

中小学体育教师评价素养量表的编制及信效度检验

郭洪亮^{1, 2}, 许文鑫³, 郭永波⁴, 于永晖⁵, 姚蕾²

(1.北京体育大学博士后流动站 北京 100084; 2.北京体育大学 教育学院, 北京 100084;
3.福建师范大学 体育科学学院, 福建 福州 350007; 4.广州商学院 体育学院, 广东 广州 510320;
5.中国人民大学 体育部, 北京 100872)

摘要: 基于新课改背景编制中小学体育教师评价素养量表, 实现中小学体育教师评价素养的量化评价。运用文献资料法厘定中小学体育教师评价素养的基本内涵和提出其核心要素, 并据此采用专家咨询法、数理统计法编制中小学体育教师评价素养初始量表; 选取 784 名中小学体育教师(样本 1)的问卷进行项目分析、探索性因子分析; 选取 833 名中小学体育教师(样本 2)的问卷进行验证性因子分析、信效度以及恒等性检验; 间隔 4 周后, 对来自样本 2 的 120 名中小学体育教师(样本 3)的问卷进行中小学体育教师评价素养量表信度重测。结果表明: 中小学体育教师教学评价素养量表由评价知识、评价技能、评价态度、评价习惯、评价品格 5 个分量表构成, 共 22 个题项。二阶五因子结构模型各项拟合指标达到统计学意义; 量表的内部一致性系数、重测信度系数、内容效度系数、收敛效度系数均介于标准值区间; 量表具有跨性别、跨学段、跨地区的测量等值性。研究认为, 中小学体育教师评价素养量表具有一定的科学性和可靠性, 可以作为测量我国中小学体育教师评价素养的有效工具。

关键词: 学校体育; 中小学体育教师; 评价素养; 体育与健康课程标准; 量化评价

中图分类号: G807.2; G807.3 文献标志码: A 文章编号: 1006-7116(2025)02-0114-09

Establishment of evaluation literacy scale for physical education teachers in primary and secondary schools and its reliability and validity test

GUO Hongliang^{1,2}, XU Weixin³, GUO Yongbo⁴, YU Yonghui⁵, YAO Lei²

(1.Postdoctoral Mobile Station, Beijing Sport University, Beijing 100084, China; 2. School of Education, Beijing Sport University, Beijing 100084, China; 3.College of Sports Science, Fujian Normal University, Fuzhou 350007, China; 4.School of Physical Education, Guangzhou College of Commerce, Guangzhou 510320, China;
5.Department of Physical Education, Renmin University of China, Beijing 100872, China)

Abstract: Based on the background of the new curriculum reform, the evaluation literacy scale of primary and secondary physical education teachers was developed to realize the quantitative evaluation of the evaluation literacy of primary and secondary physical education teachers. The basic connotation and core elements of the evaluation literacy of primary and secondary physical education teachers were determined by the method of literature data, and the initial scale of evaluation literacy of primary and secondary physical education teachers was prepared by Delphi expert consultation and mathematical statistics. A total of 784 physical education teachers (sample 1) were selected for project analysis and exploratory factor analysis. The confirmatory factor analysis, reliability and validity, and identity test were conducted on the questionnaires of 833 primary and secondary school physical education teachers (sample 2). After an interval of 4 weeks, 120 primary and secondary physical education teachers from sample 2 (sample 3) were re-tested for the reliability of the evaluation literacy scale of primary and secondary physical

education teachers. The results showed that the evaluation literacy scale of physical education teachers in primary and secondary schools was composed of 5 sub-scales, including evaluation knowledge, evaluation skill, evaluation attitude, evaluation habit and evaluation character. The fit indexes of the second-order five-factor structure model were statistically significant, and the internal consistency coefficient, retest reliability coefficient, content validity coefficient and convergence validity coefficient of the scale were all in the range of standard values. The present scale has the measurement equivalence of cross-gender, cross-school and cross-region. The conclusion suggests that the evaluation literacy scale of primary and secondary school physical education teachers is scientific and reliable, and can be also used as an effective tool to measure the evaluation literacy of primary and secondary school physical education teachers in China.

Keywords: school physical education; primary and secondary school physical education teachers; evaluation literacy; physical education and health curriculum standards; quantitative evaluation

当前,培养中小学生的体育与健康核心素养已经成为我国基础教育阶段体育与健康课程与教学改革的重要方向。评价作为体育与健康课程教学的重要组成部分,是落实学生体育与健康学科核心素养的重要抓手。对此,教育部颁布的《普通高中体育与健康课程标准(2017年版)》(以下简称《课程标准2017年版》)、《普通高中体育与健康课程标准(2017年版2020年修订)》(以下简称《课程标准2020年修订版》)、《义务教育课程方案(2022年版)》(以下简称《课程方案2020年版》)、《义务教育体育与健康课程标准(2022年版)》(以下简称《课程标准2022年版》)、《基础教育课程教学改革深化行动方案》(以下简称《行动方案》),均对中小学体育教师的教学评价赋予新的内涵要义和提出更高要求,使评价与教学、学习之间的共生关系得到前所未有的体现。然而,受传统体育教学观念的影响,部分中小学体育教师时常将评价置于教学活动之外^[1];部分中小学体育教师在开展体育教学评价活动时,一方面,对学生体育学习的评价内容仍旧局限于运动技能和身体素质的发展与提高,并按照统一的评分标准对学生进行结果评价,最终以运动成绩“论英雄”^[2]。另一方面,为了提高体育教学评价的效率,不喜欢耗时费力却具有人文关怀的评价方法^[3],如增值性评价、表现性评价、档案袋评价等。此外,在评价主体方面,教师拥有控制权和主导权,但学生的评价主体地位被弱化、评价权益被忽视^[4]。显然,以上问题的存在不仅无法满足发展学生体育与健康核心素养的需要,同时也反映中小学体育教师评价素养的不健全。

中小学体育教师评价素养作为教师专业素养的重要组成部分,其高低直接决定课程改革能否切实有效地开展,决定课程改革目标在何种程度上可以实现^[5-6]。因此,提高中小学体育教师评价素养尤为必要和迫切。中小学体育教师评价素养测评工具是揭示中小学体育教师评价素养发展现状和现存问题、提出改进对策的

有力抓手和有形标尺。因此,在发展体育与健康学科核心素养、体育教师评价素养亟待提升的多重背景下,研制严谨、适用、科学的中小学体育教师评价素养测评工具无疑具有重要的理论与现实意义。

1 中小学体育教师评价素养的基本内涵与核心要素

1.1 中小学体育教师评价素养的基本内涵

1991年美国教育评价专家斯蒂金斯^[7]首次提出“评价素养”,并将教师评价素养界定为教师所应具备的基础性教育评价知识和将这些知识运用到各种学生成就测验中的技能。此定义将知识和技能作为教师评价素养的同时,主要体现“对学习进行评价”的评价取向。随后越来越多的研究者将评价素养确认为教师专业素养的组成部分^[8],并在斯蒂金斯定义的基础上对教师评价素养进行多维界定。通过梳理大致可分为以下两种观点:其一,此类观点在界定取向上,关注“评价促进学生的学习”。如Webb^[9]指出:评价素养是关于评价学生所知、所能的方法类知识,怎样解释评价结果的知识,以及如何用评价结果提高教学效率以改善学生学习的知识。斯蒂金斯于2017年在原始定义基础上提出,教师评价素养是指精确测量和报告学生学业成就的知识与技巧,使用评价过程和结果去改善学生学习^[10]。其二,此类观点关注评价促进学生学习的同时,一方面强调教师评价的发生场域,即教师主要针对学生的学习过程和结果进行评价;另一方面体现评价素养的多维性。如李顺雨^[11]指出,教师为了促进学生的学习,通过使用多种评价方法针对学生的学习过程和结果进行评价时所体现出的知识、技能、态度和价值取向的集合。郑东辉^[12]指出,教师评价素养是教师依据促进学习的评价理念开展学生评价所表现出来的关键能力和伦理品格。

由此可见,国内外学者对教师评价素养内涵的认

识莫衷一是，争论的焦点在于对评价素养属概念、对教师评价发生场域和目的取向的认识。首先，就评价素养属概念而言，从构词法的角度来讲评价素养属于偏正短语，因此将中心语素养作为评价素养的属概念是适切的。素养介入到教育领域中不再是简单的文化认知，而是一个抽象的、多维的复杂概念，是知识、技能、态度、伦理等多方面结合在一起的整体表现^[5]。其次，就评价发生场域而言：一方面，评价必须是贯穿于体育教学活动始终的，因为评价的即时性特点就决定“过程性”评价实为其本质；另一方面，核心素养导向的评价观着重强调学生评价实践活动的过程性与综合性^[12]。最后，从评价对于学习的目的来看有以下 3 种典型的取向，一是“对学习进行评价”，二是“为了学习的评价”，三是“作为学习的评价”^[13]。其中，“对学习进行评价”是为了分等和报告，是学习之后对学习结果做出鉴定与区分，无助于学生的后续学习与改进，而“为了学习的评价”和“作为学习的评价”则是明确将评价作为教学的一部分来支持学习^[13]。显然，核心素养导向的评价实践活动，要求教师所要堅持的是为了学习和作为学习的评价取向^[12]。

综上所述，本研究基于对评价素养属概念、教师评价发生场域和目的取向的分析和理解，将中小学体育教师评价素养内涵界定为：中小学体育教师通过不断学习和实践而形成的，能够对学生体育学习过程和结果进行评价且能够通过评价促进学生体育学习所需具备的知识、技能、态度及品行的综合表现。

1.2 中小学体育教师评价素养核心要素

首先，中小学体育教师评价素养是教师专业素养的重要组成部分^[5-6]。按照当前学界对教师专业素养结构的一般认识，其结构通常由知识、能力(技能)、态度(情意)、品格 4 个部分构成^[14]，这可以作为建构中小学体育教师评价素养核心要素的重要参考。

其次，追溯教师评价素养的研究历程。国内外学者基于不同研究视角或采用不同研究思路对教师评价素养要素形成不同观点，其演变轨迹主要从一维走向多维，即从斯蒂金斯最初提出的知识型^[7]到后来的知识与技能型^[8-9, 13]，再到现如今的知识、技能、态度(理念)^[6, 11-12]+ 伦理(品格)^[12, 15]、实践力^[16]、解释力^[16]、意识^[12, 16]等。但是细究发现，评价知识、评价技能、评价态度已经成为教师评价素养结构中的共有要素^[15]。

最后，中小学体育教师评价素养的要素应当具有一定的灵活性和适应性，即必须要与新课改的要求保持一致。鉴于此，本研究结合《课程标准 2017 年版》《课程标准 2020 年修订版》《课程标准 2022 年版》中的“课程理念”“评价建议”，《课程方案 2020 年修订

版》中的“课程实施”以及《行动方案》中的“重点任务”等诸多内容，凝练出符合中小学体育教师评价素养内涵的 4 个要素，即评价知识、评价技能、评价态度、评价品格。如就评价知识而言，《课程标准 2017 年版》在“评价建议”中明确要求教师知晓“评价的内容”即“围绕运动能力、健康行为、体育品德”而评。于评价技能而言，《课程标准 2022 年版》在“评价建议”中提出“采用定量评价与定性评价结合、过程性评价与终结性评价结合……积极探索增值评价，健全综合评价”的要求，《行动方案》将“提升教师教学评价能力”作为重要任务。于评价态度而言，《课程方案 2022 年版》在“课程实施”之“改进教育评价”一栏中，要求教师“重视并积极开展学习评价，倡导评价促进学习的理念”。于评价品格而言，《课程方案 2022 年版》在“课程实施”之“改进教育评价”中要求教师“严格遵守评价的伦理规范，尊重学生人格，保护学生自尊心”。此外，习惯理论认为，所有稳定的素养必然会通过习惯以思维或行动的方式外显，习惯是素养的体现，也是形成素养的主要方式^[17]。因此，评价习惯理应成为中小学体育教师教学评价素养的核心要素。而且，中小学体育教师具备良好的评价习惯即“每课一评”是对接体育课时预设目标是否达成、达成度如何及存在哪些问题的必要通道^[18]。

综上所述，基于对教师专业素养要素的借鉴、国内外教师评价素养文献分析和比较，并结合新课改对中小学体育教师教学评价提出的新要求，本研究将评价知识、评价技能、评价态度、评价习惯和评价品格作为中小学体育教师评价素养的核心要素，并构成中小学体育教师评价素养量表的 5 个维度。

2 研究对象与方法

2.1 调查对象

样本 1：主要用于中小学体育教师评价素养量表的项目分析、探索因子分析。采用方便抽样的方式，于 2024 年 5—6 月通过问卷星平台(www.wjx.com)分别向北京、山东、广东 3 个省市的中小学体育教师发放问卷，901 份剔除无效问卷 117 份(测谎题答错)，有效问卷为 784 份，有效率为 87.0%。其中，男性 379 人(48.4%)，女性 405 人(51.6%)，小学教师 214 人(27.2%)，初中教师 305 人(38.9%)，高中教师 265 人(33.8%)，平均教龄 7.2 年。教育测量学指出量表开发所需的样本量与观测条目之比应为 10 : 1^[19]，中小学体育教师评价素养初始量表题项共为 26 个，因此样本量符合要求。

样本 2：主要用于中小学体育教师评价素养量表的验证性因子分析、信效度检验。采用方便抽样的方

式,于2024年7月通过问卷星平台(www.wjx.com)分别向上海、福建、江苏3个省市的中小学体育教师发放问卷978份,剔除无效问卷145份(测谎题答错),有效问卷为833份,有效率为85.2%。其中,男性479人(57.5%),女性354人(42.5%),小学教师276人(33.1%),初中教师299人(35.9%),高中教师258人(31.0%),平均教龄6.1年。样本量符合教育测量学要求。

样本3:主要用于重测信度分析。采用方便抽样的方式,时隔4周通过问卷星平台(www.wjx.com)从样本2中随机选取120名中小学体育教师进行第二次正式测试。剔除无效问卷17份(测谎题答错),有效问卷为103份,有效率为85.8%。其中男性59人(57.3%),女性44人(42.7%),小学教师39人(37.8%),初中教师38(36.9%)人,高中教师26人(25.2%),平均教龄7.6年。

2.2 研究方法

专家咨询法:本研究通过两轮专家咨询法最大范围内征求、汇集学校体育学研究领域和体育教师发展研究领域的15位资深专家对初步编制的中小学体育教师评价素养量表评价指标及题项进行评估,其中教授11名,副教授4名,博士生导师11名,硕士生导师4名。两轮咨询函回收率为100%,专家积极度高;专家判断依据系数 Ca 介于0.8~1.0之间,表明理论依据和实践经验在很大程度上决定着专家意见。熟悉程度系数 Cs 介于0.8~1.0之间,说明专家对该领域的熟悉程度高。权威度系数 Cr 介于0.8~1.0之间,平均值为0.899,也反映出专家权威度高。

数理统计法:本研究的中小学体育教师评价素养量表为李克特5级量表,所有题项的计分范围为“1~5分”:按5(非常符合)、4(符合)、3(一般)、2(不符合)、1(非常不符合)进行赋值。量表总得分的计算方式为“总分/项目数”,取值范围为“1~5分”^[20],得分越高说明中小学体育教师评价素养越高。采用SPSS 27.0统计软件对样本1调查数据进行项目分析、探索性因子分析,对样本2和样本3调查数据进行信效度检验。采用AMOS 23.0统计软件对样本2数据进行验证性因子分析与恒等性检验。

3 结果与分析

3.1 中小学体育教师教学评价素养量表的设计

首先,将构成中小学体育教师评价素养的5个核心维度确定为一级评价指标;其次,根据各一级评价指标的内涵,选取相应的可观测变量即测量题项。可观测变量主要通过以下两种路径获取:一是借鉴其他学科领域中教师评价核心素养评价工具题项,并更改使用情境和核心词汇。二是参考《课程标准2017年版》

《课程标准2020年修订版》《课程标准(2022版)》《课程方案2022年版》中的相关内容。初步形成由5个一级评价指标和28个题项组成的初始量表。

3.2 中小学体育教师评价素养量表评价指标的完善与确定

1)评价指标完善:德尔菲专家咨询第1轮。

本研究将初始量表编制成专家咨询问卷后,采用问卷星的方式邀请15位专家对各评价指标进行定性评估,评估原则主要包括评价指标是否及义、表达是否清楚、阅读是否流畅、是否晦涩难懂、有无歧义、是否适用于调研人群以及各个维度之间是否相对独立等^[21]。本轮共计发放问卷15份,回收问卷15份,有效问卷15份,回收有效率100%。

回收问卷后发现,5个一级评价指标均获得15位专家的肯定。同时部分专家就测量题项提出以下意见:部分专家建议在评价知识维度下增加“评价方法的使用方法”题项,部分专家建议在健身技能维度下增加“反思技能”题项以及“评价目标设计”题项,另有部分专家认为评价习惯“会自动进行评价”与“评价是我不用刻意去记就会去做的事情”表述重复,违背独立性原则。针对以上意见,经过查阅相关资料以及与部分不同专家反复沟通,本研究决定采纳,并对存在问题的评价指标进行删除,形成包含5个一级评价指标、28个测量题项的中小学体育教师评价素养量表。

2)评价指标确定:德尔菲专家咨询第2轮。

将经过第一轮修改和调整后的中小学体育教师评价素养评价指标编制成李克特5级量表,请各位专家对每一评价指标按5(非常重要)、4(比较重要)、3(一般重要)、2(一般不重要)、1(非常不重要)进行赋值。本轮共计发放问卷15份,回收问卷15份,有效问卷15份,回收有效率100%。待回收问卷后,根据变异系数对各评价指标进行取舍。如果测量题项的变异系数大于或等于0.25,则说明该题项的专家意见协调程度不够^[22]。变异系数计算公式如下所示,其中 S_i 表示*i*指标的标准差(专家认同度), M_i 表示*i*指标的平均值(专家意见集中度), V_i 表示*i*指标的变异系数(专家意见协调度)。通过计算,结果显示中小学体育教师评价素养5个一级评价指标、28个测量题项平均值均在4.3分以上,变异系数均小于0.25,协调系数较好($P<0.001$),表明专家对各评价指标具有较高认可度且意见比较一致。

$$V_j = \frac{S_j}{M_j} \quad (1)$$

3.3 项目分析

首先,对量表的28个题项进行临界比值(CR)分析,具体为:对28个题项求和并进行高低排序,取前

27% 被试者计为高分组，后 27% 被试者计为低分组。随后对高低分组的同一题项的得分进行独立样本 *T* 检验。结果显示，28 个题项均达到高度显著水平 ($P < 0.001$)，CR 值均大于 3.0，表明 28 个测量题项具有良好的鉴别度且全部符合要求^{[23][34]}。接下来采用同质性检验法，对 28 个测量题项与总分的相关性进行分析。若相关系数小于 0.4，表示个别题项与量表构念只是一种低度关系，题项与整体量表的同质性不高，此题项应该删除^{[23][34]}。检验结果发现，28 个测量题项相关系数均大于 0.4。综上所述，经过项目分析后中小学体育教师评价素养量表 28 个题项均保留。

3.4 探索性因子分析

通过项目分析后，对量表进行 KMO 和 Bartlett 球形度检验，结果发现 KMO 值为 0.980，Bartlett 球形度检验结果为 $\chi^2 = 21\,911.088$, $df = 378$, $P = 0.000$ ($P < 0.001$)。根据 Kaiser 的观点，KMO 值为 0.9 以上，Bartlett 球形度检验达到 0.05 显著性水平适合进行因子分析^{[23][76]}。接下来，采用主成分分析法和最大方差法，提取出特征根值大于 1 的因子，并遵循吴明隆等提出的 6 条题项

判断标准^{[23][59]}对题项进行筛选。经过 3 轮探索后，结果显示 KMO 值以及 Bartlett 球形度检验的相关数值均符合测量学要求，能够从众多公因子中提取特征根值大于 1 的因子为 5 个，5 个因子旋转特征值分别为 10.096、4.510、3.763、1.521、1.016，5 个因子累积方差解释率 77.728%，达到 60% 的标准^{[23][36]}(见表 1)，表明因子分析结果理想。同时删除 4 道双重载荷较高的题项(A4 与 B7、B8 与 B9)，删除了 2 道因素负荷量低于 0.45 的题目(C3、E4)。

根据因子分析结果分别对 5 个因子进行命名，其中因子 1 包含 A1、A2、A3、A5、A6 五个题项，命名为“评价知识”；因子 2 包含 B1、B2、B3、B4、B5、B6、B10 七个题项，命名为“评价技能”；因子 3 包含 C1、C2、C4、C5 四个题项，命名为“评价态度”；因子 4 包含 D1、D2、D3 三个题项，命名为“评价习惯”；因子 5 包含 E1、E2、E3 三个题项，命名为“评价品格”。最终，5 个因子和 22 个题项共同形成中小学体育教师评价素养量表。

表 1 探索性因子分析结果

条目	成分	1	2	3	4	5
A1	我完全理解体育教学评价的目的	0.803				
A2	我知道体育教学评价的功能	0.779				
A3	我知道当前体育教学评价目标的内容	0.765				
A5	我知道发展学生体育与健康学科核心素养可以使用哪些评价方法	0.702				
A6	我掌握了过程性评价、增值性评价、表现性评价等方法的使用方法	0.700				
B1	我能够确定与叙写指导评价活动的学习目标		0.819			
B2	我能够根据体育教学内容设计出合理的教-学-评一体化方案		0.810			
B3	我能够根据评价结果，进行合理有效的教学改进和决策		0.772			
B4	我能够使用智能手环、加速度计智能可穿戴设备等监测设备记录学生的课堂行为表现和运动负荷，准确分析和评价学生的运动能力		0.795			
B5	我经常使用书面、口头、记录表等不同方式将评价结果反馈给学生，帮助学生改进学习		0.761			
B6	我会反思课堂评价中信息收集和处理的优势和不足		0.734			
B10	我能够采用不同的评价方法引导学生参与评价		0.727			
C1	我始终秉承促进学习的评价理念			0.714		
C2	我愿意拿出时间来认真研读新课标中的学业要求、学业质量标准以及相应的评价案例			0.709		
C4	我非常认可新课标倡导的过程性评价、增值性评价、表现性评价等评价方法			0.735		
C5	我十分乐意在体育教学中开展评价活动			0.668		
D1	每课一评是我开展体育教学活动中不可或缺的一部分				0.718	
D2	体育教学评价是我不用刻意记就会去做的事情				0.655	
D3	即使课堂教学时间不足，我也会利用课外时间对学生体育学习表现进行评价				0.623	
E1	我拒绝采用统一标准对学生体育学习表现进行评价					0.702
E2	我会针对优等生、中等生、后等生采用不同的评价方式					0.661
E3	在体育教学评价活动中，我不会用评价语言伤害学生					0.622
特征值		10.096	4.510	3.763	1.521	1.016
累计方差解释率%		64.636	69.767	73.144	75.607	77.728

3.5 验证性因子分析

根据探索性因子分析结果, 本研究构建一阶五因子结构模型与二阶五因子结构模型, 随后导入样本 2 数据, 采用 AMOS 23.0 统计软件中的极大似然模型估计法对以上两个竞争模型进行检验和比较。同时依据修正指数表中的参数对模型进行修正。修正结果表明, 如果在 e1 和 e2、e6 和 e10、e11 和 e12、e14 和 e16 之间建立共变关系, 卡方自由度比、均方根渐进误差等拟合指标会符合或接近标准值, 但是二阶五因子结构模型各项适配度指标均优于一阶五因子结构模型。同时, 根据 Doll 建议的目标系数 t 值算法来检验二阶模型的解释力, 计算公式为一阶因子有关模型的卡方值除以二阶因子模型卡方值, t 值越接近 1 则表示二阶模型越能取代一阶模型^[24-25]。通过计算, t 值为 0.809, 即中小学体育教师评价素养二阶五因子结构模型解释了一阶的 80.9%。因此, 可认为与一阶五因子模型相比, 二阶五因子结构模型为最佳模型(见图 1)。此外, 该结果既表明中小学生体育教师评价素养量表具有良好的结构效度, 也佐证中小学生体育教师评价素养结构框架的合理性。

3.6 中小学体育教师评价素养量表的恒等性验证

由于本研究调查对象包含性别、学段及地区, 对此, 在秉持宽松策略的原则下(模型不做任何限制), 采用结构模型恒等性(MGSEM)分析检验中小学体育教师评价素养量表在不同性别、学段、地区之间的测量等值性。通过检验发现, 模型因素负荷、截距、共变异数与测量误差比较中 P 值均大于 0.05 未达到显著性水平, 且 ΔNFI 、 ΔIFI 、 ΔRFI 均小于 0.05, ΔCFI 小于 0.01。以上结果不仅表明中小学体育教师评价素养量表具有跨性别、跨学段、跨地区的恒等性, 而且也为中小学体育教师评价素养模型可作为较为普遍的相关研究基底模型^[25]提供重要依据。

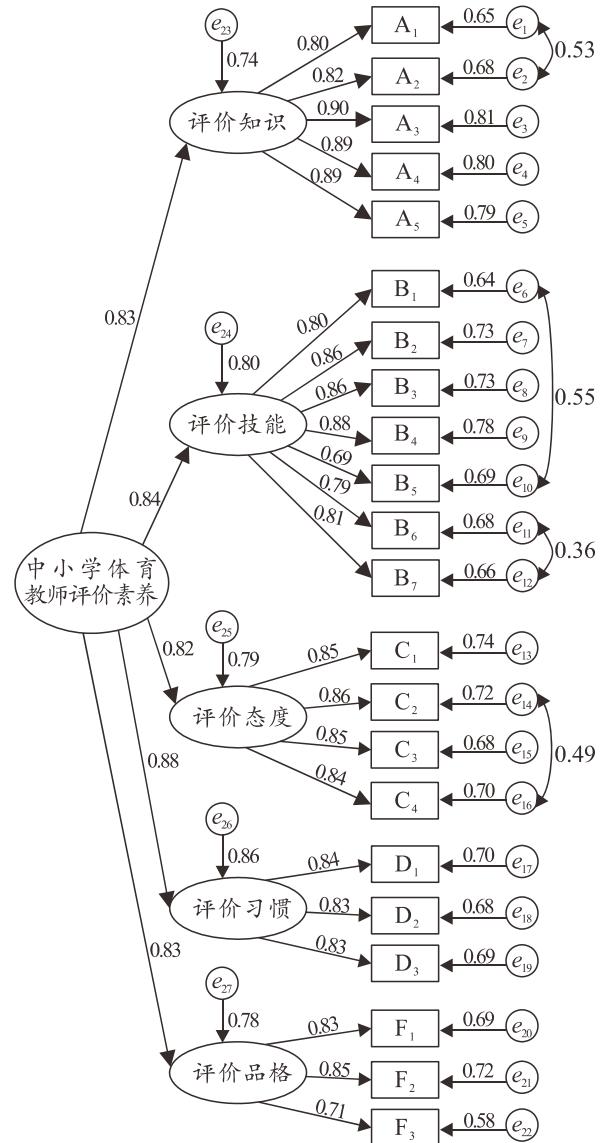
3.7 信效度检验

在本研究中信度检验指标主要为: 内部一致性系数(Cronbach's α 系数)和重测信度。如表 2 所示, 通过对样本 2 数据检验, 中小学体育教师评价素养总量表 Cronbach's α 系数为 0.883, 5 个分量表的内部一致性系数介于 0.806~0.893 之间。通过对样本 3 数据检验, 中小学体育教师评价素养总量表重测信度相关系数为 0.841($P<0.001$), 5 个分量表重测信度相关系数介于 0.818~0.871 之间($P<0.001$)。内部一致性系数介于 0.8 与 0.9 之间表示量表具有稳定的内部一致性, $ICC > 0.80$ 表示重测信度极好^{[23][21]}。因此, 上述结果表明中小学生体育教师评价素养量表具有良好的信度。

表 2 中小学生体育教师评价素养量表信度检验结果

维度	Cronbach's α 系数	ICC	题项数量
评价知识	0.893	0.821 ¹⁾	5
评价技能	0.875	0.859 ¹⁾	7
评价态度	0.806	0.818 ¹⁾	4
评价习惯	0.877	0.871 ¹⁾	3
评价品格	0.820	0.848 ¹⁾	3
总量表	0.883	0.841 ¹⁾	22

1) $P<0.001$



Chi-square=41.262 DF=11

Chi/DF=3.751

GFI=0.986 AGFI=0.965

CFI=0.993 TLI=0.987

RMSEA=0.060

图 1 中小学生体育教师评价素养二阶五因子结构模型

在本研究中效度检验指标主要为: 内容效度、收敛效度。(1)在内容效度方面, 主要包含条目水平的内容效度指数(I-CVI)与量表水平的内容效度指数

(S-CVI), 本研究邀请 15 位的资深专家对题项与其对应的各评价指标之间的相关性进行评定。通过计算, I-CVI 为 0.853, S-CVI 为 0.868。I-CVI 不低于 0.78、S-CVI 不低于 0.8 表示量表内容效度较好^[26]。因此, 以上结果表明中小学生体育教师评价素养量表具有较好的内容效度。(2)在收敛效度方面(如表 3 所示)经过检验, 中小学生体育教师评价素养量表所有题项标准因子载荷均大于 0.500; 5 个分量表平均方差萃取 AVE 介于 0.513~0.546 之间, 组合信度 CR 介于 0.734~0.798。Hair 等指出只有同时满足以下 3 种标准才说明量表具有良好的收敛效度: (1)标准因子载荷(Estimate)均大于 0.500; (2)平均方差萃取(AVE)大于 0.500; (3)组合信度(CR)大于 0.700^{[23][24]}。因此, 以上结果表明中小学生体育教师评价素养量表具有较好的收敛效度。

表 3 收敛效度检验结果

维度	因子载荷	CR	AVE
评价知识	0.81~0.90	0.734	0.529
评价技能	0.69~0.88	0.755	0.513
评价态度	0.84~0.86	0.798	0.546
评价习惯	0.82~0.83	0.777	0.522
评价品格	0.71~0.85	0.743	0.531

4 讨论

4.1 中小学体育教师评价素养量表结构分析

中小学体育教师评价素养量表主要由评价知识、评价技能、评价态度、评价习惯、评价品格 5 个分量表构成。评价知识反映的是体育教师对评价目的、功能、内容以及方法等方面的认识和理解。换言之, 体育教师在开展核心素养导向的评价活动前, 应该熟知“为什么评”“评什么”以及“怎么评”等方面的知识。“为什么评”旨在强调评价目的, 即对学生的体育学习行为进行观察、诊断、反馈、引导和激励, 以判断基于核心素养的课程目标的达成度, 给教师和学生提供即时、多元、有效的反馈, 促使学生更积极地学与教师更有效地教^{[27][25]}。

“评什么”主要围绕运动能力、健康行为、体育品德 3 个核心素养展开。其中, 运动能力评价内容主要包括基本运动技能、体能、专项运动技能等的掌握程度, 以及在日常生活中运用所学运动知识与技能解决实际问题的能力, 在展示或比赛中的表现等^{[27][25]}; 健康行为评价内容主要包括体育锻炼习惯、营养膳食、合理作息、安全防范、情绪调控、适应能力等^[28]; 体育品德评价内容主要包括学生在学练、展示或比赛中表现出的体育精神、体育道德和体育品格等^{[27][26]}。“怎么评”主要指向的是满足学生体育与健康学科核心素养发展所需要的评价方法。评价知识既是体育教师科学有效开展核心素养导向评价活动的重要前提, 也是实现教-学-评一体化和避免

盲目评价的先决条件。

评价技能是指体育教师将评价知识运用于评价实践活动中能力。根据《课程标准 2020 年修订版》《课程标准 2022 年版》的要求, 核心素养导向的体育教学评价活动必须具备以下 5 个技能: 一是会撰写指导评价活动且契合体育与健康学科核心素养的学习目标; 二是运用科学的评价方法准确收集与体育与健康学科核心素养相关的学习信息; 三是依据学习信息对教学进行反思并针对性地改进; 四是将产生的评价结果与初始学习目标比较, 解释反馈给学生, 既要让学生从评价中知道“我的实际水平”, 也要让学生知道“我在哪些方面运动表现较好”“我在哪些方面运动表现存在不足”“接下来我该如何提高”; 五是引导学生参与体育学习评价, 学生由被评价者成为课堂评价的主体既是整体培育学生核心素养的时代之需, 也是课堂评价范式转型的重要诉求。一方面, 对于促进学生的自我调节学习和终身学习均具有重要意义, 另一方面也有利于将反馈信息传递给体育教师, 促进体育教师的教学质量提高和专业改进^[29], 最终实现以评促教。总而言之, 作为体育教师评价活动的核心环节, 评价技能不仅是体育教师将评价知识转换为评价行为的关键和评价能力的集体表现, 而且也是体育教师有效开展核心素养导向评价活动不可或缺的基本手段, 解决体育教师“想评而不会评”的根本基础^[30]。评价态度是指体育教师对教学评价持有的看法以及由此做出行动选择的一种心理倾向^{[16][44]}, 其内蕴着体育教师的价值判断和情感因素, 表征着体育教师的评价立场, 即体育教师如何看待评价以及在开展核心素养导向的评价活动时认同什么、坚持什么。此外, 对于中小学体育教师而言, 真正能够促使其摒弃传统评价理念、实施核心素养导向的评价活动, 并非在于其个人的好奇心和利益的驱使, 而在于其是否真正认同评价改革的价值以及由此产生的主观意愿。由此言之, 体育教师评价态度是开展核心素养导向评价活动的重要驱动, 决定着体育教师的评价行为选择^{[16][44]}。

评价习惯是体育教师形成的自觉稳定的评价行为方式, 主要体现在评价行为稳定性和评价行为自动化两个方面。评价行为稳定性意味着体育教师要“每课一评”。评价行为自动化意味着体育教师评价行为是不需要外在环境的制约、高度自觉的行为方式^[31]。体育与健康学科核心素养的形成是一个时间长久、由量变到质变的过程。显然, 这一过程离不开评价的反馈、导向和激励作用。如果体育教师评价行为缺失恒常性、稳定性, 那么既无法实现课时目标的有效连接, 也无法协调教学程序、改进教学方法、作出正确的决策。评价品格是指体育教师在教学评价活动中所应该具备的道德品质。评

价品格隐喻着体育教师从评价的设计、信息的收集到测评结果的解释、运用及交流沟通, 都必须建立在客观公正、合理合法、合乎伦理的基础上^[15]。评价是把双刃剑, 评价向善, 如给予学生评价知情权和参与权, 尊重学生的多样性, 保护学生的评价隐私, 有助于消除学生对评价的恐惧心理, 吸引学生参与到评价中来并利用评价来改善自己的学习^[12]。如果评价在不经意间伤害了学生, 如伤人的评价语言、人与人之间的横向比较, 不仅会对学生的学习产生消极影响^[15], 而且也与《课程标准 2020 年修订版》《课程标准 2022 年版》倡导“关爱学生”的教育评价理念是背道而驰的。因此, 真正实施核心素养导向的评价活动离不开“尊重个体差异”“因材施评”在教学实践上的运用, 即评价品格是体育教师必备的评价素养。

4.2 中小学体育教师评价素养量表具有一定的科学性和可靠性

评价指标的科学性直接决定着测评工具的科学性。本研究首先采用保障评价指标体系准确性、科学性的德尔菲专家咨询法对中小学体育教师评价素养量表进行定性评估。心理测量学指出, 项目分析的主要目的在于检测编制的量表或测试个别条目的适切或者可靠程度^[22], 基于此, 本研究采用临界值比法、题总相关法等对量表题项的适切性进行定量评估。探索性因子分析主要目的是找出影响观测变量的因子个数, 以及各个因子和各个观测变量之间的相关程度, 建立量表的或问卷的结构效度^[32]。对此, 本研究通过主成分分析法减少题项的数目, 找出量表的潜在结构, 将中小学体育教师评价素养量表中具有错综复杂关系的变量综合为 5 个核心因子。验证性因子分析主要目的是检验量表建构效度的适切性与真实性^[32]。本研究通过验证, 发现构建的二阶五因子结构模型各项适配度指标均符合标准值, 而且在模型修正过程中没有删去过多的路径和题项。综上, 多种方法相结合的选择途径不仅从不同角度和目的对题项进行筛选, 而且也保证被选题项的质量和量表的科学性^[33]。

测量恒等性是指测验工具或评量方法适用于不同对象, 或于不同时间地点上使用时, 测验分数或评量结果应具有一定的恒等性^[34]。为此, 本研究借助多群组验证性因素分析考察中小学体育教师评价素养量表跨性别、跨学段、跨地区的测量恒等性。结果发现, 因素负荷、截距、共变异数与测量误差 4 个指标取值均符合理想范围, 这一结果表明跨性别、跨学段、跨地区不会对中小学体育教师评价素养量表的测量结果和稳定性产生影响。相关研究指出, 测量工具时间太长可能会造成测试对象的消极情绪, 进而“抵触”测评过程, 降低测

评的有效性^[35]。通过对多名被测试者进行访谈, 得出中小学体育教师评价素养量表测量题项数量适中, 而且题项长短适宜, 自评完成时间介于 4~6 分钟之间, 易被接受。综上所述, 以上研究方法以及操作过程具有较强的科学性, 进而也为中小学体育教师评价素养量表的科学性提供重要保障。

在本研究中, 内部一致性系数、重测信度、内容效度、收敛效度均在理想值区间。结果表明中小学体育教师评价素养量表具有良好的可靠性和适用性, 可以为我国中小学体育教师评价素养的测量提供工具支持。

4.3 局限与展望

一方面, 由于人力、物力、时间等方面的限制, 本研究仅选择部分地区部分中小学体育教师作为调查对象, 造成样本量有限、选取的调查对象不够宽阔, 因此后续研究可以进一步扩大样本量和样本选择范围, 对中小学体育教师评价素养量表进行修正和验证。另一方面, 本研究主要采用自上而下型的研究思路, 即基于文献调研的方式获取中小学体育教师评价素养量表测评题项, 尽管题项具有一定科学性和适用性, 但是这种思路往往会造成收集到的资料不够全面。因此, 后续研究可通过开门问策、深入实践完善题项内容(如采用半结构式访谈法对中小学体育教师进行访谈), 以使中小学体育教师评价素养量表更好地为体育教师实施核心素养导向的评价活动服务。

参考文献:

- [1] 张细谦, 姚蕾. 体育课程与教学论[M]. 北京: 高等教育出版社, 2021: 252.
- [2] 张健. 新课程标准下体育教学评价的问题与对策研究[J]. 课程·教材·教法, 2015, 35(6): 98-102.
- [3] 郑东辉, 毛玮洁. 论教师评价伦理素养的培育[J]. 教师教育研究, 2024, 36(5): 16-21.
- [4] 范云祥, 马卫平. 体育教学评价中几对范畴的思考[J]. 上海体育学院学报, 2011, 35(4): 70-73.
- [5] 檀慧玲, 王玥, 沈漪佳. 教师评价素养发展的挑战与对策: 基于深化教育评价改革和实施新课标的省思[J]. 教育科学研究, 2023(8): 34-41.
- [6] 盛雅琦, 张辉蓉. 新时代教师评价素养的内涵解构、价值意蕴及测评框架[J]. 课程·教材·教法, 2022, 42(5): 146-152.
- [7] STIGGINS R J. Assessment literacy[J]. Phi Delta Kappan, 1991, 72(7): 534-539.
- [8] YUE T X, GAVIN T L B. Teacher assessment literacy in practice: A reconceptualization[J]. Teaching and Teacher Education, 2016(58): 149-162.

- [9] WEBB N. Assessment literacy in a standards-based urban education setting[C]/In Annual Meeting of the American Educational Research Association, NewOrleans, Louisiana, 2002: 1-5.
- [10] CHAPPUIS J, STIGGINS R J. Stuernt-involved assessment for learning[M]. 7th ed. New York: Person Education Inc, 2017: 11.
- [11] 李顺雨. 初中课堂教师评价素养研究[D]. 大连: 辽宁师范大学, 2016: 22.
- [12] 郑东辉. 新时期教师到底需要怎样的评价素养[J]. 教育发展研究, 2022, 42(4): 46-51.
- [13] EARL L M. Assessment as learning[M]. Thousand Oaks, CA: Corwin, 2013: 26, 27.
- [14] BROOKHAR S M. Educational assessment knowledge and skills for teachers[J]. Educational Measurement: Issues and Practice.2011, 30(1): 3-12.
- [15] 朱传世. 教师评价素养框架设计初探[J]. 中国教师, 2024(5): 35-38.
- [16] 赵雪晶. 我国中学教师教学评价素养研究[D]. 上海: 华东师范大学, 2012.
- [17] 朱永新. 习惯养成是核心素养形成的行动路径——新教育实验推进“每月一事”的理论与实践[J]. 课程•教材•教法, 2017, 37(1): 4-15.
- [18] 邵伟德, 何鲁伟, 邹旭铝, 等. 核心素养融入体育课堂教学的逻辑与策略——以 2022 年版课程标准为视角[J]. 首都体育学院学报, 2023, 35(1): 10-20.
- [19] 罗伯特, 德威利斯. 量表的编制: 理论与应用[M]. 席仲恩, 杜钰, 译. 重庆: 重庆大学出版社, 2018: 68.
- [20] 孟高慧, 刘畅. 大学生学术信息组织水平量表的开发与应用[J]. 图书情报工作, 2021, 65(12): 73-82.
- [21] 孟万金, 苗小燕, 官群, 等. 中国中小学生家庭幸福感量表编制[J]. 教育研究与实验, 2020(3): 91-96.
- [22] 陈华友, 周礼刚, 刘金培, 等. 统计预测与决策[M]. 北京: 科学出版社, 2018: 13.
- [23] 吴明隆. 问卷统计分析实务—SPSS 操作与应用[M]. 重庆: 重庆大学出版社, 2010.
- [24] DOLL W J, XIA W, TORKZADEH G A confiematic factor analysis of the end-user coomputing satisfaction instruction[J]. MIS Quarterly, 1994, 18(4): 453-461.
- [25] 晏骏, 谭广鑫, 李信厚, 等.《粤港澳青少年民族传统体育文化认同量表》的编制与检验[J]. 体育科学, 2023, 27(6): 25-31.
- [26] 史静璋, 莫显昆, 孙振球. 量表编制中内容效度指数的应用[J]. 中南大学学报(医学版), 2012, 37(2): 49-52.
- [27] 教育部. 义务教育体育与健康课程标准(2022 年版)[M]. 北京: 北京师范大学出版社, 2022.
- [28] 季浏. 我国《义务教育体育与健康课程标准(2022 年版)》解读[J]. 体育科学, 2022, 42(5): 3-17.
- [29] 郭洪瑞. 学生评价素养的国外研究及其启示[J]. 中国考试, 2022(12): 70-79.
- [30] 杨国海. 教师评价素养的内涵及框架[J]. 当代教育科学, 2011(4): 11-19.
- [31] 乔玉成. 青少年锻炼习惯的养成机制及影响因素[J]. 体育学刊, 2011, 18(3): 88-94.
- [32] 吴明隆. 结构方程模型—AMOS 的操作与应用[M]. 重庆: 重庆大学出版社, 2009: 212.
- [33] 孙莹, 鲍鑫中. 一种基于方差最大化的组合赋权评价方法及其应用[J]. 中国管理科学, 2011, 19(6): 142-148.
- [34] REISE S P, WIDAMAN K F, PUGH R H. Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance[J]. Psychological Bulletin, 1993(114): 552-566.
- [35] 陈思同, 刘阳, 唐炎. 体育素养测量与评价的现状、挑战及未来[J]. 体育学刊, 2019, 26(5): 111-117.

