**Advances in Education** 教育进展, 2025, 15(5), 169-176
Published Online May 2025 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2025.155737

Hans汉斯

# 师范生教育测量与评价素养现状及培育策略

**何　壮，冯小燕，袁淑莉***

贵阳学院教育科学学院，贵州　贵阳

收稿日期：2025年4月2日；录用日期：2025年5月1日；发布日期：2025年5月8日

## 摘　要

**本研究以师范生教育测量与评价素养为对象，通过文献分析、问卷调查，构建包含评价意识、态度、知识、能力的四维测量框架，探讨其现状及影响因素。研究发现，师范生评价素养整体处于中等水平，评价知识维度得分最低，理科生、参与过命题实习或修读相关课程的学生表现更优。研究结果揭示了课程体系、实践机会与数学基础对评价素养的关键作用，为师范教育课程改革和实践优化提供了实证依据，对提升教师职前培养质量和推动教育评价改革具有重要意义。**

## 关键词

**教育测量与评价，师范生，评价素养**

# The Current Situation and Cultivation Strategies of Educational Measurement and Evaluation Literacy of Normal Students

**Zhuang He, Xiaoyan Feng, Shuli Yuan***

School of Education Science, Guiyang University, Guiyang Guizhou

Received: Apr. 2nd, 2025; accepted: May 1st, 2025; published: May 8th, 2025

## Abstract

**This paper focuses on the assessment literacy of pre-service teachers, using literature analysis, questionnaire surveys, and empirical research to construct a four-dimensional measurement framework covering assessment awareness, attitude, knowledge, and ability. The study finds that pre-service teachers' assessment literacy is generally at a moderate level, with the lowest scores in the knowledge dimension. Science-major students, those who participated in proposition practice, or took relevant courses performed better. The results highlight the critical roles of curriculum systems,**

*通讯作者。

文章引用: 何壮, 冯小燕, 袁淑莉. 师范生教育测量与评价素养现状及培育策略[J]. 教育进展, 2025, 15(5): 169-176.
DOI: 10.12677/ae.2025.155737

**practical opportunities, and mathematical foundations in shaping assessment literacy. This research provides empirical evidence for curriculum reform and practical optimization in teacher education, offering significant implications for improving pre-service teacher training quality and promoting educational assessment reform.**

## Keywords

**Educational Measurement and Evaluation, Normal School Students, Assessment Literacy**

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

## 1. 引言

教师的教育测量与评价素养简称评价素养。教师评价素养作为教师专业素养的关键一环，对于提高教学质量、促进学生全面发展具有至关重要的作用。随着教育评价改革的不断深化，教师评价素养的研究逐渐受到广泛关注。在这一背景下，对教师评价素养的研究不仅有助于提升教师的专业素养，还能为学生提供更为优质的教育服务。师范教育是教师职前教育最核心阶段，师范教育阶段要顺应时代的变革，系统的培养学生的评价素养。

## 2. 研究综述

### 2.1. 评价素养的概念

教师评价素养是指教师选择并运用适当的评价方式收集、分析和使用评价信息促进学习、改进教学的意向、知识和能力[1]。这一概念强调了评价在教学过程中的核心作用，不仅关注评价结果，更重视评价过程对教学和学习的影响。教师评价素养的特点在于其全面性和动态性，强调评价应贯穿于教与学的始终，促进学生的全面发展。

### 2.2. 评价素养标准

评价教师的评价素养水平，首先要参照一定的标准。国外研究中影响力较大的评价素养标准有：美国教师联合会、全国教育测量委员会等共同提出的“1990 年标准”，美国新教师评价与支持联盟提出的“InTASC 标准”，欧洲教育评价联合会提出的“AEA-E 评价标准”。

#### 2.2.1. 1990 年标准

美国 1990 年的教师评价素养标准是较早提出的教师评价素养框架之一。该标准主要包括七条内容，涵盖了选择适合教学决策的评价方法、管理评价结果、解释评价数据、设计评价工具、实施评价程序、沟通评价结果以及使用评价结果改进教学等方面。这些标准旨在确保教师能够有效地运用评价手段来指导教学实践，提高教学质量。

这一标准主要侧重于终结性评价知识和技能的培养，强调教师在评价结果的解读和应用方面的能力。然而，它相对缺乏对学生参与的形成性评价的涵盖，这在一定程度上限制了其在促进学生全面发展方面的作用。尽管如此，该项标准仍然是早期有关教师评价素养实证研究开发工具的重要依据。

#### 2.2.2. InTASC 标准

美国新教师评价与支持联盟 InTASC (Interstate New Teacher Assessment and Support Consortium)标准

是近年来提出的一个更为全面的教师评价素养框架。该标准涵盖了学习者与学习、教学内容知识、教学实践和专业责任四个重点领域，旨在全面提升教师的评价能力。

InTASC 标准强调以学习者为中心，提出了评价素养的三要素：评价表现、评价知识和评价伦理。这一标准不仅关注教师的评价技能，还注重教师的评价态度和价值观，强调评价的公平性和伦理性。该标准对后续国内研究者开展实证研究、开发评价工具产生了深远影响。

#### 2.2.3. AEA-E 标准

欧洲 AEA-E (Association of Educational Assessment-Europe, AEA-E)教育评价标准框架由七个核心要素组成，包括评价目标的确定、评价任务的制定、评价方法的选择、评价数据的收集与分析、评价结果的反馈与交流、评价过程的监控与调整以及评价伦理与责任。AEA-E 标准框架强调评价工具与评价标准的整体性，具有较强的操作性。它不仅为教师提供了具体的评价指导，还强调了评价过程中的反思与改进，有助于教师持续提升评价素养。AEA-E 标准提出的时间最晚，因此，该标准融合了前人研究的优点，并在标准的基础上更进一步，提供了评价工具。配套的评价工具由评价标准、评价方法、评价证据三部分构成。“评价标准”说明需要规范的评价行为表现，“评价方法”提供评价方法的使用建议，“评价证据”对评价中需要的证据的样态和类型作了举例说明。AEA-E 标准是三个标准中最具可操作性的，为国内相关研究提供了参考。

### 2.3. 评价素养测量工具

国内研究者对主流的教师评价素养评价工作做了详细评述，孙佳林等人认为，当前国外主流的评价工具主要分为三种类型：旨在评价教师评价知识的客观测试工具，教师对自身评价素养进行自我诊断的工具，以及评价教师评价素养的工具[2]。

#### 2.3.1. 客观测试工具

评价素养问卷 TALQ (Teacher Assessment Literacy Questionnaire)是基于 1990 年标准开发的一种客观测试工具，主要用于评估教师的评价素养。该测试涵盖了评价方法的选择、评价结果的管理等方面，通过选择题和简答题等形式来测量教师的评价能力。

评价素养量表 ALI (Assessment Literacy Inventory)是基于 TALQ 改编的一种评价工具，专门针对职前教师和在职教师。ALI 在 TALQ 的基础上进行了扩展和深化，增加了对教师评价实践的评估内容，能够更全面地反映教师的评价素养。

#### 2.3.2. 自我诊断工具

评价实践清单 API (Assessment of Practices Inventory)是一种评价实践清单，主要用于评估教师对自身评价技能的自信程度。该清单包含多个条目，教师需要根据自己的实际情况进行评分，从而了解自己在各个评价环节中的优势和不足。

评价观念清单 CoA (Conceptions of Assessment Inventory)是一种评价观念清单，主要用于评估教师对评价目的的认识。该清单通过一系列问题来引导教师反思自己的评价理念和实践，帮助教师树立正确的评价观念。

#### 2.3.3. 评价工具

课堂评估清单 ACAI (Assessment of Classroom Assessment Inventory)是一种教师课堂评价工具，涵盖了评价方法、评价实践信心调查等多个方面。该工具通过问卷调查的形式来收集数据，能够全面反映教师的评价素养水平。

### 2.4. 国内评价素养的实证研究

国内最早的教师评价素养实证研究可以追溯到2008年，郑东辉等人使用改编后的教师评价素养问卷对某省份四个地区的教师开展了问卷调查。研究者结果表明中小学教师的评价素养现状不容乐观，存在较多问题：教师的评价知识缺乏，评价技能较低，评价方法单一，对评价的理解不到位，评价实践功利化倾向严重。

最新的研究聚焦在高中教师群体，杨卓等人构建了涵盖评价理念、评价知识和评价技能三个一级指标的教师评价素养框架，并依据框架制定了高中教师评价素养调查问卷。团队使用该问卷在全国范围内抽样，结果表明当前我国高中教师整体评价素养表现较好，但评价知识相对薄弱，评价理念显著优于评价技能[3]。影响高中教师评价素养水平的主要因素包括评价素养价值认同、性别、地区、学校层次等。

国内学者对评价素养的研究逐渐增多，关注的群体以在职教师为主。职前教育作为师范教育的重要环节，尚未有学者关注到这一群体的评价素养。因此，本研究以这一群体为研究对象，探讨该群体的评价素养现状及培育策略。

## 3. 研究设计

### 3.1. 评价框架

本研究对评价素养的测量框架设计从师范教育、教师专业化的视角展开，评价素养的构成要素包括评价意识、态度、知识、能力，共计四个维度。根据David McClelland的胜任力素质模型，能力和知识属于外显部分，态度和意识属于内隐部分。评价意识是指评价者能够认识到评价的意义和价值，主动将评价作为一种工具，有目的的支配评价行为。评价态度指的是评价者对外部评价(如统考、高考)和内部评价(如自己组织的课堂评价等)的内在心理倾向。评价知识是指评价者开展评价活动的基本知识，如标准化测验、评价流程与方法、教育目标分类理论等。评价能力是评价者综合运用各类评价知识后表现出来的实践行为，如评价方案设计、评价解释及反馈、评价结果应用等。

### 3.2. 评价工具

本研究参照顾咸霜[4]针对教师群体的研究中使用的量表，将其修订为针对师范生的量表。修订后的师范生评价素养量表共4个维度，评价意识维度6题，评价态度维度5题，评价知识维度6题，评价能力维度7题，共计24题。题目内容为反映评价素养各种行为的陈述句，题目采用李克特5级量表，1代表完全不符合、5代表完全符合，得分越高代表评价素养越高。

量表在本研究中的$\alpha$系数为0.945，分半信度为0.938，信度数据均大于相关研究建议的标准，说明信度良好。使用结构方程模型法进行效度验证，卡方与自由度之比为3.668，CFI = 0.914，TLI = 0.903，RMSEA = 0.093，效度数据均达到相关研究建议的标准，效度良好。

### 3.3. 抽样对象

本研究在贵州省某高校的11个师范类专业中进行了抽样，共收集有效问卷312份，涵盖了学前、小教、心理学、语数外等学科教学相关师范类专业。评价素养的培养需要经过多门课程，还要在实习环节有针对性的开展评价实践，因此，本次抽样仅针对参加过实习的大三或大四学生。

## 4. 师范生评价素养现状

### 4.1. 师范生评价素养发展水平

为了更直观的报告被试群体在各维度和总分上的得分水平，以描述师范生评价素养的发展水平，本

研究引入了因子分的概念，即维度均分除以题目数量。以评价意识维度为例，该维度的平均分为 21.15，共有 6 个题目，因子分为 21.15/6 = 3.53。因子分直观地表现了该题在李克特量表 1~5 个等级中所处的位置。结果显示，抽样群体的评价素养总分因子分为 3.69，处在“一般”和“比较符合”之间，为中等偏上水平；四个维度因子分均处在这一区间，得分依次为评价态度 3.9、评价能力 3.71、评价意识 3.53、评价知识 3.16。

### 4.2. 师范生评价素养的影响因素

进一步对可能影响师范生评价素养的因素进行分析，数据显示性别、年级、实习方式(集中/分散)、实习学段(幼儿园/小学/中学)等因素上，不同群体学生之间得分无显著差异。但人才培养的课程体系中“是否开设了研究方法课程”、“是否开设了教育测量与评价相关课程”“生源类型”“实习期间是否参与过学校的评价活动”等因素上，不同群体学生的得分存在显著差异。

不同招生类型学生的均分比较结果如表 1 所示，高中阶段为理科生的学生群体在四个维度和总分上的得分均显著高于文科生。结果表明高中阶段的数学学习是大学阶段评价学习的基础。同时，物理、化学、数学等师范类专业，虽为师范类，但招生只面向理科生。这些专业的人才培养方案中设置了更多的高等数学、统计学等课程，课程体系的差异进一步强化了评价素养的差异。

**Table 1.** Influence of admission type on assessment literacy
**表 1.** 招生类型对评价素养的影响

| 维度 | 招生类型 | 平均分 | 标准差 | t |
|---|---|---|---|---|
| 评价意识 | 理科 | 21.72 | 4.47 | 3.946** |
| | 文科 | 19.09 | 5.99 | |
| 评价态度 | 理科 | 20.23 | 3.70 | 5.601** |
| | 文科 | 16.93 | 5.94 | |
| 评价知识 | 理科 | 22.77 | 4.10 | 4.601** |
| | 文科 | 19.72 | 6.77 | |
| 评价能力 | 理科 | 22.82 | 4.17 | 3.781** |
| | 文科 | 20.34 | 6.45 | |
| 总分 | 理科 | 91.12 | 15.40 | 4.693** |
| | 文科 | 79.52 | 25.05 | |

注：*代表 $p < 0.05$，**代表 $p < 0.01$，下同。

从课程开设的角度进行分析，发现开设了教育研究方法类课程的专业在评价态度维度上得分显著高于未开设该课程的专业，如表 2 所示。以测验法为代表的研究方法是该门课程的核心内容，进一步分析相关课程的教学大纲，可以发现这些课程中有关测验法的讲授内容主要是标准化测验。因此，修读了该课程，了解标准化测验的学生更倾向于使用组织测验的方法检验教学效果。但受限于课程上讲授的深度，学生的素养提升也仅仅体现在态度上。

**Table 2.** Influence of availability of research methods courses on assessment literacy
**表 2.** 是否开设研究方法课程对评价素养的影响

| 维度 | 研究方法课程 | 平均分 | 标准差 | t |
|---|---|---|---|---|
| 评价意识 | 是 | 21.53 | 3.68 | 1.085 |
| | 否 | 20.91 | 5.67 | |

续表

| | | | | |
|---|---|---|---|---|
| 评价态度 | 是 | 20.41 | 3.35 | 2.954** |
| | 否 | 18.91 | 5.04 | |
| 评价知识 | 是 | 22.23 | 3.66 | 0.353 |
| | 否 | 22.03 | 5.69 | |
| 评价能力 | 是 | 22.52 | 3.58 | 0.702 |
| | 否 | 22.13 | 5.57 | |
| 总分 | 是 | 90.14 | 12.95 | 1.19 |
| | 否 | 87.61 | 21.51 | |

进一步聚焦教育测量与评价相关内容的学习，部分专业将相关知识融入研究方法等课程，部分专业将其作为专门课程开设。数据结果表明，开设专门课程的专业在评价态度维度上的得分显著高于其他专业。在另外三个维度上，虽然平均得分并无显著差异，但开课的专业得分也呈现了高于未开课专业的趋势。这种累积效应导致在总分上出现了显著差异，如表 3 所示。

**Table 3.** Influence of availability of educational assessment courses on assessment literacy
**表 3.** 是否开设教育评价课程对评价素养的影响

| 维度 | 教育测评课程 | 平均分 | 标准差 | t |
|---|---|---|---|---|
| 评价意识 | 是 | 21.60 | 3.17 | 0.549 |
| | 否 | 21.11 | 5.13 | |
| 评价态度 | 是 | 21.63 | 2.14 | 2.981** |
| | 否 | 19.26 | 4.63 | |
| 评价知识 | 是 | 22.94 | 3.12 | 1.049 |
| | 否 | 22.01 | 5.13 | |
| 评价能力 | 是 | 23.03 | 3.22 | 0.953 |
| | 否 | 22.20 | 5.01 | |
| 总分 | 是 | 92.34 | 9.59 | 2.087** |
| | 否 | 88.18 | 19.29 | |

实习期间的工作经历也有可能对评价素养产生正向促进作用。数据分析结果显示，实习期间参与学校命题工作的学生在评价意识、评价态度、评价知识和总分上均显著高于未参与的学生如表 4 所示。命题是一项整合了学科、评价、管理等多项内容的复杂工作，参与命题是对学生评价素养的全面历练。

**Table 4.** Influence of participation in exam proposition on assessment literacy
**表 4.** 是否参与命题对评价素养的影响

| 维度 | 参与命题 | 平均分 | 标准差 | t |
|---|---|---|---|---|
| 评价意识 | 未参与 | 19.69 | 5.85 | −3.038** |
| | 参与过 | 21.64 | 4.53 | |

续表

| | | | | |
|---|---|---|---|---|
| 评价态度 | 未参与 | 18.36 | 5.13 | −2.632** |
| | 参与过 | 19.90 | 4.19 | |
| 评价知识 | 未参与 | 21.10 | 5.80 | −2.058** |
| | 参与过 | 22.44 | 4.61 | |
| 评价能力 | 未参与 | 21.48 | 5.64 | −1.687 |
| | 参与过 | 22.55 | 4.55 | |
| 总分 | 未参与 | 84.19 | 21.74 | −2.435** |
| | 参与过 | 90.07 | 17.11 | |

### 4.3. 结论与讨论

系统分析上述数据分析结论，可以发现学生群体的评价素养发展具备以下特征。

首先，学生的评价素养总体水平和各维度的水平都仅处在中等。并且，这一结论还是在主观自评的情境下得出的。评价素养是师范生未来就业必须具备的，因此，主观自评数据可能已经存在一定的社会称许效应。实际情况可能低于中等水平。

其次，各维度之间的得分并未呈现出较大差异，从因子分上来看，得分相对更高的是评价态度，最低的是评价知识。评价态度属于评价素养的内隐部分，得分最高的现状说明，学生在经过系统学习评价相关课程和实习期间评价体验之后，对评价活动有了更科学的认识，愿意在教学活动中采取评价的方式评估效果。评价知识属于评价素养的外显部分，得分最低，说明学生对自己评价相关知识的掌握情况并不自信，这可能与课程学习和实习过程中实践机会较少有关。

最后，数据显示数学知识是培养评价素养的重要基础。高中阶段或本科课程中系统学习基本数学知识的学生群体表现出了更好的评价素养，实习期间参与评价活动也能显著促进评价素养的发展。应用型大学在本科阶段的评价素养培养环节，对基础的数学知识要求并不高，但受限于师资和生源等多种因素，很多专业的课程体系都忽略了数学课程。在实习环节，受到实习时间和实习学校分配的岗位等因素影响，学生参与命题、考试数据分析等环节的机会较少。多重要素叠加，也影响了师范生评价素养的发展。

## 5. 师范生评价素养培育建议

### 5.1. 一贯式的培育视角

从教师教育全过程的角度来看，师范生处在教师专业发展的“职前”阶段。一名优秀教师的成长还将经过“入职”和“职后”阶段。师范生的评价素养培育需要面向毕业后 5 年甚至更长时间的评价素养发展需求，建立横向贯通、纵向协同的机制。横向贯通需要整合心理学、教育学、学科课程与教学论、数学、统计学、计算机等课程，强化评价设计和评价数据分析能力；纵向协同需要系统规划从大一到大四、从理论到实践、从见习到实习的素养培育路径，低年级注重评价意识、中年级注重评价知识、高年级实践评价知识、实习阶段锻炼评价能力。

### 5.2. 教学内容

研究结果表明评价知识是最弱的板块，对各师范专业各课程教学大纲分析后发现，评价素养的教学内容尚未形成体系。教学内容的系统化设计是师范生评价素养发展的重要基石，根据相关教学内容的特点可以将教学内容体系重构为“基础”“特色”“创新”三个层次。布鲁姆为代表的教育目标分类理论、

必要的数学和统计学、统计软件操作等内容为基础内容；各专业结合本学科的课程与教学论、命题技术、研究设计原理等课程为特色内容；增值评价原理、人工智能技术应用、教育大数据挖掘原理、教育评价的伦理规范等课程为创新内容。科学的教学内容设计可以帮助师范生丰富知识的广度与深度，为职业发展奠定基础。

### 5.3. 课程体系

研究结果凸显了当前人才培养方案中与评价素养有关内容缺乏系统性和完整性。各学科受到师资、实验室等人才培养资源限制，课程体系中普遍缺少《教育测量与评价》《课程与教学论》和 SPSS 等常用统计软件操作课程。加之文科招生为主的师范专业，学生数学基础相对较弱，课程体系的短板对其评价素养的发展影响更加明显。

### 5.4. 教学实践

教育评价对设计和操作能力要求很高，相应的，在师范教育阶段需要为学生提供更多的实践机会。研究结果表明，实习期间参与学校的教育评价活动对评价素养的发展有促进作用。但受到实习学校分工的影响，学生能接触到的评价活动以课后作业、随堂测验等形式为主。学生直接参与学校的考试命题、过程性评价、增值评价等项目的机会有限。针对这一问题，高校可以通过项目合作、资源库建设等方式来补充。通过这种方式，让学生在模拟的教学环境中不断练习和提升评价技能。

师范生评价素养是未来从教之后实践教育评价改革政策的重要基础，为此需要高校从培育视角、教学内容、课程体系、教学实践等方面开展教学内容和课程体系的改革，助力师范生素养发展，助力师范教育高质量发展。

## 6. 结语

本研究系统揭示了师范生教育测量与评价素养的发展现状与关键影响因素，在已有研究基础上构建了四维评价框架并验证了其科学性。研究发现，师范生评价素养整体处于中等水平，评价知识短板明显，理科背景、课程学习与实践经历对素养提升具有显著促进作用。进而提出了培育策略从贯通式培养体系、课程体系优化与实践能力强化三方面切入，为师范教育改革提供了实证依据与路径参考。师范生的教育评价素养是我国教育评价改革落地的重要影响因素，未来团队将持续开展更全面的教育测量与评价相关课程追踪研究，为教师职前培养质量提升提供更系统的理论支撑。

## 致　谢

感谢贵阳学院教育科学学院 2021 级应用心理学班、2024 级教育硕士班全体同学在本研究过程中的贡献。

## 基金项目

贵州省高等学校教学内容和课程体系改革项目：“会测善用”导向的师范生教育测量与评价课程设置与建设实践(2023250)，贵阳学院微专业建设项目“教育与心理测评大数据”。

## 参考文献

[1] 郑东辉. 美国教师评价素养研究述评[J]. 全球教育展望, 2011, 40(6): 46-51+57.

[2] 孙佳林, 郑长龙. 国外教师评价素养标准及评价工具研究述评[J]. 中国考试, 2023(3): 77-85.

[3] 杨卓, 李佳, 刘惠玲, 等. 我国高中教师评价素养现状调查研究[J]. 中国考试, 2024(5): 57-68.

[4] 顾咸霜. 专业化视域下中小学教师评价素养研究[D]: [硕士学位论文]. 重庆: 重庆师范大学, 2023.