ICS 03.180
CCS A 18

T/CS

团 体 标 准

T/CS 683—2026

多模态人工智能教育测评动态优化与质量管控规范

Specification for dynamic optimization and quality control of multimodal artificial intelligence educational assessment

2026 - 09 - 05 发布　　2026 - 09 - 05 实施

中国商品学会　　发 布

目　　次

前言 II
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 总体要求 2
5 系统架构与功能要求 3
6 数据采集与处理要求 3
7 模型开发与验证要求 4
8 动态优化要求 6
9 质量管控要求 6
10 验证方法 7

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。

本文件由浙江机电职业技术大学提出。

本文件由中国商品学会归口。

本文件起草单位：浙江机电职业技术大学、国科时代数字资源(杭州)有限公司、成都数之联科技股份有限公司、黑龙江交通职业技术学院、江西省教育技术与装备发展中心、广西制造工程职业技术学院、双旗国际科技(北京)有限公司、安徽应用技术职业大学、广州现代信息工程职业技术学院、欣荣惠信息技术有限公司、合肥焱基域科技有限公司、石家庄学院、苏州凯因赛斯智能科技有限公司、黑龙江霆拓科技有限公司。

本文件主要起草人：罗辰、刘延明、伏彦林、赵荔、高萍、李嘉、朱玥龙、李敏、冯嘉倩、张继光、陈超、褚云霞、李世伟、钟海东、郭世华。

多模态人工智能教育测评动态优化与质量管控规范

1 范围

本文件规定了多模态人工智能教育测评系统在动态优化与质量管控方面的总体要求、系统架构与功能要求、数据采集与处理要求、模型开发与验证要求、动态优化要求、质量管控要求以及验证方法。

本文件适用于多模态人工智能教育测评系统的设计、开发、部署、运行、优化与质量管控活动。本文件也适用于教育测评服务提供者、教育机构、第三方评估机构对多模态人工智能教育测评系统进行评价和审核。

2 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 22239 信息安全技术 网络安全等级保护基本要求

GB/T 25000.10 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第 10 部分：系统与软件质量模型

GB/T 25000.51 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第 51 部分：就绪可用软件产品（RUSP）的质量要求和测试细则

GB/T 35273 信息安全技术 个人信息安全规范

GB/T 36342 智慧校园总体框架

GB/T 36351.1 信息技术 学习、教育和培训 教育管理数据元素 第 1 部分：设计与管理规范

GB/T 40552 教育服务效果测评指南

GB/T 42888 信息安全技术 机器学习算法安全评估规范

GB/T 45087 人工智能 服务器系统性能测试方法

GB/T 45225 人工智能 深度学习算法评估

GB/T 45288.2 人工智能 大模型 第 2 部分：评测指标与方法

3 术语和定义

下列术语和定义适用于本文件。

3.1

多模态人工智能教育测评 multimodal artificial intelligence educational assessment

整合文本、语音、图像、视频、行为数据等多种模态数据，运用人工智能算法对学习者知识、能力、表现及学习过程进行测量与评价的活动

3.2

动态优化 dynamic optimization

在多模态人工智能教育测评系统的运行过程中，依据实时采集的测评数据、反馈信息和性能监测结果，对测评模型、测评策略和系统参数进行持续调整与改进的过程。

3.3

质量管控 quality control

为确保多模态人工智能教育测评系统及其输出结果的准确性、可靠性、公平性、安全性和可追溯性而实施的一系列管理活动和技术措施。

3.4

测评模型 assessment model

用于处理多模态输入数据并输出测评结果的人工智能算法模型，包括但不限于深度学习模型、大语言模型和多模态融合模型。

3.5

多模态数据融合 multimodal data fusion

将来自两种及以上模态（包括文本、语音、图像、视频、行为数据等）的数据进行整合与联合分析的过程。

3.6

测评置信度 assessment confidence

测评模型对其输出结果可靠程度的量化度量值，取值范围为[0，1]，值越接近 1 表示模型对该次输出的把握度越高。

4 总体要求

4.1 基本原则

4.1.1 多模态人工智能教育测评系统的动态优化与质量管控应遵循科学性、准确性、公平性、安全性、可追溯性和持续改进原则。

4.1.2 多模态人工智能教育测评系统应符合 GB/T 36342 中关于智慧校园总体框架的规定。

4.1.3 多模态人工智能教育测评系统的软件质量应符合 GB/T 25000.10 中规定的质量模型要求。

4.1.4 多模态人工智能教育测评系统的就绪可用软件产品质量应符合GB/T 25000.51 中规定的质量要求和测试细则。

4.1.5 多模态人工智能教育测评系统的网络安全保护应达到 GB/T 22239 规定的第二级及以上安全保护等级。

4.2 系统生命周期要求

4.2.1 多模态人工智能教育测评系统的动态优化与质量管控应覆盖系统规划、设计、开发、部署、运行、维护和退役的全生命周期。

4.2.2 在每个生命周期阶段应建立相应的质量管控措施和动态优化机制。

4.2.3 系统生命周期的每个阶段应形成可追溯的文档记录。

4.3 文档要求

4.3.1 多模态人工智能教育测评系统的运营方应建立并维护以下文档：

a) 系统架构设计文档；
b) 数据采集与处理规范文档；
c) 测评模型开发与验证文档；

d) 动态优化策略与实施记录文档；
e) 质量管控计划与执行记录文档；
f) 系统运行日志与异常处理记录文档。

4.3.2 上述文档的保存期限不应少于该学习者所处教育阶段法定或标准学制年限与 5 年之和；对于无固定学制的非学历教育，保存期限不应少于测评数据产生之日起 15 年。

5 系统架构与功能要求

5.1 系统架构

5.1.1 多模态人工智能教育测评系统应采用分层架构，至少应包括数据采集层、数据处理层、模型服务层、测评应用层和质量管控层。

5.1.2 各层之间的接口应采用标准化协议，接口文档应完整、准确、可追溯。

5.1.3 系统架构应支持水平扩展，在并发用户数达到设计容量的 120%时，系统响应时间不应超过设计指标的 1.5 倍。

5.2 功能要求

5.2.1 多模态人工智能教育测评系统应具备以下核心功能：

a) 多模态数据采集功能，支持文本、语音、图像、视频和行为数据的实时采集；
b) 多模态数据预处理与融合功能；
c) 测评模型推理与结果生成功能；
d) 测评结果解释与可视化展示功能；
e) 测评置信度计算与输出功能；
f) 模型性能实时监测功能；
g) 动态优化触发与执行功能；
h) 质量管控指标计算与报告生成功能。

5.2.2 上述核心推理功能（第 c 项、第 e 项和第 f 项）的平均响应时间不应超过 3 s，95 百分位响应时间不应超过 5 s；其他辅助功能的平均响应时间不应超过 5 s。

5.2.3 系统应支持至少 1000 个并发测评会话。该指标依据 GB/T 45087 中关于人工智能系统性能测试的方法确定。

6 数据采集与处理要求

6.1 数据采集要求

6.1.1 多模态数据的采集应符合 GB/T 35273 中关于个人信息安全的规定。

6.1.2 采集学习者生物特征数据（包括面部图像、声纹、指纹等）前，应获得学习者的明示同意。

6.1.3 数据采集设备的技术指标应符合以下要求：

a) 图像采集设备的分辨率不应低于 1920×1080 像素，帧率不应低于 30 帧/秒；
b) 音频采集设备的采样率不应低于 44.1 kHz，量化位数不应低于 16 位；
c) 文本输入设备的响应延迟不应超过 100 ms。

6.1.4 数据采集过程应记录采集时间、采集设备标识、采集环境参数等元数据。

6.2 数据预处理要求

6.2.1 多模态数据预处理应包括数据清洗、数据标准化、数据增强和异常值处理等环节。

6.2.2 数据标准化应采用 Z-score 标准化方法或 Min-Max 归一化方法，具体方法的选择应在系统设计文档中明确记录。

6.2.3 数据预处理的每个环节应记录处理参数和处理结果，确保处理过程可追溯。

6.3 多模态数据融合要求

6.3.1 多模态数据融合应采用特征级融合、决策级融合或混合融合策略。

6.3.2 融合策略的选择应在系统设计文档中明确说明，并附选择依据。

6.3.3 融合后的数据质量应采用平均互信息（AMI）下降率或下游关键任务性能衰减率进行量化评估。采用 AMI 下降率时，以各模态原始数据间的平均互信息值为基准，融合后数据与原始数据间 AMI 的相对下降率不应超过 5%，按式（1）计算：

$$\eta = \frac{\mathrm{AMI}_{\text{原始}} - \mathrm{AMI}_{\text{融合}}}{\mathrm{AMI}_{\text{原始}}} \times 100\% \leq 5\% \quad \cdots\cdots (1)$$

式中：

η ——AMI 相对下降率；

$\mathrm{AMI}_{\text{原始}}$——各模态原始数据间的平均互信息值；

$\mathrm{AMI}_{\text{融合}}$——融合后数据与原始数据间的平均互信息值。

6.3.4 采用下游任务性能衰减率（以 F1 分数为基准）时，以未融合各模态单独输入时的最优 F1 分数为基准，融合后任务的 F1 分数相对衰减率不应超过 2%，按式（2）计算：

$$\delta = \frac{\mathrm{F1}_{\text{最优}} - \mathrm{F1}_{\text{融合}}}{\mathrm{F1}_{\text{最优}}} \times 100\% \leq 2\% \quad \cdots\cdots (2)$$

式中：

δ ——下游任务 F1 分数相对衰减率；

$\mathrm{F1}_{\text{最优}}$——未融合各模态单独输入时的最优 F1 分数；

$\mathrm{F1}_{\text{融合}}$——融合后任务的 F1 分数。

6.3.5 所采用的评估方法应在系统设计文档中明确固定，并在验证方法中予以复现。

6.4 数据存储与安全要求

6.4.1 多模态教育测评数据的存储应符合 GB/T 36351.1 中关于教育管理数据元素的规定。

6.4.2 敏感数据（包括面部图像、声纹、行为轨迹等）应采用加密存储，加密算法应符合国家密码管理局批准的算法标准。

6.4.3 数据访问应实施基于角色的访问控制（RBAC），不同角色的数据访问权限应在系统设计文档中明确界定。

6.4.4 数据存储的备份频率不应低于每日一次，全量备份的保存期限不应少于 30 d。

7 模型开发与验证要求

7.1 模型开发要求

7.1.1 多模态人工智能教育测评模型的开发应符合 GB/T 45225 中关于深度学习算法评估的规定。

7.1.2 模型开发应采用版本控制管理，每个版本的模型应记录以下信息：

a) 模型版本号；

b) 开发时间；

c） 训练数据集信息（数据来源、数据量、数据分布）；
d） 模型架构与超参数配置；
e） 训练过程中的损失函数与优化器配置；
f） 模型验证结果。

7.1.3 模型训练数据集的规模应满足以下要求：

a） 每个测评任务的训练样本数不应少于 10000 条；
b） 多模态数据的各模态样本数应保持均衡，任意模态样本数与所有模态样本数算术平均值之间的相对偏差不应超过 20%，按式（3）计算：

$$\rho = \frac{N_{\mathrm{i}} - N_x}{N_x} \times 100\% \leq 20\% \quad \cdots\cdots (3)$$

式中：

ρ ——样本数相对偏差；

N_{i}——第 i 个模态的样本数；

N_x——各模态样本数的算术平均值。

c） 若某模态数据因采集条件、样本稀缺或教育场景特殊性等客观原因无法达到均衡要求，应在模型开发文档中明确说明原因，并采用数据增强、重加权或合成数据等补偿措施，且需通过消融实验验证该模态对最终测评指标的贡献度。

7.2 模型验证要求

7.2.1 模型验证应符合 GB/T 45288.2 中规定的评测指标与方法。

7.2.2 模型验证应采用交叉验证方法，折数不应少于 5 折。

7.2.3 模型验证的评估指标应包括但不限于：

a） 准确率，不应低于 0.85；
b） 精确率，不应低于 0.80；
c） 召回率，不应低于 0.80；
d） F1 分数，不应低于 0.80；
e） AUC 值，不应低于 0.85。

7.2.4 上述指标适用于标准化程度较高的知识能力测评任务。对于开放式、创造性、主观性等高复杂度测评任务，经质量管控层评估备案后，可适当下调阈值，但准确率不应低于 0.75，F1 分数不应低于 0.70，且须在系统设计文档中明确定义任务复杂度等级及对应的阈值调整依据。

7.2.5 模型验证应在至少三个独立的数据集上进行，包括训练集、验证集和测试集，三者不应有重叠。

7.2.6 模型验证结果应形成正式报告，报告应包括验证方法、验证数据、验证指标和验证结论。

7.3 模型安全评估要求

7.3.1 模型的算法安全评估应符合 GB/T 42888 中关于机器学习算法安全评估的规定。

7.3.2 模型安全评估应包括以下方面：

a） 模型鲁棒性评估，应评估模型应对对抗样本攻击的防御能力；
b） 模型公平性评估，应检测并消除对特定群体的偏见；
c） 模型可解释性评估，应提供测评结果的可解释依据；
d） 模型隐私保护评估，应确保模型不泄露训练数据中的个人信息。

7.3.3 模型安全评估应每年至少进行一次，评估结果应形成正式报告。

8 动态优化要求

8.1 优化触发条件

多模态人工智能教育测评系统应在以下任一条件满足时触发动态优化：

a) 测评模型的准确率连续 30 d 下降超过 3 个百分点；
b) 测评模型的 F1 分数连续 30 d 下降超过 3 个百分点；
c) 测评置信度的平均值连续 30 d 低于 0.75；
d) 系统收到用户投诉或反馈的数量在 7 d 内超过测评总次数的 1%；
e) 训练数据累计新增量超过上一次优化完成时训练数据量的 10%；
f) 教育测评的内容或标准发生变更。

8.2 优化策略

8.2.1 动态优化应采取增量学习策略或全量重新训练策略。

8.2.2 增量学习策略适用于数据分布变化较小的场景；全量重新训练策略适用于数据分布发生显著变化的场景。数据分布变化判定方法如下：采用两样本 Kolmogorov-Smirnov 检验（KS 检验）或最大均值差异（MMD）检验对特征空间分布进行检测，显著性水平 α 取 0.05；当检验 p 值大于 0.05 时，判定为数据分布变化较小，采用增量学习策略；当检验 p 值小于或等于 0.05 时，判定为数据分布发生显著变化，采用全量重新训练策略。具体采用的检验方法应在动态优化策略文档中固定并明确说明。

8.2.3 优化过程中应保留至少三个历史版本的模型，以便在优化效果不佳时进行回滚。

8.2.4 优化后的模型应满足第 7 章规定的模型验证要求后方可上线部署。

8.3 优化频率

8.3.1 多模态人工智能教育测评模型的动态优化频率不应低于每季度一次。

8.3.2 在触发优化条件满足时，应在触发后的 72 h 内启动优化流程。

8.3.3 优化流程从启动到完成不应超过 14 d。

8.4 优化记录

每次动态优化应形成完整的优化记录，包括：

a) 优化触发原因；
b) 优化策略选择及依据；
c) 优化过程中使用的数据信息；
d) 优化前后的模型性能对比数据；
e) 优化后模型的验证结果；
f) 优化实施人员与审核人员信息；
g) 优化完成时间与上线时间。

9 质量管控要求

9.1 质量指标

9.1.1 多模态人工智能教育测评系统应建立以下质量指标体系：

a) 测评准确性指标：测评结果与人工评判结果的一致性系数（Cohen’s Kappa）不应低于 0.80；

b) 测评一致性指标：同一学习者重复测评结果的组内相关系数（ICC，采用绝对一致性定义）或皮尔逊相关系数不应低于 0.85；若测评数据不满足正态分布，应采用斯皮尔曼相关系数，且不应低于 0.85。具体采用的方法应在质量管控计划中明确固定；
c) 测评时效性指标：从最后一条多模态数据完成上传到测评报告在应用层呈现完毕的平均时间不应超过 5 s；
d) 系统可用性指标：系统年可用率不应低于 99.5%；
e) 数据完整性指标：采集数据的完整率不应低于 99%；
f) 用户满意度指标：用户满意度调查的满意率不应低于 85%。

9.1.2 上述质量指标应每月统计一次，并形成质量报告。

9.2 质量监控

9.2.1 多模态人工智能教育测评系统应建立实时质量监控机制，监控频率不应低于每 10 min 一次。

9.2.2 质量监控的内容应包括：

a) 系统运行状态（CPU 使用率、内存使用率、网络延迟等）；
b) 模型推理性能（响应时间、吞吐量等）；
c) 测评结果分布（得分分布、置信度分布等）；
d) 异常事件（系统异常、数据异常、模型异常等）。

9.2.3 质量监控数据应保存不少于 180 d。

9.3 质量审核

9.3.1 多模态人工智能教育测评系统的质量审核应符合 GB/T 40552 中关于测评审核的规定。

9.3.2 内部质量审核应每季度进行一次，外部质量审核应每年进行一次。

9.3.3 质量审核的内容应包括：

a) 系统架构与功能是否符合本文件要求；
b) 数据采集与处理是否符合本文件要求；
c) 模型开发与验证是否符合本文件要求；
d) 动态优化活动是否符合本文件要求；
e) 质量指标是否达到本文件规定的要求。

9.3.4 质量审核应形成正式审核报告，报告应包括审核范围、审核方法、审核发现、不符合项和改进建议。

9.4 不符合项处理

9.4.1 对质量审核中发现的不符合项，应按照以下时限完成整改：

a) 严重不符合项（影响系统安全或测评结果正确性），应在 24 h 内完成整改；
b) 一般不符合项（不影响系统安全但影响系统性能），应在 7 d 内完成整改；
c) 观察项（建议改进项），应在 30 d 内完成整改或制定整改计划。

9.4.2 不符合项的整改过程应形成记录，包括整改措施、整改责任人、整改完成时间和整改验证结果。

10 验证方法

10.1 系统架构与功能验证

10.1.1 通过审查系统架构设计文档和接口文档，验证系统是否采用分层架构以及各层接口是否采用标准化协议。

10.1.2 通过功能测试，验证系统是否具备第 5.2.1 条所列的各项核心功能。
10.1.3 通过性能测试（依据 GB/T 45087 规定的方法），验证系统在并发用户数达到设计容量 120%时的响应时间是否不超过设计指标的 1.5 倍。
10.1.4 通过并发测试，验证系统是否支持至少 1000 个并发测评会话。
10.1.5 通过响应时间测试，验证核心推理功能的平均响应时间是否不超过 3 s，95 百分位响应时间是否不超过 5 s；验证其他辅助功能的平均响应时间是否不超过 5 s。

10.2 数据采集与处理验证

10.2.1 通过审查数据采集流程文档和用户授权记录，验证数据采集是否符合 GB/T 35273 的规定。
10.2.2 通过设备检测，验证图像采集设备的分辨率是否不低于 1920×1080 像素、帧率是否不低于 30 帧/秒。
10.2.3 通过设备检测，验证音频采集设备的采样率是否不低于 44.1 kHz、量化位数是否不低于 16 位。
10.2.4 通过延迟测试，验证文本输入设备的响应延迟是否不超过 100 ms。
10.2.5 通过审查数据预处理日志，验证数据预处理是否包含数据清洗、数据标准化、数据增强和异常值处理等环节，以及各环节的处理参数和处理结果是否记录完整。
10.2.6 通过审查融合策略文档和数据融合质量评估报告，验证多模态数据融合是否采用特征级融合、决策级融合或混合融合策略，并验证所采用的融合质量评估方法（平均互信息下降率或下游任务性能衰减率）是否在系统设计文档中明确固定。当采用平均互信息下降率时，下降率是否不超过 5%；当采用下游任务性能衰减率时，衰减率是否不超过 2%。
10.2.7 通过审查数据存储方案和访问控制列表，验证敏感数据是否采用加密存储，以及数据访问是否实施基于角色的访问控制。
10.2.8 通过检查备份日志，验证数据备份频率是否不低于每日一次，全量备份保存期限是否不少于 30 d。

10.3 模型开发与验证验证

10.3.1 通过审查模型开发文档和版本控制记录，验证模型开发是否符合 GB/T 45225 的规定，以及每个版本的模型是否记录了版本号、开发时间、训练数据集信息、模型架构与超参数配置、损失函数与优化器配置、模型验证结果等信息。
10.3.2 通过审查训练数据集统计信息，验证每个测评任务的训练样本数是否不少于 10000 条；验证各模态样本数差异是否不超过 20%，若因客观原因超出差异范围，是否在开发文档中说明了原因并采取了数据增强、重加权或合成数据等补偿措施，且通过消融实验验证了该模态的贡献度。
10.3.3 通过审查模型验证报告，验证模型验证是否采用不低于 5 折的交叉验证方法。
10.3.4 通过审查模型验证报告中的评估指标数据，验证准确率是否不低于 0.85、精确率是否不低于 0.80、召回率是否不低于 0.80、F1 分数是否不低于 0.80、AUC 值是否不低于 0.85；对于经质量管控层备案的高复杂度测评任务，验证其是否采用了经备案的调整阈值且不低于最低允许值。
10.3.5 通过审查验证数据集划分记录，验证模型验证是否在至少三个独立的数据集（训练集、验证集、测试集）上进行，且三者无重叠。
10.3.6 通过审查模型安全评估报告，验证模型安全评估是否符合 GB/T 42888 的规定，以及是否涵盖模型鲁棒性评估、模型公平性评估、模型可解释性评估和模型隐私保护评估。
10.3.7 通过审查安全评估记录的时间戳，验证模型安全评估是否每年至少进行一次。

10.4 动态优化验证

10.4.1 通过审查系统监控日志和优化触发记录，验证系统是否在规定的触发条件满足时启动了动态优化流程。
10.4.2 通过审查优化策略决策记录，验证优化策略的选择是否符合数据分布变化检测结果（采用 KS 检验或 MMD 检验，显著性水平 α 取 0.05，p 值大于 0.05 时采用增量学习，p 值小于或等于 0.05 时采用全量重新训练），且检验方法是否在动态优化策略文档中固定并一致。
10.4.3 通过审查模型版本库，验证优化过程中是否保留了至少三个历史版本的模型。
10.4.4 通过审查优化后模型的验证报告，验证优化后的模型是否满足第 7 章规定的模型验证要求后方可上线部署。
10.4.5 通过审查优化记录的时间戳，验证动态优化频率是否不低于每季度一次。
10.4.6 通过审查系统排程日志和触发记录，验证当季度内无触发条件满足时，系统是否在季度末前按计划启动了周期性例行优化流程；同时验证触发条件满足后是否在 72 h 内启动了优化流程。
10.4.7 通过审查优化启动记录与优化完成记录的时间差，验证优化流程是否在 14 d 内完成。
10.4.8 通过审查优化记录文档的完整性，验证每次动态优化是否记录了优化触发原因、优化策略选择及依据、优化过程中使用的数据信息、优化前后的模型性能对比数据、优化后模型的验证结果、优化实施人员与审核人员信息、优化完成时间与上线时间等信息。

10.5 质量管控验证

10.5.1 通过审查质量报告，验证测评结果与人工评判结果的一致性系数（Cohen's Kappa）是否不低于 0.80。
10.5.2 通过审查重复测评数据和质量报告，验证同一学习者重复测评结果的组内相关系数（ICC，绝对一致性定义）或皮尔逊相关系数是否不低于 0.85；若数据不满足正态分布，验证其斯皮尔曼相关系数是否不低于 0.85，且具体采用的方法是否在质量管控计划中明确固定。
10.5.3 通过审查系统性能监控数据，验证从最后一条多模态数据完成上传到测评报告在应用层呈现完毕的平均时间是否不超过 5 s。
10.5.4 通过审查系统运行记录，验证系统年可用率是否不低于 99.5%。
10.5.5 通过审查数据采集日志，验证采集数据的完整率是否不低于 99%。
10.5.6 通过审查用户满意度调查报告，验证用户满意率是否不低于 85%。
10.5.7 通过审查质量监控系统的配置和日志，验证质量监控频率是否不低于每 10 min 一次，以及质量监控数据是否保存不少于 180 d。
10.5.8 通过审查质量审核报告，验证内部质量审核是否每季度进行一次，外部质量审核是否每年进行一次。
10.5.9 通过审查不符合项处理记录，验证严重不符合项是否在 24 h 内完成整改，一般不符合项是否在 7 d 内完成整改，观察项是否在 30 d 内完成整改或制定整改计划。
10.5.10 通过审查不符合项处理记录的完整性，验证整改过程是否记录了整改措施、整改责任人、整改完成时间和整改验证结果。