

基于学生体验视角与证据理论的体育教学评价研究

彭玉林¹, 张诗涵¹, 李彬彬², 程亚飞³

(1.河南财经政法大学 体育学院, 河南 郑州 450046; 2.河南农业大学 体育学院, 河南 郑州 450046;
3.郑州师范学院 体育学院, 河南 郑州 450044)

摘 要: 基于证据理论方法构建体育教学综合评价模型, 建立新的识别框架和信度结构, 利用投影方法和 SWA 算法对评价意见进行集成, 完成对体育教学评价中定性指标的准确测评。研究表明, 以学生体验视角构建的评价模型对体育教学评价指标给出客观评价意见, 还得到定性指标的区间概率分布测度结果, 利用信度结构方法解决对于定性指标定义模糊粗糙的问题, 通过投影转换和 SWA 方法的集成方式以具体数据的形式体现, 完整刻画学生评价者对于定性指标偏好设定的精确程度。研究认为, 基于证据理论方法的体育教学评价模型更加符合学生在评教中的思维认知过程, 充分体现以人为本的学生视角评价思想, 评价结果保证最终结果兼备科学性和可信性, 有利于评价模型在反馈、督导和评优等各项体育教学管理工作中的普遍适用。

关 键 词: 学校体育; 学生评教; 证据理论; 体育教学评价

中图分类号: G807 **文献标志码:** A **文章编号:** 1006-7116(2025)06-0139-08

A study on physical education teaching evaluation based on the students' experience perspective and the evidence theory

PENG Yulin¹, ZHANG Shihan¹, LI Binbin², CHENG Yafei³

(1.School of Physical Education, Henan University of Economics and Law, Zhengzhou 450046, China;
2.School of Physical Education, Henan Agricultural University, Zhengzhou 450046, China;
3.School of Physical Education, Zhengzhou Normal University, Zhengzhou 450044, China)

Abstract: A new identification framework and reliability structure and a comprehensive evaluation model of physical education teaching are constructed based on the evidence theory approach. The integration of evaluation opinions has been achieved using the projection methods and the SWA algorithm, and then completing accurate assessment of qualitative indicators in physical education teaching evaluation. The results show that the evaluation model constructed by the students' experience perspective provides objective assessment opinions in physical education teaching evaluation indicators, and also obtains assessment results of interval probability distribution with the qualitative indicators. The vague and rough issue of definition with qualitative indicators has been solved by the methods of reliability structure. By the integration method of project transformation and SWA and presented by the form with specific data, completely portrays the accurate degree on preference settings of qualitative indicators by student evaluators. The conclusion holds that the physical education teaching evaluation model based on the evidence theory approach would aligns more in line with students' cognitive process of thinking in teaching evaluation, fully embodying a student-centred evaluative perspective the prioritizes human beings, and evaluation results ensuring the final results are both scientific and credible features, which is conducive to prevalent application of evaluation model in feedback, supervision, and teachers' annual evaluation in various aspects of the physical education teaching management missions.

Keywords: school physical education; students' evaluation on teaching; evidence theory; physical education teaching evaluation

收稿日期: 2025-01-09

基金项目: 河南省 2024 年度教育课程改革研究重点项目(2024-JSJY-ZD-025); 河南省高等教育教学改革研究与实践项目(2021-SJGLX420); 河南省高等教育教学改革研究与实践(研究生教育类)项目(2023-SJGLX321Y)。

作者简介: 彭玉林(1982-), 男, 副教授, 博士, 硕士生导师, 研究方向: 学校体育, 儿童青少年体质健康。E-mail: pengyulin@126.com
通信作者: 程亚飞

中共中央国务院印发《深化新时代教育评价改革总体方案》中提出“强化体育评价”的政策导向,并明确教育评价改革要坚持科学有效的要求,强化过程评价、健全综合评价,提高教育评价的科学性、专业性、客观性。体育教学评价是衡量体育课程教学水平 and 效果的有效手段,也是改善体育教学质量和提升学生体质健康水平的重要方法。当前,我国体育教学评价体系在实践中面临多重挑战:一方面,虽然学生在体育教学活动中居于主体地位,但是在评价实践中,往往过度依赖量化指标,忽视学生在课堂参与、兴趣培养、情感体验等定性维度的表现,主观评价又缺乏系统性的证据支撑,导致评价结果的客观性不足。另外,受限于认知能力、表达水平以及传统评价话语权的限制,学生对教学的体验往往不能准确反映评价意见,造成最终评价结果的科学性不足。本研究重点解决学生在评价中对于定性指标的不确定判断问题,突破传统量化评价的局限性,将学生体验纳入评价框架,丰富体育教学评价的理论内涵;以证据理论方法构建基于学生体验视角的体育教学评价模型,为处理体育教学评价中的不确定性问题提供新的方法论参考;整合教育学、心理学、信息科学等多学科理论,构建具有跨学科特征的体育教学评价体系,为改善教学质量并最终提升学生体质健康水平提供理论支持。

1 关于学生体验视角教学评价的由来与争议

自 20 世纪 70 年代初以来,以学生体验视角的教学评价都是教育科学研究的重要课题,并被认为是一种改善教学质量的有效工具。随着人们对学生教学评价研究的持续深入,学生评价意见对于教学质量提升的作用也呈现出不同观点^[1-4]。支持者认为,开展学生评价体现教学中学生主体地位的理念,是教育以人为本思想的表现^[5],定期且规范的学生教学评价有利于提高教学质量,学生评教结果也一直是教学管理机构决策的重要参考依据^[6]。20 世纪 90 年代美国高校约有 90% 的教学管理者采用各种方式的学生评价来开展教学质量评估^[7]。研究者认为教学评估是一项复杂的系统工作,而以学生体验视角的教学评价是一个重要内容^[8],同时强调学生评价结果的有效性必需建立在提前对参评学生进行沟通和指导^[9]。支持意见笃信学生评教结果的科学性来自于评价工具的可靠性^[10-11],评价工具也早已不是简单的主观感受和反馈意见问卷,各种基于教学理论的复杂指标体系和心理量表被开发出来,用于搜集测量学生评教的各类量化数据和主观信息^[12-13]。目前国际上常用的学生评价 SET 量表多达 200 多个^[14]。

学生体验视角教学评价的怀疑与争论出现在 20

世纪末,相关研究对于学生评教结果的有效性、评价工具的可信性和评价环境的可控性等问题都出现广泛质疑,并且开展大量实证研究^[15]。质疑研究主要集中在学生教学评教的 4 个方面:评价执行的特征、课程教学的特征、教师自身的特征和学生自身的特征^[16-17]。开展教学评价的目的是管理教学质量,但研究表明学生评教更多地是体现对教师满意度的无意义量化,并且导致制约教师教学技能发挥的问题出现^[18]。而且大多数学生的评分表缺乏心理测量标准方面的效度和信度验证,评价工具能否在评价中全面准确反映实际情况,是学生评价执行中的现实难题^[19-20]。

在课程教学特征方面,研究认识到课程本身的特征,例如课程性质、课程内容、课程难度、学科领域甚至其他教学环境因素(教室场地和教学设备),都可能对学生教学评价起到影响,进而影响评价结果^[21-22]。教师自身特征对于学生教学评价结果的影响则更大,直接证据表明与教师教学无关的背景特质和因素都可能会使学生的评价产生偏差,这些因素包括性别、年龄、性格甚至是容貌和声音^[23-24]。在教师特征方面被讨论最多的评价有效性无关因素是学生常常会根据他们所获得最终成绩对教师的教学进行评分,这就使得趋于严厉教学风格的教师不能获得客观评价^[25]。

在体育教学中,学生自身特征对于学生体验视角的教学质量评价有着复杂联系和深刻影响,学生在体育教学过程中期待的回馈是教学内容改善和教学方法改进,但是现实中评教结果往往只用于学校对于教师的管理,没有反馈的评教逐渐影响学生评教的积极性,而认真积极的态度又是评教结果客观科学的关键保证^[26]。如果学生在体育教学评教过程中不能明确评价目的,无法理解评价内容,不能正确使用评价方法和工具,这些问题都会导致最终结果出现偏差,所以学生评教应当充分考虑学生特征,规范设计评价工具,科学选择评价方法^[27]。

综上所述,学生教学评价尽管多有质疑和争论,但是仍然是教学质量评价系统中的一项重要内容,它的执行成效是影响教学质量的重要因素。已有文献认为,学生评教是面向课程教学的直接反馈,评价实践最重要的是尽量排除教学无关因素的干扰,以科学合理的测评工具获取学生对教学工作客观准确的核心认知信息^[28]。因此,如何完成对定性指标的合理表达和科学测定一直是体育教学评价中的重点和难点问题。

2 基于证据理论方法的体育教学评价模型

2.1 证据理论方法基础

证据理论是在传统概率论基础上提出的一种不确

定性推理方法,是以识别框架为基础解决不确定因素的数学方法^[29-30]。证据理论方法处理形式灵活,不要求了解已知问题的先验信息就能完成方法推理,在评价中对于具有“未知性”属性指标的处理更加准确合理。另外,证据理论方法通过整合信度结构和评价等级两种函数,以非线性信息融合的方式汇集证据矩阵,达到证据间支持加强和抵消抑制的效果,符合现实中不确定性问题的真实表现^[31]。因此,证据理论方法由于其强大的功能已经在医疗评价、系统评价和工程评价等领域得到广泛应用。

作为一种处理不确定性问题的数学方法,证据理论的理论内核与体育教学评价的实践需求具有高度契合性。其核心优势体现在3个方面:一是无需先验信息即可完成推理,这与体育教学评价中常面临的学生个体差异大、历史数据不足等问题相适配;二是对未知性指标的处理更精准,通过构建信度结构与评价等级的非线性映射关系,能够将学生模糊的主观感受转化为可量化的证据矩阵;三是多源证据融合机制,通过 Dempster 组合规则整合学生自评、互评、教师观察等多维度数据,实现证据间的支持加强与冲突消解。体育教学评价的核心难点在于定性指标的量化困境。

2.2 体育教学评价模型的理论构建

证据理论方法为体育教学评价模型提供多层次的系统构建方案:首先是建立识别框架,以体育教学核心目标为导向,将学生体验感受为内容的定性指标纳入识别框架,形成包含基础命题和复合命题的评价维度体系;然后进行证据获取与预处理,通过问卷调查、行为观察、生理指标监测等多源数据采集,运用隶属度函数将原始数据转化为基本概率分配(BPA),同时通过专家咨询确定各证据源的可信度权重;最后完成非线性证据融合,采用 Dempster 组合规则对多源证据进行合成,通过计算冲突系数和信任函数,动态调整不同证据间的支持度,识别冲突来源并赋予更可靠证据更高权重,最终输出综合评价结果。

在体育教学评价实践中,对于非确定性指标的学生评价倾向于使用定性描述来表达主观偏好,因为这样的表述更接近学生的真实感受,也能够更全面反映学生的评价意见,但是未合理量化的评价表述不符合评价指标信息的标准化要求,因此如何对所形成的评价主观意见信息进行科学处理,为评价体系中定性指标的测度提供准确信息,是构建科学体育教学评价模型的核心要素。

本研究基于学生评教的现实问题及其产生根源,根据证据理论方法以一种新的信度结构测度学生对于体育教学中各种不确定类型的指标,将学生主观意见

从预设评语等级中的元素扩充为其子集,设置符合半可加性的评价意见分布,对学生在定性指标的偏好表达上给予多维度的选择权利,利用扩充指标信息内容的方法,提高指标综合测度的准确性。最终使用投影方法和 SWA 算法对评价意见进行集成,从而得到学生整体对于体育教学的科学评价结果。

3 体育教学评价模型的构建

本研究采用无反馈单轮次的阶段型群体评价对教学指标进行测度。设方案集 $A = \{a_r | r=1,2,\dots,R\}$, 指标集 $Y = \{y_k | k=1,2,\dots,K\}$, 学生集合 $E = \{e_t | t=1,2,\dots,T\}$ 。研究问题可以表述为:在群体评价识别框架下,利用证据理论将学生集合 e_t 对方案 $a_r (1 \leq r \leq R)$ 在指标 $y_k (1 \leq k \leq K)$ 下(各种类型)主观偏好信息,统一转化为信度结构 $s_{k,r,t}$,并在此基础上对群体意见序列 $\{s_{k,r,t}\} (1 \leq t \leq T)$ 进行集结,最终得到群体(学生集合)对定性指标 y_k 在方案 a_r 下的评价 $s_{k,r}$ 。

3.1 证据理论方法

假定一个环境 Θ 是一个互斥、完备且可穷举的集合,表示为如下形式:

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\} \quad (1)$$

当 Θ 中的元素被视为某一证据可能支持的命题(或某一问题可能的答案),且仅有唯一的命题(可以由一个或多个元素组成)得到支持,则环境 Θ 被称为一个识别框架。

定义1:设环境 Θ 为一个识别框架,如果集函数 $m:2^\Theta \rightarrow [0,1]$ (2^Θ 为环境 Θ 的幂集),且满足 $m(\emptyset)=0$ 和 $\sum_{A \subseteq \Theta} m(A)=1$,则称函数 m 为识别框架 Θ 上的一个基本可信度分配。

$m(A)$ 表示证据支持 Θ 中命题 A 发生的程度。若 $A \subseteq \Theta$ 且 $m(A)>0$,则称 A 为证据的焦元, $m(A)$ 主要凭借主观经验得出。

定义2:假设识别框架 Θ 下的两个证据 E_1 和 E_2 ,相应的基本可信度分配分别为 m_1 和 m_2 ,对应的焦元分别为 A_i 和 B_j ,设 $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j) \leq 1$,则证据合成法则为:

$$m(A) = \begin{cases} \frac{\sum_{A_i \cap B_j = A} m_1(A_i)m_2(B_j)}{1-K} & A \neq \emptyset \\ 0 & A = \emptyset \end{cases} \quad (2)$$

上述合成记为 $m_1 \oplus m_2(A)$,系数 K 反映了证据之间的冲突程度,当 $K=1$ 时表示两个证据完全冲突。

在定性指标评价中,证据理论方法在识别框架(语言评价等级集合) $H = \{H_n | n = 1, 2, \dots, N\}$ 的基础上 ($H_1 \prec H_2 \prec \dots \prec H_N$, 其中 \prec 表示“劣于”), 通过构建信度结构:

$$s_{kr} = \{(H_n, \beta_{kr}^n), n = 1, 2, \dots, N; (H, \beta_{kr}^H)\} \quad (3)$$

对评价者主观信息进行测度。其中, β_{kr}^n ($0 \leq \beta_{kr}^n \leq 1$) 表示评价者对方案 a_r 在定性指标 y_k 下被评价为评语等级 H_n 的信任程度, 简称信度; $\beta_{kr}^H = 1 - \sum_{n=1}^N \beta_{kr}^n$ 为表示由于评价者的“无知”而未被分配的剩余信度, 将其分配给整个识别框架 H 。

3.2 学生评价意见偏好的表达

1) 学生评价意见偏好表达形式。

单个群体成员对定性指标的偏好表达一般分为比较和直接两种形式。如 AHP 方法是通过方案间“两两比较”后得到各项的相对重要性作为对定性指标的测度, 这种方法易受指标数量限制。因此在评价实践中, 采用直接形式表达偏好更具有适用性。

根据主观信息类型的不同, 直接偏好表达又可以分为数量和语言两种形式。数量形式一般采用主观评分的思想, 而语言形式则采用语言等级对定性指标进行测度。数量形式实质上是评价等级与量化差异统一为数量标度, 而语言形式是将评价等级与量化差异分为 2 个层次, 用于区分不同语言等级间实际差异的变化^[32]。当语言等级间实际差异一致时, 语言等级与数量形式的偏好表达实质是等价的。在体育教学评价中, 特定的定性指标不适合进行数量形式的评定, 因为学生在主观转换评价意见时会因为自身原因造成表达信息的扭曲和损失, 因此更适合利用语言形式表达主观偏好意见。

2) 学生评价意见偏好表达内容。

学生在体育教学评价中的偏好表达内容主要涉及两个方面, 一是学生自身基于语言等级偏好表达的能力, 二是对自身偏好的信任程度。

以学生体验视角的体育教学评价要充分考虑学生自身的认知水平, 以及学生运用语言自由表达的能力。在评价实践中经常会出现问卷问题过于晦涩, 学生不能充分理解而无法给出确切的答案, 这就造成信息获取的缺失, 进而影响评教结果^[33]。基于学生体验视角的思想, 就是将学生视为教学评价过程中提供主观信息的用户, 尽可能允许其不受限制、充分地表达自身对定性指标的偏好信息, 评价设计需给予学生充分的自由表达空间, 摒弃限定性和引导性的表述方式, 使学生能够毫无束缚地阐述对体育教学定性指标的看

法, 确保学生能够真实、全面地传递自身评价意见, 从而为教学评价提供丰富且准确的主观信息。

学生自身偏好信任程度在教学评价实践中是指学生对于自己所表达评教意见的理解和认可程度。单一数据形式的评价调查暗含学生对其自身的主观意见完全确信这一不合理假定, 事实上学生的回答仍有可能是模棱两可的猜测或者一定程度的妥协结果^[34]。因此, 在测度学生体育教学评教意见时, 还应深入了解学生实际情况并全面考虑体育教学各环节, 如教学内容设置、教学方法运用、师生互动模式等。通过多维度收集信息, 为问卷或量表设计提供有价值的辅助内容, 进而为有效测度学生评价意见的不确定性提供依据, 同时在整合评价意见时为设定意见的可靠性提供参考, 以提升体育教学评价的科学性与准确性。

体育教学评价需平衡定量与定性数据, 既要通过体能测试、技能达标等量化指标反映身体素质, 又要通过开放性问题、行为观察等定性方法捕捉学生的运动态度。此外, 体育教学评价还具有紧密结合运动场景和运动项目内容的特殊性, 学生在教学过程中的情感体验也是评价的重要内容。基于证据理论框架的信度结构因其复合形式, 能够满足基于大规模群体框架下单个群体成员对定性指标评价时, 偏好表达在形式和内容上的要求, 可以有效规避评价中评价意见出现的偏差, 因此非常适用于作为以学生体验视角的体育教学质量评价工具。

3.3 学生评价主观信息的测度

1) 建立识别框架。

由于信度结构建立在识别框架基础上, 因此首先要定义语言评价等级的识别框架。利用语言评价等级集合构建基于多属性决策方法的识别框架, 为测度评价等级的区间不确定及局部未知, 将集合扩展为矩阵形式(4), 其中 H_{ij} ($i = 1, 2, \dots, N$; $j = 1, 2, \dots, N$), 表达了评价者区间形式的主观判断。

$$H = \begin{Bmatrix} H_{11} & H_{12} & \cdots & H_{1(N-1)} & H_{1N} \\ & H_{22} & \cdots & H_{2(N-1)} & H_{2N} \\ & & \ddots & \vdots & \vdots \\ & & & H_{(N-1)(N-1)} & H_{(N-1)N} \\ & & & & H_{NN} \end{Bmatrix} \quad (4)$$

2) 构建信度结构。

基于识别框架中单个群体成员 e_i 对方案 a_r 在定性指标 y_k 下的主观偏好, 可以用以下信度结构形式进行测度:

$$s_{kr,t} = \{(A_i, \beta_{rk,t}^i \{A_i\}); i = 1, 2, \dots, j\}, A_i \subseteq H \quad (5)$$

其中, $0 \leq \beta_{rk,t}^i \leq 1$ 表示对评级等级集合 A_i 的信度,

j 表示识别框架 $s_{kr,t}$ 中焦元的个数。 A_i ($i=1,2,\dots,j$) 为识别框架 H 中的一个子集,当元素多于一个时,要求元素间是相邻的,如 $s_{kr,t}=\{(\{H_3,H_4,H_5\},1)\}$ 表达了区间不确定与局部未知。

基于信度结构(5)可对各种类型的主观信息进行测度。例如,主观信息为完全没有认知(数据缺失)

表示为 $s_{kr}^t=\{(\{H\},1)\}$; 不完全认知($\sum_{i=1}^j \beta_{rk,t}^i < 1$)表示为

$s_{kr}^t=\{(\{H_4\},0.7)\}$ 或 $s_{kr}^t=\{(\{H_3\},0.7),(\{H_4\},0.2)\}$; 不精确认知则表示为 $s_{kr}^t=\{(\{H_4\},0.7),(\{H_5\},0.3)\}$; 区间认知表示为 $s_{kr}^t=\{(\{H_3,H_4,H_5\},1)\}$ 等。因此,包括缺失数据在内所有类型的评价意见都可以转化成为统一的信度结构,这符合学生评教实践中的现实情况,对于定性指标学生所表现出“无知”状态的不确定性分布,这种在主观评价上细微差距的区分与量化表现正是本研究模型的优势。此外,统一的信度结构也为群体意见的集成奠定可靠性基础。

3.4 学生评价意见的集成

完成对学生主观评价意见的表达与测度后,要以信度结构表示的个体意见进行集成,并最终得到全部学生对教学定性指标的评价结果。因为信度结构存在集合形式,本研究采用投影方法首先将含有集合形式的信度结构转化为元素形式,再通过加权平均方法对信度结构完成集成。

基于投影的证据集成方法是将 N 个互斥的命题视为两两相互垂直的坐标轴,因此识别框架成为一个包含 N 个坐标轴的坐标系。其实质是通过投影将含有集合形式焦元 A_i ($i=1,2,\dots,j$) 的信度结构转化为单元形式信度结构,即焦元的单元素化。

1) 单个群组意见的投影。

基于群体评价框架,以信度结构为基础给出关于群体成员信度投影的定义。

定义 3: 设群体评价等级集合 H 为一个包含 N 个互斥元素的识别框架, $\beta_{rk,t}^i\{A_i\}$ 为群体成员 e_i 对方案 a_r 在定性指标 y_k 下评价为 A_i 的信度,则定义:

$$p_{rk,t}^{in} = \beta_{rk,t}^i\{A_i\} \cdot \cos \alpha_{in} \quad (6)$$

为信度 $\beta_{rk,t}^i\{A_i\}$ 向坐标轴 n 的投影。其中 α_{in} 表示焦元 A_i 与坐标轴 H_n ($n=1,2,\dots,N$) 间的夹角。

假设焦元 A_i 中包含有 s 个元素,且 A_i 与 s 个元素所在的 s 个坐标轴间的夹角相等且和为 90° , 则 $\beta_{rk,t}^i\{A_i\}$ 中焦元 A_i 向坐标轴 H_n 的投影为:

$$p_{rk,t}^{in} = \begin{cases} \beta_{rk,t}^i\{A_i\} \cdot \cos(\frac{90^\circ}{s}), & 2 \leq s \leq N \\ \beta_{rk,t}^i\{A_i\}, & s = 1 \end{cases} \quad (7)$$

而与其他坐标轴间的夹角分别为 90° 。

对于识别框架 H 中任意一个评价等级 H_n , 加总群体成员 e_i 意见中所有在坐标轴 H_n 上的投影信度为:

$$H_n: p_{rk,t}^n = \sum_{i=1}^j p_{rk,t}^{in} \quad (8)$$

$p_{rk,t}^n$ 表示单个群组意见经过投影后对评价等级 H_n 的支持度。但所有评价等级下的支持度与剩余可信度分配的和不一定为 1, 即 $\beta_{rk,t}\{H\} + \sum_{n=1}^N p_{rk,t}^n \neq 1$, 因此需要归一化。由于 $\beta_{rk,t}\{H\}$ 代表整个识别框架中的未分配信度, 投影后应保持不变。因此转化后的单个群组成员意见的关于评价等级 H_n 的信度分配为:

$$\beta_{rk,t}^n\{H_n\} = \frac{p_{rk,t}^n(1 - \beta_{rk,t}\{H\})}{\sum_{n=1}^N p_{rk,t}^n} \quad \text{且} \quad \sum_{n=1}^N \beta_{rk,t}^n\{H_n\} + \beta_{rk,t}\{H\} = 1 \quad (9)$$

经过投影, 本研究将含有集合形式的信度结构转化为元素形式, 即将信度结构转化为以下的形式(10):

$$s_{kr,t} = \{(H_n, \beta_{rk,t}^n\{H_n\}), n=1,2,\dots,N; (H, \beta_{rk,t}^H\{H\})\} \quad (10)$$

2) 学生群组评价意见的集成。

在集成前需要考虑每个学生评价意见的相对重要性。通过基于证据推理方法下的信度结构, 假设群组成员 E 意见的相对重要性程度设定为 $w = \{w_t | t=1,2,\dots,T\}$

($0 \leq w_t \leq 1, \sum_{t=1}^T w_t = 1$), 其中群组成员 e_t 意见的重要程度为 w_t 。通过定义基于信度结构的 SWA 算子对群体意见进行集成。

定义 4: 设 s_t ($t=1,2,\dots,T$) 为基于识别框架 $H = \{H_n | n=1,2,\dots,N\}$ 的信度结构:

$$s_t = \{(H_n, \beta_{rk,t}^n), n=1,2,\dots,N; (H, \beta_{rk,t}^H)\},$$

则定义基于这一信度结构的 SWA 算子为:

$$\text{SWA}(s_1, s_2, \dots, s_T) = \left\{ \left(H_n, \sum_{t=1}^T w_t \beta_{rk,t}^n \right), n=1,2,\dots,N; \left(H, \sum_{t=1}^T w_t \beta_{rk,t}^H \right) \right\} \quad (11)$$

其中, $w = \{w_t | t=1,2,\dots,T\}$ ($0 \leq w_t \leq 1, \sum_{t=1}^T w_t = 1$)

为针对 s_t 的加权向量。

因此, 根据 SWA 算子可将群组成员 e_t 对方案 a_r 在定性指标 y_k 下的基于信度结构的群体意见集成为一个新的信度结构:

$$s_{kr} = \left\{ \left(H_n, \sum_{t=1}^T w_t \beta_{rk,t}^n\{H_n\} \right), n=1,2,\dots,N; \left(H, \sum_{t=1}^T w_t \beta_{rk,t}^H\{H\} \right) \right\} \quad (12)$$

表示整个群体对定性指标 y_k 的测度, 由此得到整

体学生评价意见结果。

4 实证研究结果

以某高校大学体育课程为例,说明基于学生体验视角与证据理论方法的教学评价体系如何在实践中执行。本研究选取其中一个 20 人($E = \{e_t | t=1,2,\dots,T\}$)的课程教学班级为例,利用评价等级集合分别从教学态度(y_1)、教学环节(y_2)、教学内容(y_3)、教学效果(y_4)和教学方法(y_5)5 个维度对 5 名任课教师(a_1, a_2, a_3)的教学进行评价。假设 5 个指标维度权重都相等,20 个学生意见的可靠度相等($w_t = 0.05, t=1,\dots,20$)。

教学评价设计分为两轮:第一轮要求学生每位教师在不同的评价指标下给出语言评价等级,学生可

以在评价等级集合的基础上给出任意形式的答案;第二轮要求学生自身给出的评价等级的信任程度进行判断,对于多选形式,既可以给出每一个等级对应的信任程度(形成分布评价),也可以给出一个整体的信任程度(形成区间评价)。具体步骤如下:

(1)评价问卷填写。从评价方法角度,这一步骤是作为群体评价的学生,表达自身主观偏好意见的环节。表 1 为全体学生对教师 a_1 在教学态度 y_1 上表达的主观偏好。从信度数据可以发现,很多学生对自身的答案并非十分确定。其中对答案不确定的原因可能是多种的,如对教师不熟悉或对指标概念理解不太清晰等,而这些信息在传统评分形式的教学评价下并未显现。

表 1 全体学生对教师 a_1 在教学态度 y_1 上的主观偏好

学生	评价结果及信度						学生	评价结果及信度					
	H_1	H_2	H_3	H_4	H_5	H		H_1	H_2	H_3	H_4	H_5	H
e_1			0.6	0.4			e_{11}				0.5		0.5
e_2			0.9			0.1	e_{12}				1		
e_3						1	e_{13}			0.9		0.1	
e_4				1			e_{14}				0.7		0.3
e_5					1		e_{15}				0.8		0.2
e_6					0.6	0.4	e_{16}					1	
e_7				1			e_{17}				1		
e_8				0.2	0.7	0.1	e_{18}				0.5	0.5	
e_9				0.9		0.1	e_{19}			0.4	0.6		
e_{10}				1			e_{20}				0.6	0.3	0.1

表 2 全体学生主观偏好的测度

学生	信度结构	学生	信度结构	学生	信度结构	学生	信度结构
e_1	$\{(H_3, 0.6), (H_4, 0.4)\}$	e_6	$\{(H_5, 0.6)\}$	e_{11}	$\{(H_4, 0.5)\}$	e_{16}	$\{(H_5, 1)\}$
e_2	$\{(H_3, 0.9)\}$	e_7	$\{(H_4, H_5, 1)\}$	e_{12}	$\{(H_4, 1)\}$	e_{17}	$\{(H_4, 1)\}$
e_3	$\{(H, 1)\}$	e_8	$\{(H_4, 0.2), (H_5, 0.7)\}$	e_{13}	$\{(H_4, H_5, 0.9)\}$	e_{18}	$\{(H_4, 0.5), (H_5, 0.5)\}$
e_4	$\{(H_4, 1)\}$	e_9	$\{(H_4, H_5, 0.9)\}$	e_{14}	$\{(H_4, 0.7)\}$	e_{19}	$\{(H_3, 0.4), (H_4, 0.6)\}$
e_5	$\{(H_5, 1)\}$	e_{10}	$\{(H_4, 1)\}$	e_{15}	$\{(H_3, H_4, 0.8)\}$	e_{20}	$\{(H_3, H_4, 0.6), (H_5, 0.3)\}$

(2)主观评价意见测度。根据识别框架和信度结构,对表 1 中每个学生表达的偏好信息进行测度。表 2 为测度结果,表中学生 e_3 没有给出评价信息。

(3)投影转换。利用定义 3 和公式(9)将表 2 中含有集合形式评价等级的信度结构转化为单元元素形式。如信度结构 $s_{11}^{20} = \{(\{H_3, H_4\}, 0.6), (H_5, 0.3)\}$, $\beta_{11,20}^1 \{H_3, H_4\}$ 中集合焦点 $\{H_3, H_4\}$ 向坐标轴 H_3 和 H_4 的投影分别为:

$$p_{11,20}^{13} = p_{11,20}^{14} = \beta_{11,20}^1 \{H_3, H_4\} \cdot \cos\left(\frac{90^\circ}{2}\right) = 0.6 \times \frac{\sqrt{2}}{2} = 0.424, \\ p_{11,20}^{15} = 0.3$$

根据公式(8)可知:

$$H_3 : 0.42; H_4 : 0.42; H_5 : 0.3; H : 0.1$$

根据公式(9)进行标准化得到:

$$\beta_{11,20}^3 \{H_3\} = \beta_{11,20}^4 \{H_4\} = 0.33; \beta_{11,20}^5 \{H_5\} = 0.24; \\ \beta_{11,20} \{H\} = 0.1$$

即通过投影将含有集合的信度结构 s_{11}^{20} 转化为

$\{(H_3, 0.332), (H_4, 0.332), (H_5, 0.236)\}$ 的单元素形式。同理对 e_7 、 e_9 、 e_{13} 、 e_{15} 进行投影,从而将表1中所有含有集合的信度结构转化为单元素信度结构。

(4)群体评价意见的集成。利用定义4和公式(11、12),将所有基于单元素信度结构的学生评价意见($w_t = 0.05$, $t=1, \dots, 20$)进行集成,最终得到全体学生对教师 a_i 在教学态度 y_1 上的评价,以信度结构表示为如下形式:

$$s_{11} = \{(H_3, 0.15), (H_4, 0.44), (H_5, 0.27); (H, 0.14)\}$$

同理可对教学环节(y_2)、教学内容(y_3)、教学效果(y_4)和教学方法(y_5)4个维度进行测度,完成整体教学评价。

5 讨论

结合实证研究结果可以看到在新的评价模型中,学生不仅对体育教学评价指标给出客观评价意见,还得到定性指标的区间概率分布测度结果(见表2),而不是仅仅完成与否的判定或者等级评分,对于指标的区间偏好表达学生在评价实践中的真实感受,利用信度结构方法解决对于定性指标定义模糊粗糙的问题,同时模型也符合学生在评教中的思维认知过程,充分体现以人为本的学生视角评价思想。评价结果通过投影转换和SWA方法的集成方式以具体数据形式体现,保证最终结果兼备科学性和可比性,有利于评价模型在反馈、督导和评优等各项教学管理工作中的普遍适用。

以学生体验视角构建的评价模型相较传统学生评教测度模式更加全面可信、更加友善可行,有利于测度信息的科学获取。评价结果信息显示在评价指标层和准则层中不同的可能结果概率,完整刻画学生评价者对于定性指标偏好设定的精确程度,通过获取更多评价意见和表述更多主观偏好的方法,切实提高体育教学评价中定性指标测度的全面性和准确性。

基于证据理论方法的体育教学评价模型体现主观分布测度评价意见的思想,通过对识别框架中全部评价指标的平衡信度分配,扩充定性指标评价信息的内容含量,充分融合学生群体的评价意见和偏好程度,模型不仅能够明确表示体育教学评定结果,还可以构建学生评价指标的概率分配函数,排除干扰性证据因素用以反馈评价体系的修正,提升体育教学综合评价结果的可信度和评价实践指导性。

学生体验视角的体育教学评价在实践中很容易受到学生认知水平和非教学因素影响,从而造成最终评价结果的可靠性和科学性备受争议。因此,建立充分考虑学生感受的测评工具,开展科学化教学评价,是

提升体育教学质量和学生体质健康水平的前提条件。研究以一种新的信度结构用于测度主观评价意见,为教学评价定性指标的测度提供具有定距尺度的原始信息,充分考虑在体育教学评教中学生根据自身体能健康情况表达的各种主观偏好,并使用投影方法和SWA算法对基于信度结构的评价意见进行集成,通过对学生体质健康水平的表达意见综合,来获取更加科学合理的评价结果。

在体育教学评价中有大量的定性评价指标,鉴于评教学生的心智能力和认知水平,在面对这些指标时往往难以给予确定的判断,文中建立的评价模型在学生体验视角基础上,使用语言评价等级集合并允许学生对评价等级进行程度判断,充分考虑学生的主观感受和意见,全面获取评价因素尤其是定性指标的深刻信息,有效解决学生评教中的主要争议问题,切实提升学生体育教学评教结果的可靠性和科学性,也为针对提升学生体能健康能力的教学改革提供理论依据。

参考文献:

- [1] 孙鳌,孙由之. 学生评教中的博弈与变革[J]. 高教发展与评估, 2020, 36(5): 47-56+117.
- [2] GROSS R B, SMALL A C. A survey of faculty opinions about student evaluations of instructors[J]. Teaching of Psychology, 1979(6): 216-219.
- [3] DENEVE H M F. University teachers' thinking about lecturing: Student evaluation of lecturing as an improvement perspective for the lecturer[J]. Higher Education, 1991, 22(1): 63-91.
- [4] HAMMONDS F, GINA J, AMMONS G, et al. Student evaluations of teaching: Improving teaching quality in higher education[J]. Perspectives: Policy and Practice in Higher Education, 2017(21): 26-33.
- [5] 邱德明. 教育以人为本研究述评[J]. 上海教育科研, 2018(6): 23-27.
- [6] LITTLEFORD L N, ONG K S, TSENG A, et al. Perceptions of European American and African American instructors teaching race-focused courses[J]. Journal of Diversity in Higher Education, 2010, 3(4): 230-244.
- [7] 吴立军,田启波. 评“教”还是评“人”?——基于学生评教的有效性研究[J]. 高教探索, 2020(8): 57-65.
- [8] 余海波. 基于学生评教的高校教师教学能力提升[J]. 国家教育行政学院学报, 2017(6): 77-81.
- [9] SAMUEL M L. Flipped pedagogy and student evaluations of teaching[J]. Active Learning in Higher Education, 2021, 22(2): 159-168.

- [10] TROUT P. What the numbers mean: Providing a context for numerical student evaluation of course[J]. *Change*, 1997(29): 24-30.
- [11] ALAUDDIN M, KIFLE T. Does the student evaluation of teaching instrument really measure instructors' teaching effectiveness? An econometric analysis of students' perceptions in economics courses[J]. *Economic Analysis and Policy*, 2014, 44(2): 156-168.
- [12] 张俊杰, 李会超, 郭成根, 等. 现代体能训练理念与方法融入高校公共体育的现状与对策[J]. *中国学校卫生*, 2021, 42(11): 1605-1608+1612.
- [13] 郭凌云, 吴凤彬. 多元化健康教育促进大学生体育锻炼行为的效果评价[J]. *中国学校卫生*, 2021, 42(1): 50-53, 57.
- [14] BEDGGOOD R E, DONOVAN J D. University performance evaluations: What are we really measuring?[J]. *Studies in Higher Education*, 2012, 37(7): 825-842.
- [15] SPROULE R. The underdetermination of instructor performance by data from the student evaluation of teaching[J]. *Economics of Education Review*, 2002, 21(3): 287-294.
- [16] BROCKX B, VAN ROY K, MORTELMANS D. The student as a commentator: Students' comments in student evaluations of teaching[J]. *Procedia—Social and Behavioral Sciences*, 2012, 69: 1122-1133.
- [17] GILBERT R O, GILBERT D R. Student evaluations of teaching do not reflect student learning: An observational study[J]. *BMC Medical Education*, 2025, 25(1): 1-7.
- [18] HASSAN E. Investigating substantive and consequential validity of student ratings of instruction[J]. *Higher Education Research & Development*, 2009, 28(3): 319-333.
- [19] GOOS M, SALOMONS A. Measuring teaching quality in higher education: Assessing selection bias in course evaluations[J]. *Research in Higher Education*, 2017, 58(4): 341-364.
- [20] SPOOREN P, BROCKXB, MORTEHNANS D. On the validity of student evaluation of teaching: The state of the art[J]. *Review of Educational Research*, 2013, 83(4): 598-642.
- [21] UTTL B, WHITE C A, WONG G D. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related[J]. *Studies in Educational Evaluation*, 2017, 54: 22-42.
- [22] GALLAGHER H A. Vaughn elementary's innovative teacher evaluation system: Are teacher evaluation system: Are teacher evaluation scores related to growth in student achievement?[J]. *Peabody Journal of Education*, 2004, 79(4): 89-107.
- [23] MACNELL L, DRISCOLL A, HUNT A N. What's in a name: Exposing gender bias in student ratings of teaching[J]. *Innovative Higher Education*, 2015, 40(4): 291-303.
- [24] MITCHELL K, MARTIN J. Gender bias in student evaluations[J]. *Political Science & Politics*, 2018, 51(3): 648-652.
- [25] PACOL C A. Sentiment analysis of students' feedback on faculty online teaching performance using machine learning techniques[J]. *WSEAS Trans. Inf. Sci. Appl*, 2024, 21: 65-76.
- [26] 谭亚丁, 王国军, 刘石军, 等. 基于 WSR 系统论的高校体育类专业术科课学生评教合作博弈的破解路径[J]. *安徽体育科技*, 2022, 43(3): 71-75.
- [27] 张和平, 谭娟, 何素艳. 优秀体育教师维度、概念模型与评价体系: 基于学生视角的质性研究[J]. *武汉体育学院学报*, 2021, 55(10): 86-92+100.
- [28] 金梅, 方宗祥. 中美高校学生评教差异研究——基于国内外两所高校的实证研究[J]. *大学教育*, 2021(7): 47-50+60.
- [29] DEMPSTER A P. A generalization of Bayesian inference[J]. *Journal of the Royal Statistical Society*, 1968(30): 205-247.
- [30] YANG J B, SINGH M G. An evidential reasoning approach for multiple-attribute decision making with uncertainty[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1994, 24(1): 1-18.
- [31] FAN C L, SONG Y F, LEI L. Evidence reasoning for temporal uncertain information based on relative reliability evaluation[J]. *Expert Systems with Applications*, 2018, 12(113): 264-276.
- [32] 乔晗. 基于证据理论的综合评价定性指标测度优化研究[J]. *统计与信息论坛*, 2017, 32(6): 16-24.
- [33] 宋旭红. 高校学生评教管理功能的价值回归[J]. *中国高等教育*, 2020(6): 57-59.
- [34] STEIN S J, GOODCHILD A, MOSKAL A, et al. Student perceptions of student evaluations: Enabling student voice and meaningful engagement[J]. *Assessment & Evaluation in Higher Education*, 2020, 8: 1-15.