doi:10.3772/j.issn.2095-915x.2026.04.002

# 多属性学术评价值的误差形成机制及对策研究

开放科学（资源服务）标识码（OSID）

曾喜平[1] 俞立平[2]

1. 上海科学技术职业学院 商贸管理学院 上海 201800;
2. 浙江工商大学 统计与数学学院 杭州 310018

摘要：[ 目的 / 意义 ] 多属性学术评价值的误差影响学术评价质量以及评价结果的运用，从评价结果角度开展的检查和深度分析研究一直较为缺乏。[ 方法 / 过程 ] 将误差理论引入到多属性学术评价值的问题分析中，认为多属性学术评价值主要存在系统误差，在对多属性学术评价利益相关方分析的基础上，探讨了多属性学术评价值的误差产生机制，并提出了相应对策。[ 结果 / 结论 ] 多属性学术评价值系统误差的产生环节包括评价值不合理、只有排序评价结果、评价值绝对优劣价值判断和比较错误、跨年评价值比较错误、评价结果数据分布与偏倚问题等。评价值的系统误差问题既是评价者的问题也是评价结果应用者问题。通过加强学术评价过程控制、采取定值评价方法、利用 sigmoid 函数处理评价指标或评价结果、慎重进行跨期比较等手段可有效降低系统误差。评价机构、政府与学术界在评价质量控制中应发挥好各自作用。对于多属性学术评价值的质量问题需要采取一套标准的范式进行检查。

关键词：学术评价；多属性评价；指标体系；评价结果；系统误差

中图分类号：G302；G35

# Research on Error Checking and Generation Mechanism of Multi-attribute Academic Evaluation Results

ZENG Xiping[1] YU Liping[2]

1. Shanghai College of Science & Technology, Business School, Shanghai 201800, China;
2. School of statistics and mathematics, Zhejiang Gongshang University, Hangzhou 310018, China

**Abstract:** [Objective/Significance] The error of multi-attribute academic evaluation results affects the quality of academic evaluation and the application of evaluation results. Research on inspection and in-depth analysis from the perspective of evaluation results has been lacking. [Methods/ Processes] Introducing the error theory into the problem analysis of multi-attribute academic evaluation results, it is considered that there are mainly systematic errors in the multi-attribute academic evaluation results. Based on the analysis of the multi-attribute academic evaluation stakeholders, this paper explores the error generation

**基金项目** 国家社会科学基金项目“学术期刊评价 —— 指标创新与方法研究”（21FTQB016）。

**作者简介** 曾喜平（1980-），硕士生，副研究员，主要研究方向为教育评价、人力资源管理领域。俞立平（1967-），通信作者，博士，教授，主要研究方向为技术经济、科技评价，E-mail：yvliping@126.com。

**引用格式** 曾喜平 , 俞立平 . 多属性学术评价值的误差形成机制及对策研究 [J]. 情报工程 , 2026, 12(4): 13-35.

doi:10.3772/j.issn.2095-915x.2026.04.002

mechanism of multi-attribute academic evaluation results and put forward corresponding countermeasures. [Results/Conclusions] The study found that the generation of systematic errors in multi-attribute academic evaluation results includes unreasonable evaluation values, only ranking evaluation results, judgment and comparison errors of absolute advantages and disadvantages of evaluation values, errors in comparison of evaluation values across years, data distribution and bias of evaluation results, etc. The systematic error of the evaluation value is not only the problem of the evaluator, but also the problem of the applicator of the evaluation results. Systematic errors can be effectively reduced by strengthening the control of the academic evaluation process, adopting fixed value evaluation methods, using sigmoid function to process evaluation indicators or evaluation results, and prudently making intertemporal comparisons. Evaluation institutions, government and academia should play their respective roles in evaluation quality control. For the quality of multi-attribute academic evaluation values, a set of normative paradigms should be adopted for inspection.

**Keywords:** Academic Evaluation; Multi-Attribute Evaluation; Index System; Evaluation Results; Systematic Error

## 引言

多属性学术评价方法已经成为一种主流的评价手段。学术评价方法包括单指标评价、复合指标评价、指标体系多属性评价三大类，从评价内容的信息丰富度来看，单指标评价信息含量较低，指标体系多属性评价信息含量最高[1-2]。随着国家对学术评价越来越重视，政府在学术评价中往往希望搜集到更多的信息，从而更好地做好科研机构、科研项目、科研人才的管理工作。广大高等院校、科研院所也需要充分利用学术评价成果以加强单位的科研管理，因此多属性学术评价得到了越来越多的重视。从典型的评价实践来看，目前主流的世界大学评价、教育部学科评价、代表性的学术期刊评价、主要的科技人才评价等往往也采用指标体系多属性评价方法。

多属性学术评价值往往存在不同程度的误差。评价误差不同于其他单个评价对象的评估，其影响巨大，涉及到评价的科学性与公平公正。评价结果的误差分析视角是多方面的，从广义上或从评价工作本身讲，涉及到评价指标选取、权重赋值、指标标准化、评价方法等各个环节；从狭义上，误差就局限于评价结果分析与应用环节。实际上，评价机构在学术评价时会严肃认真对待评价的每个环节，做好质量控制，力求减少误差。而从狭义评价结果角度，找到判断误差的框架与方法却是普遍被忽视的问题。比如评价结果并不能进行跨期比较，但在实际分析时进行了跨期比较，单个评价值并不能进行绝对优劣的价值判断，评价结果值数据偏倚严重等等。

多属性学术评价值的误差问题是潜在的，这加剧了其危害性。即使努力做好评价工作每个环节，由于评价方法的复杂性，迄今已有数十种主要的多属性评价方法，每种评价方法原理不同，侧重点不同，评价者很少有熟悉较多方法并且能进行方法比较的，往往选择自己熟悉的并且具有较高科学性的评价方法。从评价结果的应用来看，很少有评价者全面考虑评价结果的综合应用存在的问题，或者应用的可能途径问题，从而反过来对学术评价进行改进。从政府对评价机构的监测管理与评价本身的研

究来看，一个有影响力的学术评价结果一旦公布，需要对其误差问题进行质量检查，而如何系统开展该项工作，尚缺乏有效的技术手段。以上问题必须引起足够的重视，需要进行全面系统的分析。开展相关问题的讨论，在理论上有利于推进学术评价与误差理论，在实践中有利于提供学术评价结果误差分析的方法论与检查范式，从而提高学术评价质量，加强学术评价的科学性与公平公正性。

关于多属性学术评价值的误差问题，现有的研究重点主要集中在与评价相关的前期环节中。比如指标体系构建、数据标准化、评价方法的选择、评价方法的创新等等，更多集中在方法层面。一部分学者从基础理论的角度探讨多属性评价方法自身的问题，另一部分学者主要从应用研究角度开展研究。学术评价是典型的应用研究，或者说应用基础研究，但是该领域的研究却较少关注学术评价结果的误差检查问题，更多注重结果的分析。一个有趣的悖论是，每篇从事具体学术评价的文献都认为自己已经处理好每个环节，评价结果是没有问题的，为什么要分析其误差。

关于宏观计量方法对评价的影响，Martin[3] 提出，只要有数据，就应该补充相应的说明，并且加以必要的理论分析。李爱群等 [4] 认为期刊评价体系由评价理论、方法、指标体系等构成，经典的理论基础、客观的评价指标体系、科学的评价方法才能得出较高可信度的评价结果。王薇 [5] 比较了定量研究方法结果的运用，指出描述统计法简洁明了，质性分析法通俗易懂但缺乏科学性，专业模型法严谨科学但往往联系实际不够，指导性不强。魏登云等 [6] 认为由于多元统计方法的理论性相对复杂，使得应用者往往回避原理简单使用软件，导致结果错误。刘贤龙等 [7] 实证发现采用指标体系确定一类期刊从累计被引数看符合加菲尔德集中定律，比单一指标评价方法更加全面。

关于多属性学术评价方法对评价结果的影响，相对较多的研究是多属性评价方法的选择问题，这是个隐性的问题，潜藏在大量多属性学术评价以及评价方法的基础理论研究中，这里不再赘述。俞立平等 [8] 研究了阿罗的不可能定理对期刊评价结果的影响，发现定值评价无影响，定序评价有较大影响。卢小莉等 [9] 基于单位核心期刊期刊列表数据，运用加权平均百分位（WAP）方法集成期刊等级结果，研究期刊列表的认可度和学科公平性问题。俞欣辰等 [10] 实证发现多属性评价会显著降低期刊评价指标的数据偏倚水平，但不同评价方法的影响不同。俞立平等 [11] 研究了评价结果标准化方法对组合评价的影响，认为对于线性评价没有必要进行标准化，而非线性评价需要进行标准化处理。

关于误差理论对评价的影响，Campbell[12] 定义数据分析方法导致的误差“方法变异”，与计量方法相关，与评价对象结构变异无关。Wright[13] 实证发现单纯的客观评价无法解决人力资源评价的误差问题。郑凯 [14] 认为应采用非参数检验方法进行测量差异分析，处理好偶然误差问题。张华 [15] 分析了网络调查的样本选择中的误差产生机制。贺建风等 [16] 提出大数据杠杆抽样方法，可较好处理异常点，降低大数据调查的系统误差。

doi:10.3772/j.issn.2095-915x.2026.04.002

从现有的研究看，关于学术评价的误差问题，较多关注评价方法层面，目前采用指标体系多属性评价方法已经得到多数学者的认同。关于评价方法之外的一些数据处理方法对评价的影响，也有一些研究涉及。至于系统误差的基础理论与应用研究，也有较多的研究成果。总体上在以下方面有待进一步研究：

第一，学术评价结果是学术评价的最终成果，也需要进行质量检验，需要提供一套分析框架。围绕这个相对较小的环节，分析其误差类型，以及其产生机制，是一项十分有意义的工作。

第二，关于学术评价结果问题分析，现有研究往往仅仅关注结果值和结果排序，较少关注评价结果的深度运用，其中涉及的评价利益相关方众多。一旦出现误差和问题，不能单纯从评价者角度分析，还需要从政府部门、评价对象、公众、评价研究人员等角度分析。

第三，针对学术评价结果的误差问题，提出改进建议。

## 1 学术评价结果系统误差的理论分析

### 1.1 学术评价结果的误差类型

根据误差的定义，在单次测量时，如果评价值与评价结果真实值存在差异，那么存在系统误差；在多次测量时，如果每次测量结果与多次测量的平均值之间存在差异，那么存在偶然误差。对于多属性学术评价而言，在保证评价指标、权重赋值、标准化方法、评价方法一致的情况下，无论测量多少次，评价值总是相同，因此多属性学术评价值主要存在系统误差，涉及评价值与真实值的差异，决定了多属性学术评价的准确度。

需要说明的是，在多属性评价中，真实值其实是不可知的，因为涉及数 10 个评价指标，评价指标的选择、权重赋值、标准化方法、评价方法、结果处理等都有可能产生系统误差，因此本文只能将真实值作为一种理论结果来进行深入分析，即将真实值作为一种研究的假定与前提存在，并不影响分析的逻辑框架与结论的一般性。实际上，这种处理方法在物理测量与统计学中是一种正常现象，比如测量长度、质量，真实值其实也是永远达不到的。

### 1.2 多属性学术评价值与利益相关方

从多属性学术评价利益相关方角度分析评价结果的系统误差是一个全新的视角，如图 1 所示，也能更加全面地加强对问题的深入分析。评价利益相关方主要包括评价者、评价结果应用者两大类。评价结果应用者包括政府部门、公众、评价对象、学术同行、媒体等，其中学术同行是特殊的评价结果应用者。

评价组织者：一般是政府部门、高校、科研院所等，即进行评价的发包方。如果评价组织者自行进行评价，那么评价组织者与评价者可以合二为一。这里需要特别强调一下政府机构评价，对于学术评价而言，其具有较大的公共物品属性，政府对高等院校、科研院所往往有较大公益性的科研投入，代表国家行使主管部门的职能，加强学术评价是天经地义的。需要说明的是，政府一般是评价者组织者，也是评价结果的应用者。

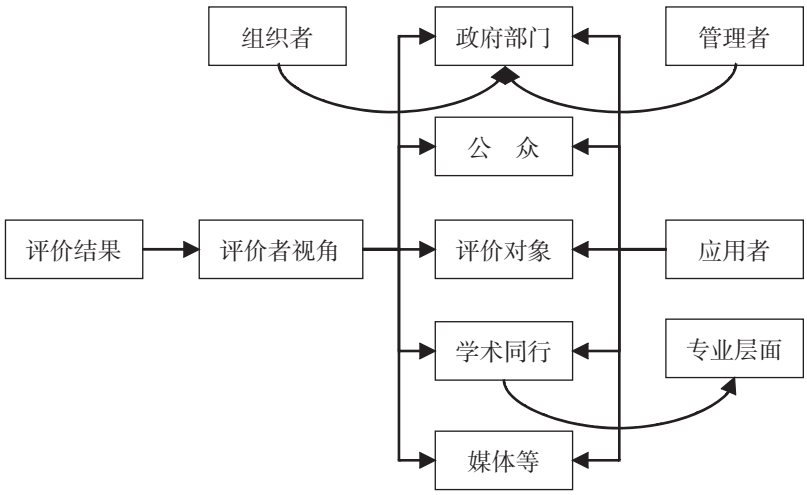


**图 1　多属性学术评价值利益相关方**

评价者：指具体从事评价学术评价工作的机构和专家，这是多属性学术评价的重要核心。评价组织者一般会委托评价者进行具体的评价。评价者一般是高校、科研院所或第三方机构。

政府部门：政府部门是学术评价结果的主要应用者。政府需要通过对学术评价结果的分析，评估政策效果，发现其中的问题，进而采取相关的政策措施。如果采纳评价结果，政府还承担着来自公众、评价对象、媒体等多方的压力，因而会特别注重评价结果的质量。当然，从学术评价管理角度，政府又是学术评价工作本身的管理者，比如“破五唯”问题就是政府倡导的，所以政府具有多重身份。

公众：公众是学术评价结果的关注者。一些影响力较大的学术评价结果，如大学评价、学科评价对公众具有广泛的影响。公众对学术评价结果也具有重要的反馈作用，本质上，公众在学术评价结果的合理性、公正性等评估中也发挥着重要的作用，如果某大学是公认的全国第一，一旦排到全国第二，对评价者也会产生巨大的压力。此外，对于一些异常的学术评价结果，公众也会非常敏感。

评价对象：评价对象是学术评价结果的直接利益相关方。目前各种学术排名众多，涉及大学、科研院所、学科、学术期刊等等。首先，每个评价对象对自己的排名均非常关心，会情不自禁地进行比较对比。其次，那些处于关键位置的学术评价结果也非常敏感，比如位于核心期刊与非核心期刊之间的那些期刊，位于 A 类与 B 类排名之间的那些学科等等。这些评价对象会非常关注评价结果质量，进而会对评价者施加或造成压力。

学术同行：学术同行是学术评价结果的理性分析者。学术同行会站在理论视角，结合实际应用，对学术评价结果进行理性审视，发现其中可能存在的问题，从而使得评价者不断优化评价方法与技术，降低评价误差，提高评价质量。此外，学术同行会不断探索评价理论，优化评价方法和技术，同时提出新的评价方法，为评价者提供各种支撑。在学术评价中，学术

doi:10.3772/j.issn.2095-915x.2026.04.002

同行主要承担评价者的职能。学术同行也有可能成为学术评价成果的特殊应用者，即借助评价结果，从事后续相关研究工作。

公共媒体：包括各种官方与自媒体，也是学术评价结果的重要关注者。甚至一些知名媒体本身，也有可能成为评价对象，典型的有泰晤士报世界大学排名、美国新闻与世界报道大学排名等。媒体为了吸引公众注意，对异常学术评价结果值非常敏感，甚至会进行一些夸大式的宣传，所以对评价者也有一定的压力，并且这种压力往往是难以预见的。

### 1.3 多属性学术评价系统误差的分析框架

多属性学术评价值的系统误差的分析框架如图 2 所示。评价者多数情况下是学术同行，或者是评价专家或相关技术人员。评价组织者主要是政府部门、高校等管理部门。共有 5 种系统误差产生途径，分别是评价值的合理性、只有排序结果的评价值、评价值绝对优劣的判断、评价值的跨年比较、评价值数据分布及偏倚等。这 5 种系统误差的产生途径，均与评价者相关，并且与评价应用者相关，而评价值数据分布与偏倚与特殊的评价应用者 —— 学术同行相关。

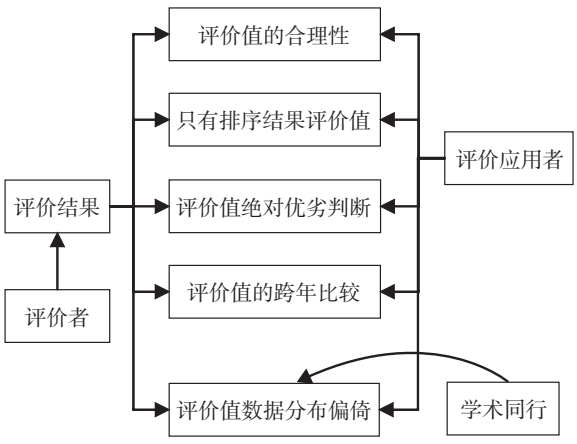


图 2 系统误差分析框架

## 2 多属性学术评价值系统的产生机制

### 2.1 评价值不合理导致系统误差

评价值不合理就是指评价值偏离真实值。评价的真实值问题是个非常复杂的问题，其判断对象包括政府部门、专家、公众、评价对象等。实际上，同一人难以对所有评价对象的真实值加以判断，更多情况是许多人往往仅对其中部分评价对象的真实值能够进行有效的判断，但这些人所有判断的并集一定程度上可以代表真实值，其中存在矛盾的采用大多数即可。目前中国社会科学评价研究院进行的期刊评价，网上征求专家对各种期刊的打分就采取的这种思路，其前提是专家必须客观公正进行打分，同时要适当考虑专家打分期刊的总数量 [17]。

造成学术评价结果值系统误差的原因是多方面的，包括指标选取、权重赋值、指标标准化方法、评价方法等。至于评价管理、评价过程各环节的误差问题，暂不进行深入讨论。如果重点从评价结果值分析的角度研究学术评价的系统误差，类似“反向工程”，从评价结果值的系统误差入手来研究系统误差的产生机制，从而可以更好地将相关问题推向深入，如图 3 所示。可以设想这么一种场景，某种学术评价结果公布后，从什么角度，采用什么方法来分析其系统误差，这是研究重点。

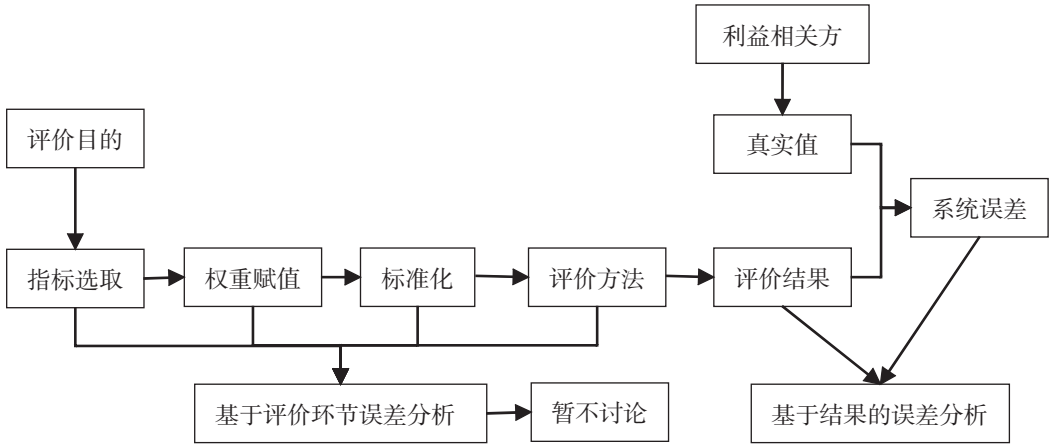


图 3　评价流程的系统误差

实际上目前几乎所有的学术评价研究结果总要面对合理性检查，再好的学术评价，要在合理性上得到认同是非常困难的，但必须得到大多数认同，不一致之处至少要能进行翔实的解释。在这样的背景下，评价的真实值是存在的，或者可以视同存在。如果学术评价结果偏离真实值太多，包括部分评价对象的排序、部分二级指标打分排序、部分问题分析等等，那么就构成了学术评价值的系统误差，其实这种现象屡见不鲜。

## 2.2　只有排序结果评价值的系统误差

在学术评价中，多数评价结果是定值评价，就是评价结果有具体的数值，还有少数评价结果只有评价对象的排序，但没有评价结果值，是非参数的。定值评价结果能够提供完全信息，也包括定序评价信息，而定序评价结果只能提供部分评价信息，其信息是残缺的，如图 4 所示。

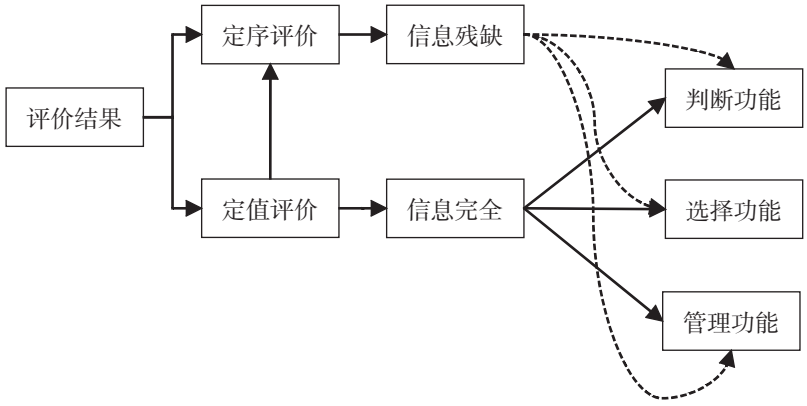


图 4　定序评价结果的系统误差分析

从学术评价的功能看，主要有判断功能、选择功能、管理功能，定值评价完全具备这三大功能，而定序评价只具备部分判断功能、部分选择功能、部分管理功能。

此外由于定序评价的信息不完全性，对其系统误差的判断也会产生影响，缺乏真实值的数据判断基础，即使有排序，评价利益相关方也难以进行深入的比较分析。比如在学术期刊

doi:10.3772/j.issn.2095-915x.2026.04.002

评价中，如果只知道 A 刊大于 B 刊，但不知道得分差距值，是很难为辅助决策提供参考。

基于以上分析，定序评价结果的信息不完全性本身就是一种系统误差，不能单纯从评价方法角度来研究定序评价方法，这是不完备的，一定要考虑其应用。

### 2.3 评价值绝对优劣的价值判断问题

对于大多数学术评价结果而言，都有具体的评价值，共有三种类型：无绝对价值判断、准绝对价值判断、绝对价值判断，如图 5 所示。

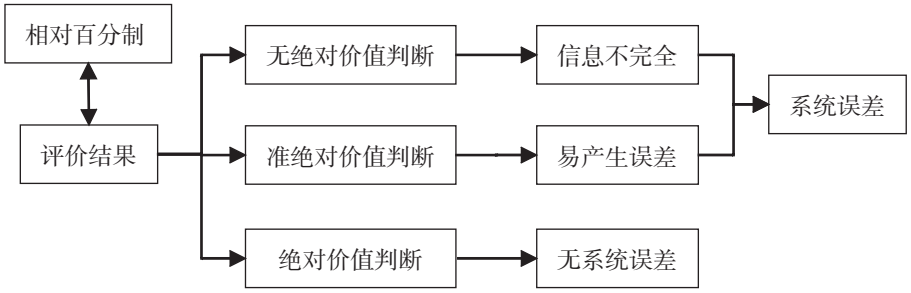


图 5 评价值绝对价值判断的误差分析

第一，评价值无绝对价值判断评价功能。就是指学术评价值越大越好，但是单个评价值缺乏直接判断标准。比如主成分分析或因子分析评价，其评价结果既有正值，也有负值，但具体到某一个评价对象，无法判断该评价对象的好坏，此外评价值为负值，公众也难以理解。其实这是评价值公布的一种信息不完全现象，若具备评价值绝对价值判断评价功能其实更好。

第二，评价值准绝对价值判断评价功能。比如加权 TOPSIS 评价值位于 0~1 之间，乘以 100 可以近似准绝对价值判断；再如加权线性评价，由于标准化值正常在 0~1 之间，因此线性加权评价值也在 0~1 之间，乘以 100 同样可以进行准绝对价值判断，目前许多评价方法均是这种情况，但准绝对价值判断本质上是错误的。

以线性加权汇总为例，标准化方法通常有极大值标准化和极差标准化两种，设 $X$ 为评价指标，$X'$ 为标准化结果，max() 为极大值函数，min() 为极小值函数，两种标准化方法的公式分别为：

$$X' = \frac{X}{\max(X)} \quad (1)$$

$$X' = \frac{\max(X) - X}{\max(X) - \min(X)} \quad (2)$$

这两种标准化方法的共性在于假定极大值为 1，相当于满分是 100。从学术评价真实值的本质出发，极大值并非是满分 100 分。比如学业考试，课程满分标准是有真实值的，每条考试题都有标准答案，某考生考试为满分意味着其答案和标准答案一致，即与真实值完全吻合。在学术评价中，比如期刊影响因子指标，某期刊影响因子为 11.9 达到极大值，但这完全不是百分制中的满分。

如果在学术评价结果分析中，将这种准百分制的结果视同百分制的结果进行绝对价值判断就是错误的。发生这种错误的有些是评价者导致，但也有可能是评价结果的应用者导致。一个有趣的问题是，如果是后者导致，那么评价者是否应该承担责任？实际上，在学术评价越来越发达的今天，必须全方位思考相关问题，

评价者自身还是要承担相关责任的。

第三，评价值绝对价值判断评价功能。实际上，迄今为止几乎所有的评价方法均没有考虑到这个问题，这是多属性评价方法亟待解决的基础问题。

## 2.4 评价值跨年比较错误导致系统误差

在学术评价中，根据评价周期和评价次数，可以分为单次评价与多次评价，单次评价就是同一机构或评价者只进行过一次学术评价，多次评价就是进行过多次评价，如图 6 所示。

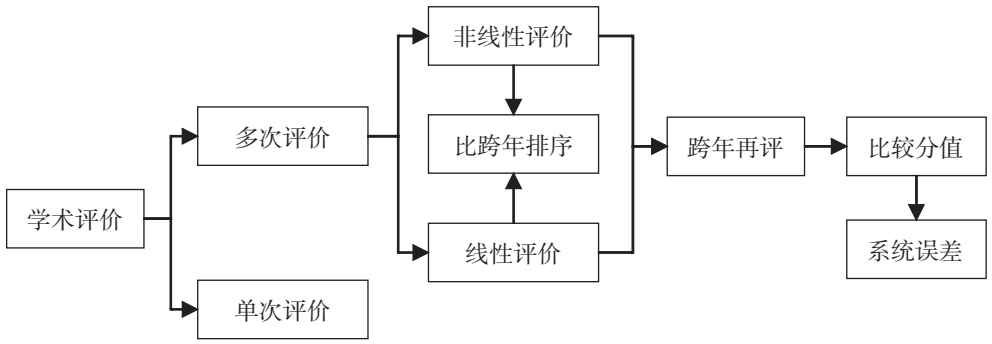


图 6 评价值跨年比较误差分析

评价结果的跨年比较包括两种，第一种是仅仅比较相对排序，比如 A 期刊去年评价结果排序第 5，今年排序第 7，进步了 2 个位次，这是一种非参数比较。如果评价方法一致，进行单个评价对象排序自我比较是可以的，但其他比较就是错误的，会造成系统误差。比如 A 期刊去年从第 5 到今年第 7，B 期刊去年从第 11 到今年第 15，不能说 B 期刊比 A 期刊进步大，因为不同位置的评价值区分度不同，越是靠前的期刊，进步越是困难，何况本质上进步大小不能仅仅看排序，要看同一参照下的分值。

第二种是比较评价结果值，即不同年度之间直接进行评价分值的比较，比如 A 期刊去年评价值为 78.5 分，今年评价值为 81.3 分。目前几乎所有的评价方法，均不可以这样比较，因为每年评价是单独进行的，每年是每年的参照系，跨年不可以比较。在实际学术评价结果的分析中，这种比较司空见惯，本质上也是一种系统误差。

那么有没有办法使得跨年的评价值直接进行比较呢？采用同一方法对跨年数据进行评价就可以了。目前学术评价方法包括线性、非线性两大类评价方法，采用这两类评价方法对跨年数据进行评价，跨年数据就具有可比较的基准，这是一种直接的思路。

对于非线性评价方法，以主成分分析评价为例，比如 2019—2021 三年采用该方法分别评价各年学术期刊，此时跨年期刊的评价值是不可以比较的，为了使得 3 年 3 次评价的评价值可比，可采用主成分分析对 3 年数据混合再次进行一次评价，然后再将评价结果分解到各年，此时进行任意跨年比较就可以了。但这样处理隐含了一个新的问题，就是 3 年混合数据评价的 2021 年的评价结果与2021年单独评价结果排序并不一致，所以这种比较一般用于学术研究。

对于线性评价方法，相对而言较少遇到非

doi:10.3772/j.issn.2095-915x.2026.04.002

线性评价方法的问题，将 3 年数据混合重新评价后不仅跨年数据可比，而且 3 年混合数据评价的 2021 年的评价结果与 2021 年单独评价结果排序基本一致度较高，但仍然有不一致问题，所以还是有系统误差。

### 2.5 评价结果数据分布与偏倚问题

从统计学角度，任何学术评价结果必须满足公众或评价对象等利益相关方的统计学需求，主要包括评价结果的数据分布与偏倚问题，这个问题是隐性的，一般公众并不关注。

在学术评价中，一些评价结果接近正态分布，满足公众预期，呈现“中间多、两头少”的钟形分布特征，如人才评价、大学评价等等。但也有一些评价，如学术期刊评价，结果存在较大的偏倚。在统计分析中，公众、媒体等往往对这个问题并不敏感，可能会犯一些错误。比如简单采用平均值进行分析，夸大数据偏倚异常值的优点或不足等。

这个问题是否属于评价问题。换个角度，如果评价者能考虑到这个问题，并且预先采取一些措施，是可以改善这个问题，只能说随着学术评价的发展，对评价者的要求越来越高了。

## 3 减小评价值系统误差的对策

### 3.1 加强学术评价过程质量控制

公众之所以觉得学术评价值不准确、没有说服力等，本质上是学术评价过程控制没有做好所致，包括指标选取、权重赋值、数据标准化、评价方法选取、结果处理等诸多环节。由于学术评价是一项系统工程，这些环节中任何一个环节处理不好对评价结果都会有一定影响，从而导致系统误差。这些环节中，最重要的环节是评价方法的筛选，其次是权重赋值。

关于评价方法的选择，目前主流的多属性评价方法有几十种，其选取问题一直没有得到有效的解决。从评价实践看，目前一些知名的学术评价，往往采用指标体系加权汇总方法，如泰晤士报世界大学排名、美国新闻与世界报道大学排名、北京大学核心期刊等，而从评价理论研究看，主要涉及非线性评价方法。线性评价方法原理简洁，方法可控，权重实现较好，评价实践应用广泛，建议作为重点予以选用。

关于权重设置，本质上是个哲学层面的问题。权重赋值方法主要分为主观赋权与客观赋权两大类，一些专家反对主观赋权方法，认为其不稳定，人为因素影响较大，不科学，但是从辞海中“评价”的定义看，评价是对人对事物的看法，本来就是主观的，权重的本意就是重要性，也是主观的，客观赋权才是有问题的。至于主观赋权中的问题，完全可以通过规范赋权方法、加强专家交流、改进主观赋权方法等措施得到有效的改善，建议采用主观赋权方法。这里的主观赋权专家主要是管理部门、评价组织者以及邀请的专家。

### 3.2 学术评价中不建议采取只有排序结果的评价方法

诸如 ELECTRE、BORDA 等评价方法，其评价结果侧重排序，不能提供完全信息，在评价中只能发挥部分判断功能、部分选择功能、部分管理功能，严重降低了学术评价的作用，这类方法在学术评价中不宜选用。对于简单评

价目的、宏观评价可用，不宜用在精细、复杂、微观评价中[17]。由于学术评价多属性评价方法众多，解决这个问题应该相对比较容易。

### 3.3 根据情况采用sigmoid函数处理指标或评价结果

sigmoid 函数图像接近成长曲线，符合事物发展规律，其值在 0~1 之间，但永远不可能为 0 或 1，0.6 表示及格，使评价值具有绝对优劣判断标准。采用 sigmoid 函数标准化，不仅解决了评价值无绝对价值判断功能的不足，而且改善了评价数据分布偏倚，发挥了一举两得的作用。

### 3.4 慎重进行评价值的跨期比较

评价结果使用者尤其是评价对象往往喜欢进行跨期比较，除非简单进行同一评价对象的跨年排序比较，其他比较则全部是错误的。

如果实在要进行评价对象的跨年比较，可以采用的方法是采用同一种评价方法对历年数据统一进行评价，然后再跨期比较，但这种方法在解决了跨期可比较问题的同时使得每年的评价结果值又出现新的差异，虽然线性评价方法在一定程度上能解决这个问题，但仍存在缺陷。

在具体的连续性学术评价中，每年的评价往往有少数改动，比如调整个别评价指标，改变标准化方法，适当修改权重组合，甚至改变评价方法，在这种情况下，除非出于评价结果的学术研究需求，否则不宜进行跨期比较。

跨期比较的前提是评价目的和评价需求没有发生根本改变，如果发生根本变化，那跨期比较就是错误的。

综上所述，评价的跨期比较只能用在学术研究层面，比如建立面板数据，研究评价的特征与规律等等，不宜大规模推广，至于评价结果应用的跨期分析几乎都是错误的。

## 4 结论与讨论

### 4.1 多属性学术评价值的误差问题需要引起足够的重视

现有的多属性评价工作往往重视评价各个环节，如指标选取、权重赋值、指标标准化、评价方法、数据处理等方面，但仅仅做好这些工作并不能减少评价结果的系统误差。系统误差的产生环节包括评价值不合理、只有排序评价结果、评价值绝对优劣比较错误、跨年评价值比较错误、评价结果数据分布与偏倚问题等。并且这些系统误差的产生许多与评价环节联系不大，因而容易被忽略，因此，对这些问题必须重视，否则会严重影响多属性学术评价质量。

### 4.2 评价值的系统误差问题既是评价者的问题也是评价结果应用者问题

传统认为，学术评价值的系统误差问题百分之百是评价机构的问题，其实问题并非如此。评价机构固然要对系统误差承担主要责任，但公众误用评价结果也存在责任。类似中国科学院 SCI 期刊分区争论，对评价者而言，该分区主要根据影响因子等划定了不同级别的期刊，但将该期刊分级标准应用到单位科研考核中出现问题，却不是中国科学院的责任。关于学术评价结果值的跨期比较问题、绝对优劣的判断问题等主要就是评价结果应用者的问题，但评

doi:10.3772/j.issn.2095-915x.2026.04.002

价机构也应承担相应责任。评价机构应该了解评价结果值的可能应用，并在评价报告中进行必要的数据处理或说明。

### 4.3 评价机构、政府与学术界在评价质量控制中应发挥好各自作用

评价机构是学术评价的第一责任人，也是学术评价质量的守门人；政府在学术评价中具有多重身份，包括评价者、评价结果应用者、评价管理者等等，主要应发挥评价导向作用，规范评价工作，慎重出台具体的评价标准；学术界承担着学术评价理论、方法与技术的探索、指导、自纠、分析等多重作用，一定要加强学术争鸣。评价机构、政府与学术界各司其职，做好本职工作，是保证学术评价质量的重要基础。

### 4.4 对于多属性学术评价值的质量问题需要采取一套规范的范式进行检查

本文提供了一套多属性学术评价值系统误差检查与分析的范式，只要进行相关环节的检查，并采取必要的措施，就可以大幅度降低多属性评价结果的系统误差。主要的措施包括加强学术评价过程的质量控制尤其是评价方法选择与权重赋值，摈弃只有排序结果的评价方法，必要时采用 sigmoid 函数对评价指标或评价结果值进行标准化，以及慎重进行评价值的跨期比较等等。

## 参 考 文 献


[1] 许安明 , 彭添焕 . 社会科学研究的大数据赋能方式探索 [J]. 科技情报研究 , 2025, 7(2): 133-142.

[2] 俞立平 , 黄梅花 . 网络下载量 : 引文量指标滞后的替代解决方案 [J]. 科技情报研究 , 2024, 6(1): 10-21.

[3] MARTIN T. Recent School Effectiveness Counter-critiques: Problems and Possibilities, British Educational and Research Journal, 2001, 27(4): 443-457.

[4] 李爱群 , 黄玉舫 , 邱均平 . 我国学术期刊文献计量评价体系的客观性与评价结果的准确性探讨 [J]. 中国科技期刊研究 , 2009, 20(4): 609-613.

[5] 王薇 . 学校评价结果解释的方法研究：描述、分析与模型 [J]. 教育理论与实践 , 2018, 38(19): 18-22.

[6] 魏登云 , 张文俊 . 体育科学研究中多元统计分析常见问题及对策 [J]. 北京体育大学学报 , 2016, 39(5): 137-144.

[7] 刘贤龙 , 胡国亮 . 综合评价结果的合理性研究 [J]. 统计研究 , 1998(1): 38-40.

[8] 俞立平 , 武夷山 . 阿罗不可能定理对科技评价结果的影响研究 —— 兼谈定序评价与定值评价 [J]. 科技进步与对策 , 2012, 29(3): 123-126.

[9] 卢小莉 , 李晶 , 宋鸿芳 , 吴登生 . 部门期刊列表评价结果的认可度与公平性研究 [J]. 科学与社会 , 2020, 10(1): 18-31.

[10] 俞欣辰 , 潘有能 . 学术期刊评价指标与评价结果的数据分布关系研究 [J]. 情报杂志 , 2015, 34(9): 102-106.

[11] 俞立平 , 潘云涛 , 武夷山 . 科技评价结果标准化方法及对组合评价的影响 [J]. 情报杂志 , 2012, 31(1): 41-46.

[12] CAMPBELL D T, O'CONNELL E J. Methods factors in multitrait-multimethod matrices: multiplicative rather than additive? [J]. Multivariate Behavioral Research, 1967, 2(4): 409-426.

[13] 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, 2010, 54(8): 875-901.

[14] 郑凯 . 测量可靠性及其估价方法分析 [J]. 体育科学 , 2007(2): 90-93.

[15] 张华 . 网络调查中的非抽样误差 —— 以开放式 Web 调查为例 [J]. 统计学报 , 2022, 3(5): 70-82.

[16] 贺建风 , 张莉维 . 异常点诊断视角下的大数据杠杆抽样方法 [J]. 统计学报 , 2022, 3(3): 70-85.

[17] 苏金燕 . 开放同行评议的接受度与评审质量 —— 基于中国人文社科领域的实证证据 [J]. 世界社会科学 , 2025(5): 178-193.

[18] 叶继元 . 学术 " 全评价 " 体系论 [M]. 北京 : 社会科学文献出版社 , 2022.


（责任编辑：李昶璇）