|  |  |
| --- | --- |
| ICS | 35.240.99 |
| CCS | |  | | --- | | D:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T.pngD:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T后面的反斜杠.png XZBX |   L 67 |

西安市质量与标准化协会团体标准

T/XZBX 0092—2025

多模态人工智能教育动态测评体系

构建指引

Guideline for construction of dynamic evaluation system of multimodal artificial intelligence education

2025 - 09 - XX发布

2025 - 09 - XX实施

西安市质量与标准化协会  发布

目次

[前言 III](#_Toc207697615)

[引言 V](#_Toc207697616)

[1 范围 1](#_Toc207697617)

[2 规范性引用文件 1](#_Toc207697618)

[3 术语和定义 1](#_Toc207697619)

[4 构建原则 2](#_Toc207697620)

[5 技术框架 3](#_Toc207697621)

[6 功能模块 4](#_Toc207697622)

[7 数据采集处理 5](#_Toc207697623)

[8 动态测评方法 6](#_Toc207697624)

[9 教育应用场景 7](#_Toc207697625)

[10 质量保障与持续改进 8](#_Toc207697626)

1. 前言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由西安市质量与标准化协会提出并归口。

本文件起草单位：上海中侨职业技术大学。

本文件主要起草人：朱 江。

1. 引言

随着人工智能技术在教育领域的广泛应用，传统以考试为核心的静态测评模式已难以全面反映学习者的真实水平和发展潜力。多模态人工智能教育动态测评体系通过整合文本、语音、图像、视频、行为数据等多种信息源，能够实现对学习过程的实时感知、动态追踪和智能分析。这种测评体系不仅可以更全面地揭示学习者的认知能力、情感状态与行为习惯，还能够为个性化教学提供即时反馈和科学决策依据。

当前，国内外在人工智能教育测评领域的探索不断深入，但在数据采集、模型构建、指标体系设计、隐私保护和教育场景适配等方面仍缺乏统一的规范与指导，导致不同平台和机构之间存在测评结果不一致、数据不可比和难以共享等问题。因此，亟需建立一套科学、规范、可操作的多模态人工智能教育动态测评体系构建指引。

本文件的制定，旨在规范多模态数据的采集与处理流程，明确动态测评的原则与方法，提出技术框架与应用场景的构建思路，为教育管理者、教师、科研机构和相关企业提供参考，推动教育测评由静态走向动态、由单一走向多元，促进教育公平与质量提升。

多模态人工智能教育动态测评体系

构建指引

* 1. 范围

本文件规定了多模态人工智能教育动态测评体系的构建原则、技术框架、功能模块、数据采集处理、动态测评方法、教育应用场景及质量保障与持续改进等内容。

本文件适用于各类教育机构、教育技术企业和科研单位在多模态人工智能教育测评体系的设计、开发、应用与管理过程中参考使用。

本文件不适用于单一模态的传统考试测评或非教育领域的人工智能测评体系。

* 1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 35273—2020 信息安全技术 个人信息安全规范

GB/T 36352—2018 信息技术 学习、教育和培训 教育云服务：框架

GB/T 41867—2022 信息技术 人工智能 术语

GB/T 43466—2023 信息技术 学习、教育和培训 中小学生信息素养评价指南

* 1. 术语和定义

下列术语和定义适用于本文件。

多模态数据 multimodal data

来源于文本、语音、图像、视频、传感器及学习行为记录等多种信息源的综合数据。

动态测评 dynamic assessment

以过程性、发展性为导向，通过实时数据采集与分析，持续跟踪学习者表现并提供即时反馈的测评方式。

个性化反馈 personalized feedback

基于学习者个体特征与测评结果，实时生成有针对性的学习建议或干预措施。

可解释性 interpretability

人工智能模型在教育测评中的运作逻辑、结果形成过程以及决策依据能够被教师、学习者和管理者理解的能力。

* 1. 构建原则
     1. 科学性

测评体系必须以教育学、心理学和人工智能的科学理论为基础，保证测评工具的信度与效度。

这一原则要求测评方法来源于可靠的研究成果，并在实践中经过验证，确保数据与结果真实、准确、可重复。

* + 1. 系统性

测评应覆盖学习者的知识掌握、能力发展、情感态度和学习行为等多个维度，形成全过程、全要素的系统化评价。

这意味着体系设计不应仅限于单点测量，而是要通过多模态数据的融合实现全面而连续的跟踪。

* + 1. 动态性

体系应强调发展性与过程性，能够反映学习者在不同时间点的变化和进步情况。

这要求测评不仅提供结果性的总结，还要关注学习过程中的即时表现，便于教师进行适时干预。

* + 1. 个性化

测评体系应重视个体差异，能够针对不同学习者的特点生成差异化的分析结果与反馈。

该原则强调人工智能技术在测评中的优势，应通过算法与模型的智能分析，帮助教师因材施教。

* + 1. 公平性

体系建设必须遵循教育公平原则，避免因算法偏差或数据不平衡导致结果失真。

这不仅要求在数据采集和模型训练时保证样本的多样性与代表性，也要求在应用过程中关注弱势群体的测评公正。

* + 1. 安全与隐私

测评体系应符合国家和国际信息安全与个人隐私保护要求。

这意味着在数据采集、存储和应用环节均需采取严格的安全措施，防止数据泄露与滥用。

* + 1. 可解释性

人工智能在教育测评中的应用必须具备透明度和可解释性，使教师、学习者和教育管理者能够理解模型的运行逻辑和结论依据。

该原则有助于提升测评体系的信任度与教育应用价值。

* 1. 技术框架
     1. 数据层

数据层是整个体系的基础，承担着学习过程多模态信息的采集任务。为了确保数据的全面性和真实性，应遵循以下要求：

1. 数据来源应包括文本、语音、图像、视频、传感器信号以及学习行为记录；
2. 采集方式应符合标准化流程，保证数据的准确与完整；
3. 对于敏感数据，应采取匿名化或脱敏措施。
   * 1. 处理层

原始数据往往存在噪声、缺失和异构问题，需要在处理层进行预处理和融合。该环节的目标是提升数据质量，使其能够为后续建模提供可靠支撑。具体包括：

1. 对采集数据进行清洗，剔除无效与错误信息；
2. 将不同来源的数据进行格式统一与标准化处理；
3. 建立规范的数据标注体系，为模型分析提供高质量样本。
   * 1. 模型层

模型层是实现智能测评的核心，依靠人工智能算法对学习者进行多维度建模。为了保证模型分析的科学性和可解释性，应注重以下方面：

1. 构建认知、情感与行为多维模型，全面刻画学习者特征；
2. 模型应支持动态更新，能够实时反映学习者状态变化；
3. 模型应具备可解释性，避免因“黑箱”效应降低测评的透明度。
   * 1. 应用层

应用层是测评体系与教育实践结合的关键环节，负责将分析结果转化为教育价值。为了确保结果能被有效利用，应做到：

1. 输出个性化反馈，帮助教师精准施教，帮助学习者改进方法；
2. 提供预警与干预功能，发现潜在学习问题并及时应对；
3. 为教育管理者提供群体性数据支持，辅助决策制定。
   * 1. 安全层

安全层贯穿整个体系，确保数据和模型运行的合规性与可靠性。为此，应从以下几个方面加强管理：

1. 建立严格的数据访问权限和身份认证机制；
2. 采用数据加密与安全传输方式，防止泄露与篡改；
3. 遵循国家和国际隐私保护标准，确保数据使用的合法性和合规性。
   1. 功能模块
      1. 数据采集模块

数据采集是测评体系的起点，其质量直接决定后续分析的科学性与可信度。为了保证采集过程规范，应满足以下要求：

1. 采集范围应覆盖文本、语音、图像、视频和学习行为等多模态数据；
2. 数据来源应包括学习管理系统、在线课堂、智能终端和传感器等；
3. 数据获取方式应遵循合法合规原则，避免过度采集。
   * 1. 数据处理模块

原始数据往往存在噪声、冗余和异构问题，需要通过处理模块提升数据质量。主要任务包括：

1. 对数据进行清洗，剔除错误和缺失信息；
2. 统一数据格式，确保不同来源的数据能够融合使用；
3. 按照既定标准进行标注，为模型训练提供支持。
   * 1. 模型分析模块

模型分析是实现智能化测评的核心环节，通过人工智能算法对学习者的多维特征进行刻画。其功能应包括：

1. 构建认知分析模型，识别学习水平和知识掌握情况；
2. 构建情感识别模型，监测学习过程中的情绪状态；
3. 构建行为分析模型，追踪学习习惯与互动特征。
   * 1. 测评反馈模块

反馈模块将分析结果转化为可操作的教育建议，是测评体系发挥价值的重要环节。其功能应包括：

1. 根据结果生成个性化学习报告，突出学习优势与不足；
2. 为教师提供实时预警信息，帮助其调整教学策略；
3. 向学习者推送个性化改进建议，实现因材施教。
   * 1. 管理监控模块

为了保证测评体系的整体运行效率与教育管理价值，需要设置管理监控模块。其主要任务包括：

1. 监控测评系统的运行状态，确保稳定可靠；
2. 为教育管理者提供群体性学习数据，支持宏观决策；
3. 定期生成系统运行报告，用于质量评估与改进。
   * 1. 安全合规模块

安全与合规是多模态测评体系建设的底线。为避免数据滥用和算法偏差，该模块需具备以下功能：

1. 建立权限控制机制，确保数据访问安全；
2. 采用加密存储和传输技术，防止信息泄露；
3. 建立伦理审查机制，避免因算法偏差引发教育不公平。
   1. 数据采集处理
      1. 数据采集

数据采集是测评体系的起点，决定了信息的完整性与代表性。为了保证数据的全面性与规范性，应关注以下方面：

1. 数据来源应覆盖学习者的文本、语音、图像、视频、传感器信号及学习行为记录；
2. 采集方式应符合合法合规要求，并取得学习者或监护人的知情同意；
3. 采集过程中应保持采样频率与精度合理，避免过度采集或遗漏关键信息。
   * 1. 数据处理

原始数据往往存在噪声、冗余和缺失等问题，需要通过科学处理确保数据质量。主要工作包括：

1. 数据清洗：剔除无效或错误数据，修复缺失值；
2. 数据标准化：统一不同来源和格式的数据，使其具备可比性；
3. 数据融合：对多模态数据进行特征提取与融合，形成统一的输入格式；
4. 数据标注：按照统一的标签体系进行标注，确保训练数据的一致性与高质量。
   * 1. 数据存储与管理

为保证数据的可追溯性和安全性，实验室或平台应建立完善的数据管理制度：

1. 建立数据仓库或数据库，支持结构化与非结构化数据存储；
2. 建立访问权限管理制度，确保只有授权人员能够使用核心数据；
3. 定期备份与日志记录，避免因系统故障导致数据丢失；
4. 对长期存储的数据进行加密与脱敏处理，确保隐私安全。
   * 1. 数据类型与处理要点

在教育动态测评中，不同模态数据具有不同特点，其处理要点各不相同。常见数据类型及处理要点见表1。

1. 多模态数据类型与处理要点

| 数据类型 | 数据来源 | 处理要点 |
| --- | --- | --- |
| 文本数据 | 作业答案、课堂对话、学习日志 | 分词、去噪、情感分析，提取知识点关联 |
| 语音数据 | 课堂发言、口语练习、讨论录音 | 降噪、语音识别、韵律特征提取 |
| 图像数据 | 手写作业、板书、实验操作 | 图像增强、特征提取、目标识别 |
| 视频数据 | 上课录像、实验演示、学习行为 | 动作识别、姿态分析、行为模式建模 |
| 行为数据 | 登录记录、点击操作、学习路径 | 时间序列建模、行为模式分析、特征聚合 |
| 传感数据 | 心率、眼动、脑电等可穿戴设备 | 信号滤波、特征提取、与学习状态的相关建模 |

该表对不同数据类型的处理要点进行了总结，为教育测评系统在数据层面的实现提供了参考。

* 1. 动态测评方法
     1. 过程性测评

过程性测评关注学习全过程中的表现，能够全面记录学习者的进步轨迹。

通过采集作业、课堂互动和学习日志等数据，持续追踪学习者表现。强调学习过程中的努力程度和改进方向，而不仅是最终成绩。

可用于发现学习中的薄弱环节，为个性化教学提供支持。

* + 1. 形成性测评

形成性测评强调在学习过程中给予即时反馈，帮助学习者及时调整学习策略。

利用人工智能分析学习者答题和互动情况，提供实时诊断。生成个性化反馈报告，帮助学习者理解错误原因。

教师可根据反馈结果及时调整教学方法，实现教学与测评一体化。

* + 1. 预测性测评

预测性测评注重通过已有数据预测学习者未来表现，具有前瞻性。

通过建模学习行为模式，预测学习者可能出现的困难或风险。结合历史数据，推测学习者在未来任务中的表现趋势。

为教师提供预警，帮助其提前制定干预措施。

* + 1. 综合性测评

综合性测评是多模态测评方法的整合应用，强调全面性和多维度评价。

将文本、语音、视频、行为等多模态数据结合分析，形成学习者的完整画。兼顾认知、情感和行为三个维度，避免单一数据带来的片面性。

可为教育管理者提供群体性、系统性的教育质量分析结果。

为便于对比和选择，常见动态测评方法及其适用场景总结见表2。

1. 动态测评方法与适用场景

| 方法类型 | 主要特点 | 适用场景 |
| --- | --- | --- |
| 过程性测评 | 强调学习全过程数据的收集与跟踪 | 学习轨迹分析、能力发展评估 |
| 形成性测评 | 实时反馈，强调即时诊断与改进 | 课堂互动、作业批改、个性化教学 |
| 预测性测评 | 基于数据建模，预测未来表现趋势 | 学习预警、干预措施设计 |
| 综合性测评 | 融合多模态数据，多维度全面分析 | 学习者画像、教育质量整体评估 |

表2显示了不同方法的特点与应用方向，教育机构可根据具体需求选择合适的动态测评方式，或将多种方法结合使用，以提升测评的科学性与实用性。

* 1. 教育应用场景
     1. 课堂教学

课堂教学是动态测评体系应用的核心场景之一。通过对课堂多模态数据的实时分析，教师可以更直观地掌握学生的学习状态。主要应用表现为：

1. 系统可实时记录课堂发言、互动频率和作业表现；
2. 教师可根据即时反馈调整教学内容和节奏；
3. 数据分析结果能够揭示知识掌握中的薄弱环节，便于针对性辅导。
   * 1. 个性化学习

在个性化学习中，动态测评体系通过学习者画像和行为分析，实现差异化支持。其价值体现在以下几个方面：

1. 系统可依据学习风格和学习轨迹，推荐个性化学习资源；
2. 学习者可获得差异化反馈，及时调整学习方法；
3. 教师和家长可通过数据监控学习进展，支持因材施教。
   * 1. 教育管理

教育管理层面需要宏观视角，动态测评体系可以通过大规模数据分析为政策制定和资源配置提供依据。具体功能包括：

1. 提供整体教育质量报告，支持教育政策的科学制定；
2. 揭示不同学校或班级之间的差距，为教育资源均衡配置提供参考；
3. 跟踪教育改革举措的实施成效，形成数据驱动的管理机制。
   * 1. 教育科研

在科研探索中，体系为研究者提供了丰富的多模态数据资源，可推动教育学与人工智能的融合创新。其主要应用包括：

1. 基于大数据分析，发现学习过程中的潜在规律和模式；
2. 支持新型测评工具、教学方法和教育干预策略的研发与验证；
3. 促进跨学科融合研究，推动教育科学与智能技术的协同发展。

为了便于理解不同场景下的应用价值，常见教育应用场景及支持功能总结见表3。

1. 教育应用场景与支持功能

| 应用场景 | 主要作用 | 支持功能 |
| --- | --- | --- |
| 课堂教学 | 辅助教师实时调整教学 | 即时反馈、弱点识别、课程效果分析 |
| 个性化学习 | 提供差异化学习资源与指导 | 学习者画像、个性化推荐、差异化反馈 |
| 教育管理 | 服务宏观教育政策与资源配置 | 教育质量报告、差距识别、政策效果追踪 |
| 教育科研 | 推动理论研究与工具研发 | 大数据挖掘、教学方法验证、跨学科融合 |

* 1. 质量保障与持续改进
     1. 质量保障

质量保障是确保体系长期有效运行的前提。它不仅依赖于技术性能，还包括组织管理和规范执行。为此，测评体系应在以下方面提出明确要求：

1. 数据质量控制：确保采集数据真实、完整和准确，避免数据失真影响分析结果；
2. 模型性能评估：定期对模型的准确率、稳定性和可解释性进行检验，确保其教育应用价值；
3. 系统运行监控：建立运行日志和监控机制，及时发现并处理故障，保障系统稳定；
4. 用户反馈机制：通过收集教师和学习者的使用反馈，检验体系在实际教育中的效果与适用性。
   * 1. 持续改进

持续改进是体系不断适应教育发展和技术创新的必然要求。它强调通过循环迭代，不断优化测评流程和功能。具体应包括以下方面：

1. 需求调研：定期开展教育一线调研，了解教师与学习者的新需求；
2. 方法优化：根据教育学与人工智能领域的最新成果，改进数据处理和测评方法；
3. 技术升级：随着硬件和算法的发展，逐步更新系统架构和工具模块；
4. 标准修订：结合实际运行经验，适时修订规范，提升适用性与先进性。
   * 1. 闭环管理

质量保障与持续改进的最终目标，是形成完整的闭环管理模式。通过“计划—实施—检查—改进（PDCA）”循环，可以确保体系始终处于动态优化状态：

1. 在“计划”环节，设定体系目标与改进方向；
2. 在“实施”环节，落实检测、建模与应用工作；
3. 在“检查”环节，对运行效果进行评估；
4. 在“改进”环节，将评估结果转化为优化措施，并固化为新的制度。

