since the reform and opening-up, China's science and technology sector has achieved remarkable progress. Early academic evaluations in China emphasized quantity-based metrics. However, as China has risen to become the world's second-largest economy, the nation has increasingly prioritized high-quality economic development. Consequently, the orientation of academic evaluation has gradually shifted from a quantity-centric approach to a quality-centric one.

Academic evaluation relies on specific methodologies, which in turn influence evaluation outcomes. At the micro level, peer review is the primary evaluation method, while macro-level evaluation predominantly employs multi-attribute approaches based on indicator systems. Current socio-economic evaluation demands are primarily focused on macro-level assessment. Significant variations in evaluation results arise due to differences in indicators, weighting, standardization methods, and evaluation techniques, leading to numerous quality issues in academic evaluation.

Academic quality evaluation and the quality of academic evaluation itself represent two crucial dimensions for conducting effective academic assessments. Research in this field holds significant importance. Theoretically, it advances the methodology of scientific and technological evaluation, enriching approaches within information science and the philosophy of science. Practically, it enhances the quality of academic evaluation, ensuring scientific rigor and fairness through methodological improvements, thereby boosting the credibility of academic assessments. This also facilitates relevant departments in optimizing the allocation efficiency of scientific and technological resources.

## **1.Quality Evaluation and Evaluation Quality: Their Relationship**

Quality evaluation, also known as quality measurement, refers to all activities systematically monitoring whether product quality meets standards. Academic quality evaluation, specifically, assesses the quality of scientific research, focusing on evaluating the quality of scientific and technological projects, talent, and research outcomes. Its core lies in innovation, aiming to assess research quality, with research outcomes being the subject of evaluation. Since academic evaluation emphasizes basic research rather than applied research,

the dimensions of its subjects can be analyzed from multiple angles, such as the type of subjects and their scale (Deng Sanhong, 2023; Li Shuqing, 2024). Based on the dimensions of quality and quantity, academic evaluations can be categorized into those focusing on quality, those focusing on quantity, and those adopting a combined approach.

Evaluation quality refers to the overall quality of the evaluation process and the methodologies employed — a process-oriented quality that involves not only management but also methodological and technical aspects. Stufflebeam noted that different evaluations have distinct quality standards (Stufflebeam, 2001). Yu Liping et al. suggest that meta-evaluation methods can improve the quality of academic evaluation. Evaluation work encompasses not only routine management tasks but also the design of evaluation systems, mechanisms, and institutional frameworks (Yu Liping, 2012). Evaluation methods can be classified in multiple ways: by type (e.g., peer review, bibliometric methods, web metrics); by perspective (e.g., single-indicator vs. multi-indicator); or by outcome format (e.g., qualitative vs. quantitative).

The quality of evaluation management is straightforward to comprehend, primarily referring to whether the evaluation design is scientific and reasonable, whether the evaluation process is standardized, and whether the evaluation objectives are clear. The quality of evaluation methods concerns not the principles or technical issues of any single method itself, but rather the appropriateness of the methods adopted in academic evaluation and the correctness of their application. Zhang Weihua posits that the quality of comprehensive evaluation should encompass accuracy, reliability, feasibility, applicability, effectiveness, and rationality (Zhang Weihua, 2004).

The final outcome of academic quality evaluation depends not only on the clarity of evaluation objectives — meaning the evaluation must primarily target quality elements — but also on the quality of the evaluation process and the technical quality of the evaluation methods (Figure 1).

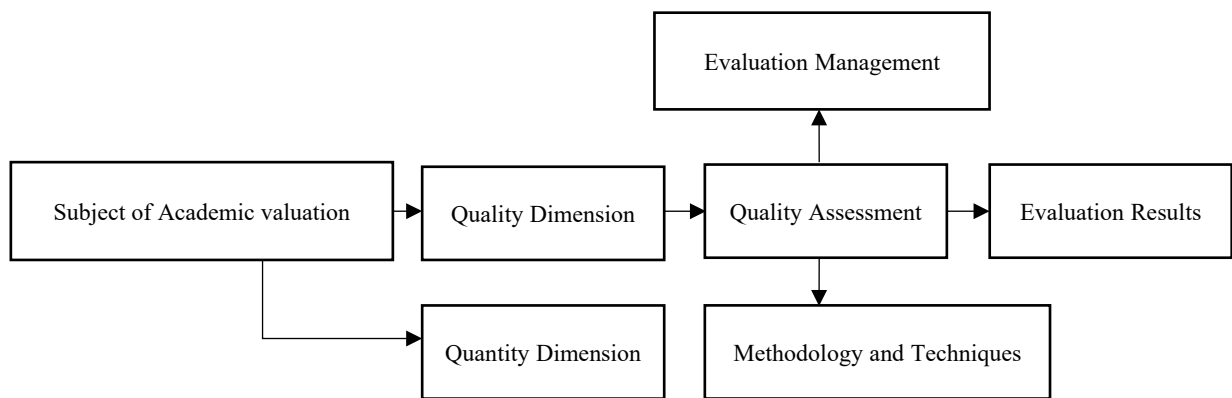


Figure 1: Academic Quality Assessment and Assessment Quality

## 2. Academic Quality Assessment

### 2.1 Quantitative and Qualitative Dimensions of Academic Evaluation Subjects

For academic evaluation, quantity and quality are two crucial dimensions that are embedded in nearly all

evaluation subjects. In other words, this classification system itself possesses broad universality (Figure 2). Academic evaluation subjects encompass project evaluation, talent evaluation, achievement evaluation, policy evaluation, and more. Each of these factors incorporates both quality and quantity elements, such as patent quality/patent quantity (Jiao Haixia, 2019), innovation quality/innovation quantity (Gong Jing, 2023), talent quality/talent quantity (Li Maning, 2023), achievement quality/achievement quantity, and policy quality/policy quantity.

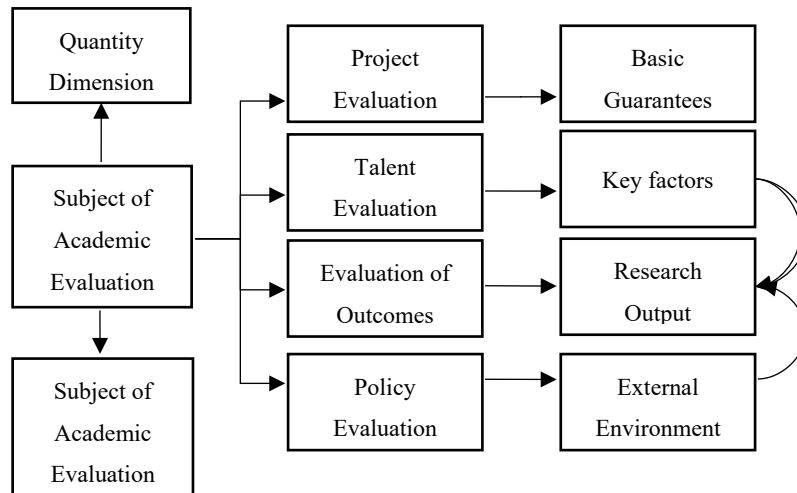


Figure 2: Academic Evaluation Targets

Among these evaluation targets, resources and projects constitute the fundamental safeguards for scientific research, talent represents the most crucial factor for innovation, policies form the external environment, and achievements mark the culmination of research. Therefore, the evaluation of scientific and technological achievements is a vital component of academic assessment and currently the most extensively studied domain.

In summary, academic evaluation is characterized by multifaceted objects and the dual dimensions of quantity and quality. The essence of innovation lies in the process in which quantity generates quality, and quantity and quality mutually transform each other (El, 2007). In evaluating scientific and technological achievements, the most critical issues are academic quality assessment and academic quantity assessment.

## 2.2 Core and Extended Metrics for Evaluating Academic Achievement Quality

Given the significance of academic achievements, this discussion focuses on core and extended metrics for assessing their quality (Figure 3). Core metrics for academic achievement quality refer to key indicators that directly reflect the quality of such achievements, such as academic innovation and the value of the outcomes. The most crucial value, of course, is market value.

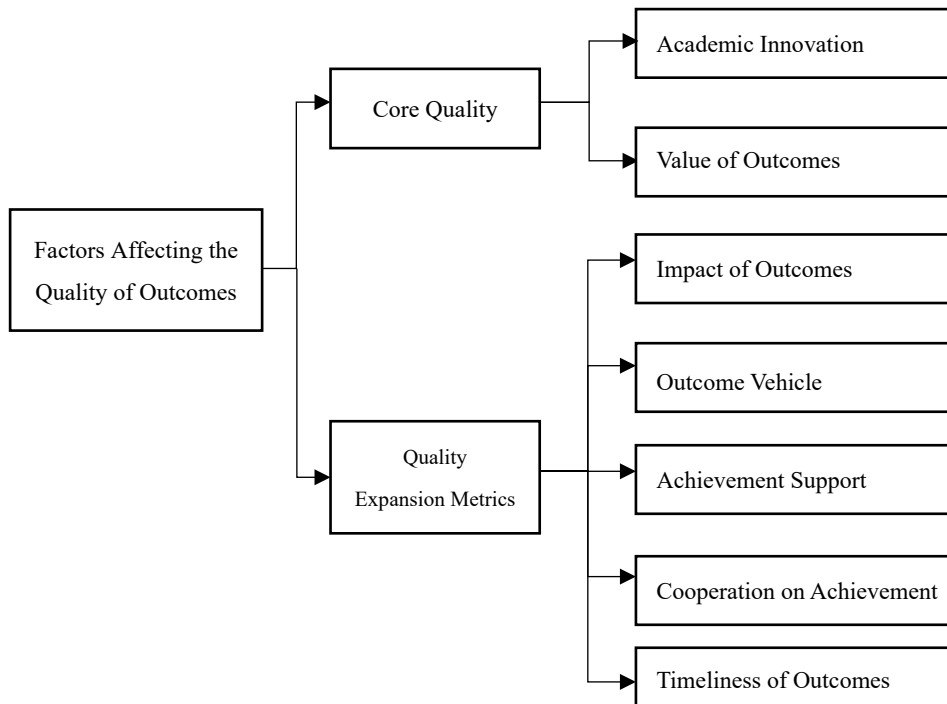


Figure 3: Elements of Academic Achievement Quality Evaluation

Expanded indicators of academic achievement quality refer to metrics that, although not direct quality evaluation measures, can be used for assessment if they exhibit high correlation with core quality indicators. For instance: Impact of achievements: The influence of an achievement is often highly correlated with its quality. - Achievement medium: Factors like journal ranking or tier generally correlate with paper quality — higher-tier journals typically publish higher-quality papers. - Achievement support: Elements such as project level and funding often correlate with quality — achievements from high-level projects tend to be of higher caliber. Collaborative research, encompassing co-innovation and team dynamics, where works produced with high-caliber authors frequently exhibit superior quality; Timeliness of outcomes, reflecting novelty and relevance to current research frontiers.

It is important to clarify that these extended quality indicators only demonstrate statistical significance in large sample sizes. Therefore, when evaluating academic achievements with a limited number of subjects, the use of extended indicators is not recommended.

Chen Song and others argue that academic quality itself is an outcome requiring evaluation, inherently necessitating peer review for scoring (Chen Song, 2022). Supplementary metrics, however, often involve concrete data. If sufficient data can be collected on a certain basis, big data and large models can be leveraged for training and learning. At least at the macro level of academic quality, this theoretically holds the potential to replace peer review.

### 3.The Essence of Academic Evaluation Quality

The essence of academic evaluation quality encompasses the standardization of evaluation organization and

management, the scientific rigor of evaluation methods, the degree to which evaluation objectives are achieved, and the reasonableness of evaluation outcomes (Figure 4).

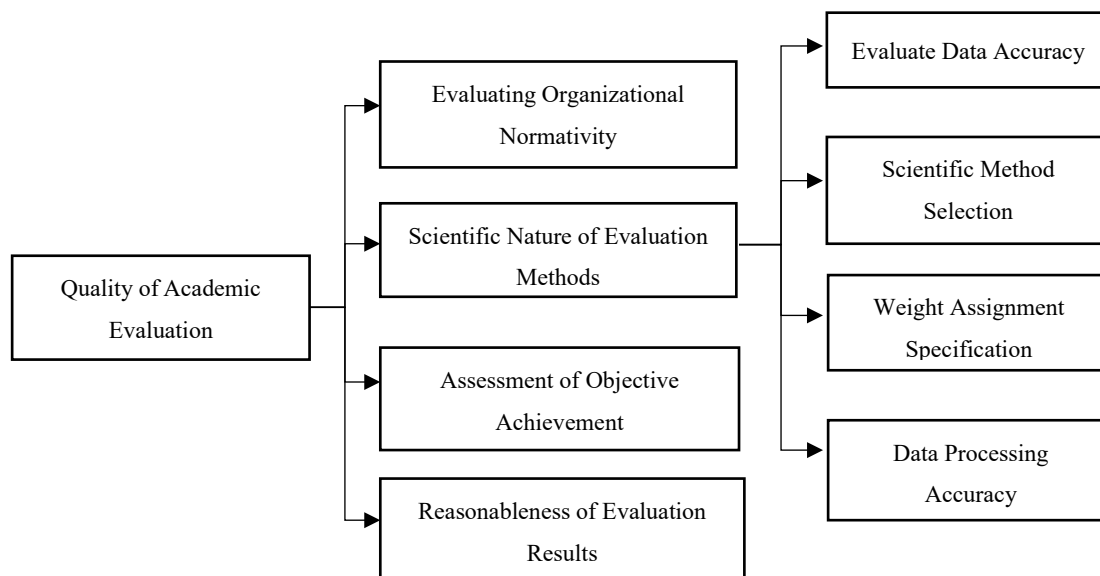


Figure 4: Defining the Quality of Academic Evaluation

#### (1) Evaluating the Standardization of Work Organization

The organization of evaluation work falls under evaluation management issues, primarily involving higher education institutions, research institutes, and similar organizations that serve as the commissioning parties for academic evaluations. While professional evaluation agencies also face organizational management challenges in evaluation work, this is beyond the scope of this paper. When organizing academic evaluations, universities and research institutions generally possess solid experience and well-established procedures for routine, ongoing assessments. However, for non-routine evaluations or those with special requirements — such as discipline evaluations where assessment standards undergo significant adjustments each cycle — slight oversight can lead to management irregularities.

#### (2) Scientific Nature of Evaluation Methods

First, the accuracy of evaluation data. Accurate data serves as the fundamental guarantee for effective academic evaluation. Data accuracy encompasses multiple stages, including standardized sampling surveys, precise indicator definitions, accurate data collection, and reliable secondary data processing. Bruce and Hillmann refer to this as metadata quality, emphasizing the need to minimize random errors in data collection, eliminate systematic errors, and strengthen data auditing to ensure data quality (Bruce, 2004).

Second, the scientific nature of method selection. First, we must establish clear criteria for defining the scientific nature of methods. Second, we should strengthen research on mainstream evaluation methods and absorb the essence of their underlying principles. Currently, relevant scientific and technological evaluations organized by government departments such as the Ministry of Education and the Ministry of Science and Technology generally employ an indicator system where each indicator is assigned a specific weight, followed

by weighted aggregation. This approach has certain rationality and warrants further research. For instance, Ji Yimu et al. found that multi-decision models demonstrate high feasibility in encyclopedia entry evaluation, with the method significantly influencing the evaluation outcome (Ji Yimu, 2021).

Third, standardization of weight assignment. Chu Guangxing and Li Bangyi proposed that quality diagnosis principles for academic evaluation weight assignment can be extended to all multi-attribute academic evaluation domains (Chu Guangxing, 2006). Weight assignment methods can be categorized into subjective weighting, objective weighting, and combined subjective-objective weighting. Academic evaluation serves research management, which inherently reflects administrators' perspectives and objectives — making it fundamentally subjective. This explains why mainstream evaluations by the Ministry of Education and the Ministry of Science and Technology employ subjective weighting. Subjective weighting involves critical details such as expert selection, meeting standardization, and scientific normalization methods, all of which require rigorous quality assurance.

Fourth, precision in data processing. Data processing in academic evaluation encompasses indicator data refinement, standardization methods, evaluation method calculations, and post-processing of evaluation scores. Two key issues warrant emphasis: First, standardization methods yield different results, significantly impacting evaluation outcomes. Therefore, standardized, optimized standardization methods must be adopted. Second, precision in result processing is crucial. For instance, if evaluation results do not follow a normal distribution, should a transformation be applied? If so, which method should be employed?

### (3) Achievement of Evaluation Objectives

Yu Liping and Zeng Huajun argue that the authority of academic evaluation outcomes is influenced by key factors such as the purpose of evaluation, the subject being evaluated, and the evaluators themselves (Yu Liping, 2024). Every evaluation serves specific objectives. Upon completion, both evaluation administrators and evaluators should reflect on whether the existing evaluation has achieved its intended purposes and fulfilled its guiding function. Evaluation objectives are often formulated strategically, carrying profound policy implications. Interpreting these objectives requires a thorough grasp of the underlying policies, which is essential for understanding the evaluation's purpose and direction.

### (4) Reasonableness of the evaluation results

Academic evaluation results are expressed in two ways: scores and rankings. Naturally, where scores exist, rankings follow, though scores are more common. In reality, both the public and experts share a certain consensus regarding evaluation rankings, particularly for top-tier candidates. This stems primarily from the fact that even when evaluation results do not follow a normal distribution, they typically exhibit a “bulge in the middle, fewer at the extremes” pattern. This characteristic allows for clear differentiation among outstanding candidates. In practice, if the top three rankings in an evaluation receive broad acceptance, the results can be considered reasonable.

#### 4. Analysis of Misconceptions in Academic Quality Assessment and the Quality of Academic Evaluation

##### 4.1 Misconceptions in Academic Quality Evaluation: Perspectives on Evaluation Targets and Scale

###### (1) The Relationship Between Academic Quality Assessment and Scale

The evaluation of academic achievement quality can be categorized into three major types based on scale, evaluation workload, and organizational structure: single-paper evaluation, representative work evaluation, and macro-level paper evaluation (Figure 5).

Single-paper evaluation strictly focuses on the essence of innovation quality, employing peer review as its method. Although peer review is influenced by various factors such as the organization of the evaluation process and the expertise of reviewers, it remains recognized as the most effective method for assessing paper quality. In the context of academic journal acceptances, peer review constitutes the sole method.

Representative work evaluation also targets the quality of research papers. Representative works typically constitute a limited set of achievements, such as five or ten papers. Since the initiative of “Abolish the Five Exclusives” (i.e., reforming the evaluation policy by breaking the overemphasis on the five quantitative metrics: papers, titles, degrees, awards, and enrollment), promoting representative work evaluation has been encouraged, which is certainly justified. At its core, peer review should still be employed. When evaluating a small number of subjects, peer review remains the most effective method in a strict sense.

Macro-level paper evaluation transcends pure quality assessment by encompassing both quantitative and qualitative dimensions. When evaluating vast numbers of papers, peer review becomes impractical. Two alternative approaches emerge: First, using journal prestige as a proxy for paper quality — statistically analyzing the journals and publication counts of papers. At the macro level, this proves highly effective. The second involves using an indicator system for evaluation, which synthesizes metrics based on bibliometrics, webometrics, and other indicators. The first method is primarily used for statistical analysis and evaluation of scientific and technological achievements, while the second is mainly applied to evaluate academic journals or research teams. Of course, it can also be used to study the intrinsic relationships between indicators.

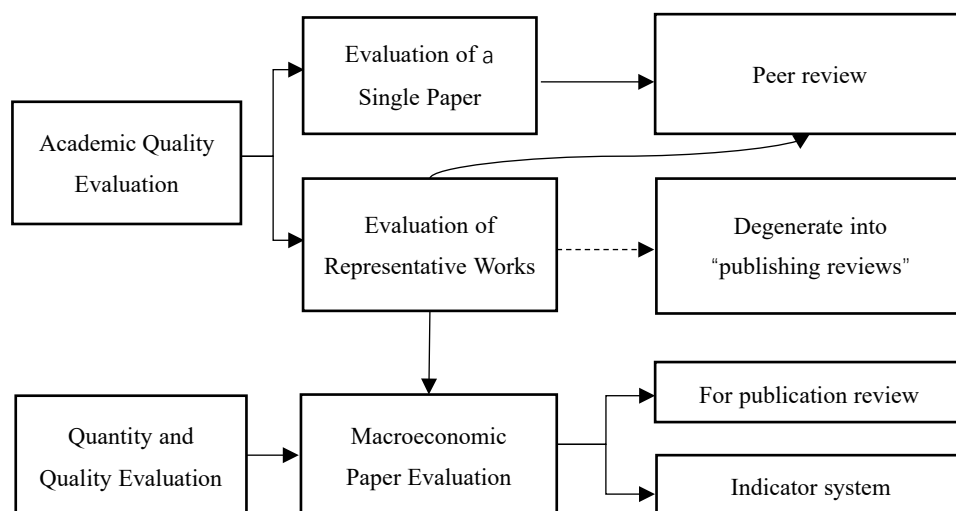


Figure 5: Relationship Between Academic Quality Evaluation and Scale

(2) Common Misconceptions About Evaluating Representative Works

Misconceptions in academic quality evaluation stem from the volume of subjects being assessed. When evaluating a small number of subjects — such as graduate theses, academic papers, or professional title submissions — peer review is entirely feasible. However, when dealing with a large volume of subjects, representative work evaluation faces significant challenges. For simplicity, we use discipline assessment as an example (Figure 6). In the Ministry of Education's fifth round of discipline evaluations, each discipline required 40 representative works. With hundreds of disciplines nationwide, larger disciplines might have tens of thousands of representative works. Peer review becomes fundamentally unfeasible under such circumstances due to logistical constraints: the organizational effort required is prohibitive, the evaluation costs are unsustainable, the timeframe is unrealistic, and the workload on reviewers is overwhelming. Distributing such a vast number of papers to individual experts also lacks meaningful comparability across institutions. In such scenarios, representative work evaluation effectively becomes a macro-level paper assessment, inevitably reverting to the “evaluating papers by their journals” approach.

For discipline evaluation, there are two evaluation methods: one is based on so-called representative works to 'evaluate papers by journals,' and the second method is based on all papers to 'evaluate papers by journals.' Conducting 'evaluation by journals' on all papers does not significantly increase the workload, but when it comes to overall evaluation, it is definitely much more accurate than sampling. For example, when selecting representative works, if University A has 100 top papers, but University B only has 40 top papers, and only 40 representative works can be chosen for evaluation, University A's score may be similar to University B's. However, according to the overall evaluation, University A demonstrates a significantly higher actual level than University B. When conducting large-sample evaluations, abandoning representative work evaluation is the most defensible approach.

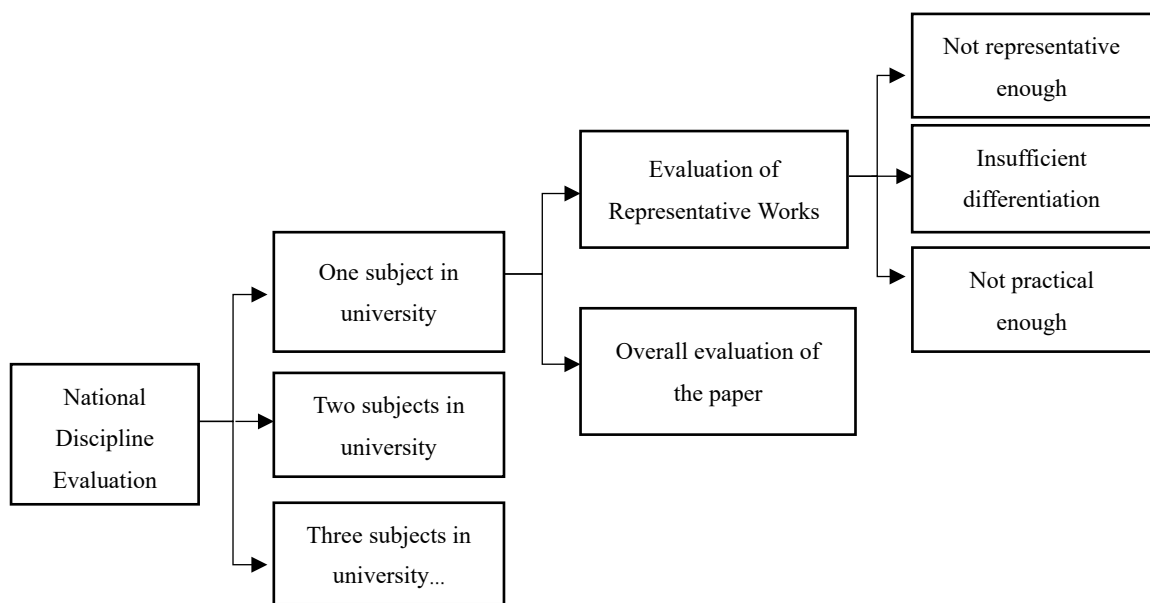


Figure 6: Issues in Subject Evaluation

One misconception about representative work evaluation is that when evaluating a large number of papers, ostensibly promoting representative work evaluation while reverting to quantitative counting. This phenomenon is a typical case of acting under a false pretext, as seen in discipline evaluations.

The second misconception about representative work evaluation is that, in small-sample evaluations, due to limited workload, one may have to adopt an indicator system for evaluation. In these situations, the scarcity of direct evaluation indicators compels the use of more indirectly related ones. In other words, using expanded quality indicators to supplement core quality indicators, such as evaluating representative papers based on their associated funding projects. However, projects are merely a guarantee for research, not innovation itself. This is also inappropriate because expanded quality indicators only make sense in large samples and overall evaluations; they are meaningless for small samples.

The third misconception about representative work evaluation is that its original intention is quality-oriented. Due to the large number of high-level representative works, it inevitably turns into an evaluation of 'quantity and quality,' deviating from the direction of quality evaluation. This requires careful balancing.

#### 4.2 Common Misconceptions About the Quality of Academic Evaluation — From the Perspective of Evaluation Management and Methods

Misconceptions in the quality of academic evaluation include errors in evaluation management, evaluation indicators and data, selection of evaluation methods, weighting scheme, and calculation processing (Figure 7). These categories are delineated as follows.

##### (1) Misconceptions in Evaluation Management

First, there is a lack of systematic and in-depth research on the system and mechanism of evaluation. Chu Guangxing and others believe that insufficient attention has been paid to the quality control mechanism of the scientific and technological evaluation process and its control strategies, which has become one of the reasons for inaccurate evaluation results. Evaluation work is a part of management work. As an evaluation management department or organizer, one should adopt the concept of meta-evaluation and reflect on the evaluation itself, avoiding unnecessary evaluations to reduce the workload at the grassroots level. For regular evaluations, emphasis should be placed on improving standardization, scientific rigor, and fairness.

Second, there is a misguided understanding of policies. Policies serve as guiding principles for evaluation, and a poor understanding of policies can compromise academic evaluation. For example, the fundamental purpose of the "breaking the five-only" policy is to perform quality evaluations and reduce indicators that are not indicative of academic quality or indeed those that simply denote academic quantity. If, in actual academic evaluations, the banner of "breaking the five-only" is raised but the quality of evaluation is still neglected, this is a clear misconception. Furthermore, if objective circumstances make it difficult to conduct effective quality evaluations, this should be clearly stated in the evaluation report; failure to do so is also a misconception.

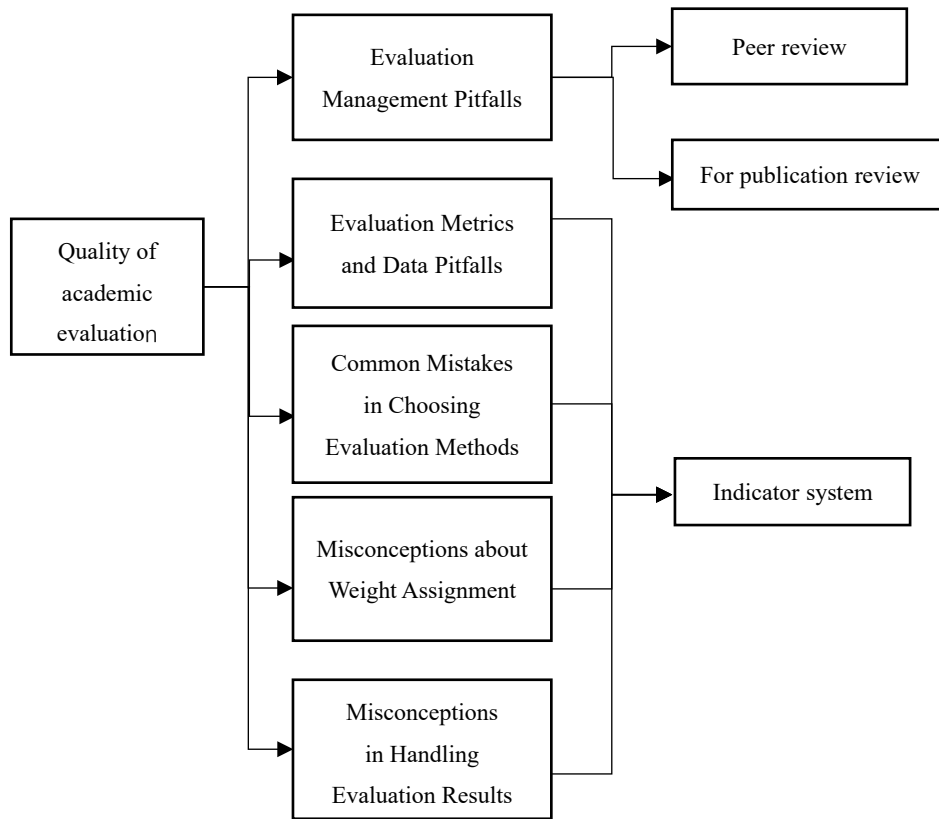


Figure 7: Misconceptions in Academic Quality Assessment

Third, the deployment of the evaluation work is rough. For instance, grassroots government departments might lack evaluation experience and are ill-equipped to undertake evaluation tasks, resulting in overly simplified processing of academic evaluation work and the production of low-quality or even erroneous evaluations. For relevant provincial government departments, the deployment of evaluation work is not simply about issuing and forwarding higher-level documents. Beyond the nationally unified evaluations, local areas should adapt their academic evaluation practices according to local conditions, the characteristics of scientific research, and regional heterogeneity, scientifically planning their academic evaluation work. For example, Jiangsu has a concentration of universities and could develop its own set of standards for evaluating basic research. Guangdong, as a major province in information technology, could also develop its own practices for IT R&D evaluation and the commercialization of results.

(2) Evaluation Metrics and Data Pitfalls

First is the arbitrary selection of evaluation indicators. Academic evaluation is a serious undertaking where indicator selection forms the foundation and must be treated with due rigor. For instance, in journal assessment, no empirical evidence links a higher or lower number of authors to superior paper quality, thus rendering this indicator inappropriate. The same applies to the number of regional distributions; there is no evidence that papers with collaborators from more provinces are of higher quality. Even if this has some statistical significance at a macro level, this indicator should be used cautiously in micro-level evaluations.

Second, the issue of heterogeneous and inconsistent indicators. This refers to inconsistencies in the

evaluation purpose, the objects being evaluated, or the content being evaluated. For instance, using total citation counts to evaluate academic journals can be unfair because some journals have only been published for 5 years, while others have been around for 20 years. Similarly, selecting impact factor and total citation count to evaluate journals is problematic because the impact factor evaluates papers published in the past two years, while total citations consider all papers since the journal's inception, meaning their evaluation targets are inconsistent.

Third, pitfalls in indicator data. Common phenomena mainly include the following aspects: ① Converting raw data into ranked data sacrifices data quality and makes the evaluation coarse. ② Using ranked data instead of raw data is also a mistake. ③ Completely dismissing subjective scoring data is not advisable; as long as subjective scoring is standardized and fair, it can also be an important source of data.

### (3) Misconceptions in Weight Assignment

First, the issue of one-sidedly pursuing objectivity in weight assignment. Weight, by definition, represents importance, a subjective concept. Given that evaluation is also subjective in nature, completely objective weights are debatable. Even if the results of weight assignments from different batches of experts are inconsistent, weights may still converge through standardized expert weight assignment procedures. Overemphasizing the sophistication and complexity of weight assignment methods, or excessively stressing objective weights, can only lead academic evaluation away from the true purpose of research management. For example, if evaluation is conducted using just two indicators — academic quantity and academic quality — under the principle of quality orientation, which assignment method can ensure that the weight of academic quality exceeds that of academic quantity?

Second, the issue of natural weight caused by divergent mean values after standardizing the evaluation indicators. For instance, if evaluation is conducted using only two primary indicators — academic quantity and academic quality — with equal assigned weights, due to data bias, the average for academic quantity might be 80 while the average for academic quality is 40. In actual evaluation, the weight of academic quantity would therefore be greater, and the two would not be equal as intended. This is a hidden problem with significant impact, as it can distort evaluation weights and cause actual weights to deviate severely.

### (4) Misconceptions in Evaluation Methods

First, pursuing the flamboyance and complexity of the method's principles or formula calculations is misguided. A method with more complexity and formulas is not necessarily more scientific. It is critical to carefully evaluate the method's principles, the scenarios in which it is used, and its applications to form a comprehensive judgment.

Second, not all well-established evaluation methods, such as principal component analysis, factor analysis, or TOPSIS, are necessarily scientific. Principal component analysis and factor analysis focus on dimensionality reduction, with their main use being to grasp the key elements in the development of things; evaluation is merely a byproduct. Although TOPSIS has been widely studied, its practical application remains

limited and warrants further investigation. There are also many nonlinear evaluation methods that need further research.

Third, weighted-aggregation evaluation methods are not straightforward. These methods are widely used in academic evaluation and university assessment, employing elementary mathematical principles from the high school curriculum. However, they conceal many issues that significantly affect evaluation results, such as the choice of standardization methods and the bias in indicator data when evaluating a small number of subjects.

#### (5) Misconceptions in Handling Evaluation Results

In common linear academic evaluations, the maximum value method or range method is generally used for standardization. In this way, the evaluation score does not have an absolute meaning. For example, a score of 60 does not necessarily mean that the level has reached the 'passing line.' Many current practical academic evaluations basically take this for granted, which is completely wrong.

### **5. Some Reflections on Academic Quality Assessment and the Quality of Academic Evaluation**

#### 5.1. Academic quality evaluation is optional, but the quality of the academic evaluation is mandatory

Although the importance of academic quality evaluation is increasingly recognized, and evaluations based on academic quality are becoming more common, it is difficult to conduct academic quality evaluations when the number of evaluation subjects is large or when macro-level academic evaluation is involved. In such cases, evaluations tend to focus on academic quantity, or a combination of quality and quantity, becoming the mainstream. Therefore, academic quality evaluation is an evaluative orientation that can coexist with other evaluation orientations.

The quality of academic evaluation primarily focuses on methodology and management aspects. It is crucial for any academic evaluation and mandatory. Essentially, it emphasizes the scientific nature of the evaluation method and the rigor and standardization of evaluation management, which must be stressed in any academic evaluation regardless of its purpose or orientation.

#### 5.2. Quality evaluation and evaluation quality must be organically integrated

Quality evaluation is approached from the perspective of evaluation objectives, while evaluation quality is approached from the perspective of the evaluation process. To conduct objective and accurate quality evaluations, attention must be given to both the evaluation process and procedural quality; both are indispensable. Problems existing in quality evaluation and evaluation quality should be addressed in a targeted manner to ensure both maintain a high level of quality.

#### 5.3. Blind spots in academic quality evaluation on a small scale

As an academic quality evaluation, when dealing with a single evaluation subject or small-scale evaluations, peer review is almost the only method. However, peer review has its limitations and faces numerous issues, with its accuracy improvement being constrained. Conversely, for medium to large-scale academic quality

evaluations, accuracy tends to increase, and evaluation quality may even surpass that of peer review. Therefore, methods for evaluating small-scale academic quality are not suitable for large-scale promotion. For instance, strict representative work evaluations may actually benefit from traditional indicator-based evaluation methods.

#### 5.4. Theoretical research on the quality of academic evaluation needs to be strengthened

From an evaluation methodology perspective, efforts to optimize and innovate individual evaluation methods have never ceased. Additionally, some auxiliary evaluation methods, such as weight assignment methods, have received attention and achieved substantial results. However, at the system and macro levels, theoretical research on methodologies for selecting evaluation methods, assigning weights, selecting indicators, and standardizing indicators is severely lacking, hindering improvements in academic evaluation quality. Ultimately, improving the quality of academic evaluation requires addressing these theoretical issues.

**Funding:** Key Project of Zhejiang Provincial Natural Science Foundation: Research on the Transformation Mechanism of Manufacturing Industry from Quantity-oriented Innovation to Quality-oriented Innovation (Z21G030004).

**Conflict of Interest:** The authors declare no conflict of interest.

#### ORCID

Hu Jiabin<sup>ID</sup><https://orcid.org/0009-0007-0166-908X>

#### References

- Bruce, T. R., & Hillmann, D. I. (2004). *The continuum of metadata quality: Defining, expressing, exploiting*. ALA editions.
- Chen, S., An, F. F., Zhang, M., et al. (2022). Thoughts on improving the peer review mechanism of scientific journals in China. *Acta Editologica*, 34(1), 53–57. DOI: <https://doi.org/10.16811/j.cnki.1001-4314.2022.01.011>
- Chu, G. X., & Li, B. Y. (2006). The quality control mechanism and control strategy of science and technology evaluation process. *Science of Science and Management of S. & T.* (3), 40–43.
- Deng, S. H., Yang, J., Wang, H., et al. (2023). Academic evaluation from the perspective of multi-source heterogeneous data: Connotation, status and prospect. *Scientific Information Research*, 5(4), 42–56. DOI: <https://doi.org/10.19809/j.cnki.kjqbyj.2023.04.004>
- Gong, J., Yin, Z. M., & He, Y. (2023). The impact of service input of manufacturing on the quantity and quality of enterprise innovation — Based on empirical evidence of 2,274 manufacturing listed companies. *Journal of University of International Business and Economics*, (2), 137–156. DOI:

<https://doi.org/10.13509/j.cnki.ib.2023.02.007>

- Ji, Y. M., Xu, Z. Y., Liu, S. D., et al. (2021). Research on evaluation of article quality in internet encyclopedia based on multi-decision model: Case study of Baidu Encyclopedia. *Journal of Information Resources Management*, 11(5), 38–48+59. DOI: <https://doi.org/10.13365/j.jirm.2021.05.038>
- Jiao, H. X. (2019). The study on the relationship between patent quantity and quality: An empirical study based on the Yangtze River Delta. *Technology Intelligence Engineering*, 5(5), 25–36.
- Li, M. N., Wang, S. H., & Hu, Z. L. (2023). Tech talent aggregation and knowledge innovation efficiency — Analysis based on the quality and quantity of talents. *Journal of Statistics*, 4(4), 15–27. DOI: <https://doi.org/10.19820/j.cnki.ISSN2096-7411.2023.04.002>
- Li, S. Q., & Zheng, T. S. (2024). Evaluation of knowledge exchange efficiency in academic virtual community based on the characteristics of author network structure. *Scientific Information Research*, 6(1), 33–48. DOI: <https://doi.org/10.19809/j.cnki.kjqbyj.2024.01.004>
- Rietzschel, F. E., Nijstad, A. B., & Stroebe, W. (2006). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology*, 43(6), 933–946. DOI: <https://doi.org/10.1016/j.jesp.2006.10.014>
- Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183–209. DOI: [https://doi.org/10.1016/S1098-2140\(01\)00127-8](https://doi.org/10.1016/S1098-2140(01)00127-8)
- Yu, L. P., Pan, Y. T., & Wu, Y. S. (2012). An effective method to enhance science and technology evaluation quality: Meta-evaluation. *Studies in Science of Science*, 30(4), 512–516. DOI: <https://doi.org/10.16192/j.cnki.1003-2053.2012.04.005>
- Yu, L. P., & Zeng, H. J. (2024). Research on quality diagnosis of academic evaluation weight assignment. *Information Studies: Theory & Application*, 47(4), 84–89+104. DOI: <https://doi.org/10.16353/j.cnki.1000-7490.2024.04.012>
- Zhang, W. H. (2004). A preliminary study on the comprehensive evaluation of quality issues using multiple indicators. *Statistics & Decision*, (12), 126–127. DOI: <https://doi.org/10.13546/j.cnki.tjyj.2004.12.072>