科大讯飞
iFLYTEK



2021
智能教育发展蓝皮书

——智能技术赋能教育评价

讯飞教育技术研究院
认知智能国家重点实验室智能教育研究中心

版权声明

版权归科大讯飞股份有限公司所有。保留一切权利。

非经科大讯飞股份有限公司书面同意，任何单位和个人不得擅自摘抄、复制本蓝皮书内容的部分或全部，并不得以任何形式传播。

前 言

2020 年 9 月，中共中央、国务院印发的《深化新时代教育评价改革总体方案》（以下简称《总体方案》）充分体现了鲜明的时代特征，明确指出“坚持科学有效，改进结果评价，强化过程评价，探索增值评价，健全综合评价，充分利用信息技术，提高教育评价的科学性、专业性、客观性。”“创新评价工具，利用人工智能、大数据等现代信息技术，探索开展学生各年级学习情况全过程纵向评价、德智体美劳全要素横向评价。”教育部 2021 年工作要点进一步强调，全面推进《总体方案》各项任务落细落地，扎实推进新时代教育评价改革。因此，适应教育领域深化改革与发展要求，充分利用现代信息技术优势，开发评价工具和手段，创新智能技术赋能的教育评价新模式，是新时代教育评价改革与实践的重要课题。

智能技术赋能教育评价的目的是对传统教育评价进行突破和创新。在长期的教育评价实践中，结果评价、过程评价、增值评价、综合评价中均已有不同程度的技术应用，但每一种评价都客观存在着不同类型和层次的问题，很多问题往往难以用传统的手段和方式加以解决，迫切需要采用新的技术和方法进行突破。当前，智能技术在教育评价领域的应用总体上还处于起步发展阶段，相关的技术系统和应用模式还不成熟，有待进一步研发和探索。在走向智能时代背景下，围绕服务“五类主体”，聚焦“四个评价”，致力破解评价实践中的难题，基于人工智能、大数据等技术开发和应用智能化评价系统及工具，创新教育评价内容及方式，提升教育评价的数字化、智能化水平，实现科学、客观和高效的评价与反馈，促进教育事业的改革与发展。

基于以上背景和思路，我们结合近年来国内外开展技术支持的教育评价实践探索，对智能技术赋能的教育评价进行了系统研究，撰写了《2021 智能教育发展蓝皮书——智能技术赋能教育评价》研究报告。本报告是“智能教育发展蓝皮书”系列研究之 2021 年度报告。报告共分 7 章，包括理念与框架、智能技术赋能“四个评价”、典型实践案例、行业现状及发展趋势。以期为智能技术与教育评价的融合创新提供依据和参考。

本报告的研究和撰写是在科大讯飞智能教育专家委员会指导下，由讯飞教育技术研究院/认知智能国家重点实验室智能教育研究中心负责组织与实施。在报告研究撰写过程中得到了许多行业专家的指导和帮助，众多机构和一线教育工作者给予了技术支持或提供了大量案例、文献资料，在此一并表示衷心感谢！由于撰写时间仓促，掌握数据资料不够，加上作者水平所限，本报告一定存在许多不足之处，敬请批评指正！

《2021 智能教育发展蓝皮书》编写组

2021 年 10 月

讯飞教育技术研究院

目 录

第 1 章 理念与框架 1
1.1 提出与价值 1
1.2 总体框架 3
第 2 章 智能技术赋能结果评价 5
2.1 结果评价发展现状及面临挑战 5
2.2 智能技术如何赋能结果评价 5
第 3 章 智能技术赋能过程评价 13
3.1 过程评价发展现状及面临挑战 13
3.2 智能技术如何赋能过程评价 13
第 4 章 智能技术赋能增值评价 18
4.1 增值评价发展现状及面临挑战 18
4.2 智能技术如何赋能增值评价 18
第 5 章 智能技术赋能综合评价 23
5.1 综合评价发展现状及面临挑战 23
5.2 智能技术如何赋能综合评价 23
第 6 章 智能技术赋能教育评价典型实践案例 29
6.1 智能技术赋能教育评价实践概述 29
6.2 智能技术赋能基础教育质量监测 30
6.3 智能技术赋能学校发展评价 32
6.4 智能技术赋能在线课程评价 33
6.5 智能技术赋能教师课堂教学评价 36
6.6 智能技术赋能学生综合素质评价 37
第 7 章 行业现状及发展趋势 40
7.1 行业现状分析 40
7.2 技术应用趋势 45
7.3 行业发展趋势 47
工作组与致谢名单 49

讯飞教育技术研究院

第 1 章 理念与框架

智能技术赋能为教育评价改革发展提供了重要机遇。充分发挥智能技术优势，对传统教育评价进行革新与发展，探索形成新的教育评价模式，为解决评价实践中的难题提供可行路径。

1.1 提出与价值

1.1.1 智能技术赋能教育评价的提出

教育评价是一个动态发展的过程，建立客观、多元、科学、开放的教育评价体系是当前全球教育评价改革的趋势。国际成就评价协会倡导实行多维、综合评价[①]。世界各国积极探索实践，形成了 PISA、PIRLS、TIMSS、PIAAC 等国际教育评价项目，评价工具从简单的纸笔测验到纸笔、观察、计算机等多种手段相结合，评价功能从筛选走向更加关注诊断、促进发展和决策服务等。2020 年 9 月，我国颁布《总体方案》系统推进和全面深化教育评价改革。探索新型教育评价模式已经成为深化教育评价改革的必然趋势和要求。

近几年随着人工智能、大数据等智能技术的迅速发展与广泛应用，人类社会发展走向智能时代。2017 年我国国务院出台了《新一代人工智能发展规划》，从而广泛出现了“人工智能+”概念。2018 年，习近平总书记在向 2018 世界人工智能大会的贺信中指出“人工智能赋能新时代”。2019 年政府工作报告中提出拓展“智能+”，为制造业转型升级赋能。“智能技术赋能”就是运用智能技术增加正能量、增强发展能力，促进传统行业智能化升级与发展，包括教育领域的转型发展，催生了“人工智能+教育”的新型教育模式[②]。

智能技术赋能教育评价的正式提出，是教育评价领域顺应“智能+”时代要求，依靠智能技术增强发展能力，推动自身的改革与发展，以实现教育评价的数据化、智能化。正如董奇指出的，教育评价已进入新的发展阶段，智能化的教育评价新趋势已出现[③]。杨宗凯认为智能技术与教育的深度融合为教育教学评价改革创新提供了可行路径[④]。陈峰强调加快更新迭代教育评价手段，研究开发教育

① 中国教育科学研究院国际与比较教育研究所. 全球经验:教育评价改革进行时[M].重庆:西南师范大学出版社,2020:241.
② 任友群.总序[M]//刘邦奇,吴晓如.智慧课堂:新理念 新模式 新实践.北京:北京师范大学出版社,2019:2.
③ 董奇. 教育评价进入新的发展阶段,智能化评价趋势已出现[EB/OL]. [2021-03-11]. https://www.eol.cn/news/xueshu/201908/t20190807_1677069.shtml.
④ 杨宗凯. 利用信息技术促进教育教学评价改革创新[J]. 人民教育 2020(21):30-32.

测量和评价技术，充分利用“互联网+”、AI 等技术工具，探索“智评”模式，以此推行精准评价[①]。可见，走向智能时代使教育评价迎来新的发展机遇，充分发挥智能技术优势，探索技术赋能的新型教育评价模式，为教育评价改革提供了发展方向和可行路径。

1.1.2 智能技术赋能教育评价的涵义

综合国内外学者对智能技术运用于教育评价的理解，可以看到智能技术赋能教育评价将是未来教育评价发展的必然趋势。教育评价是对教育过程价值判断的活动[②]，新时代智能技术能够帮助我们在这种教育价值判断活动中，做到更准确、更全面、更及时、更自动和更智能。根据以上分析，可以认为“智能技术赋能教育评价”是在走向智能时代背景下，充分利用人工智能、大数据等智能技术的优势，对传统教育评价进行革新与发展，改进教育评价过程与方法，创新教育评价模式，提升教育评价的数字化、智能化水平，实现科学、客观和高效的评价与反馈，促进教育事业的改革与发展。

1.1.3 智能技术赋能教育评价的主要价值

（1）评价模型科学化。借助智能技术手段，通过教育评价研究人员、信息技术专业人员、教育教学人员等多方的协作，针对不同的评价目标、对象和内容，构建科学的评价指标模型、权重模型、算法模型。

（2）主体参与多元化。在智能终端、教育大数据和信息化评价工具的支持下，增强了多元主体参与性，使得教师、学生个体、同伴、专家及利益相关者等多元化主体形成“评价共同体”，多视角反映评价对象的综合表现，全面客观呈现评价结果，以保证教育评价的有效性和公平性。

（3）数据获取伴随化。利用平台系统及各类智能硬件设备对评价数据进行全过程、全方位、多维度的采集，改变过去人工采集记录的方式，实现评价数据伴随式获取。

（4）诊断分析最优化。通过采用数据融合、数据分析等技术，对多模态数据进行诊断分析，实现了多维、全局数据处理和分析的最优化，从而达成精准评估和测评。

① 陈峰. 用评价改革引领促进健康教育生态建设[EB/OL]. [2021-03-11]. http://www.zjjyb.cn/html/2020-12/02/content_29771.htm.
② 陶西平. 教育评价辞典[M].北京:北京师范大学出版社,1998:55-56.

（5）评价反馈精准化。基于智能技术实现过程化的评价数据反馈，为教师提供实时、准确的反馈信息；通过高度个性化定制的交互式数据可视化图表，将评价结果直观、生动、精准地呈现给用户；通过智能推荐引擎等技术将评价结果精准推送给用户。

1.2 总体框架

推进智能技术赋能教育评价的建设与应用，需要从整体上把握智能技术背景下技术、业务、应用和用户间的关系以及对评价产生的影响。借助教育信息化顶层设计方法论[①]，在智能教育关键技术平台参考框架的基础上[②]，结合智能技术赋能教育评价活动的实现过程，可以提出和构建智能技术赋能教育评价生态体系总体框架，如图 1-1 所示。

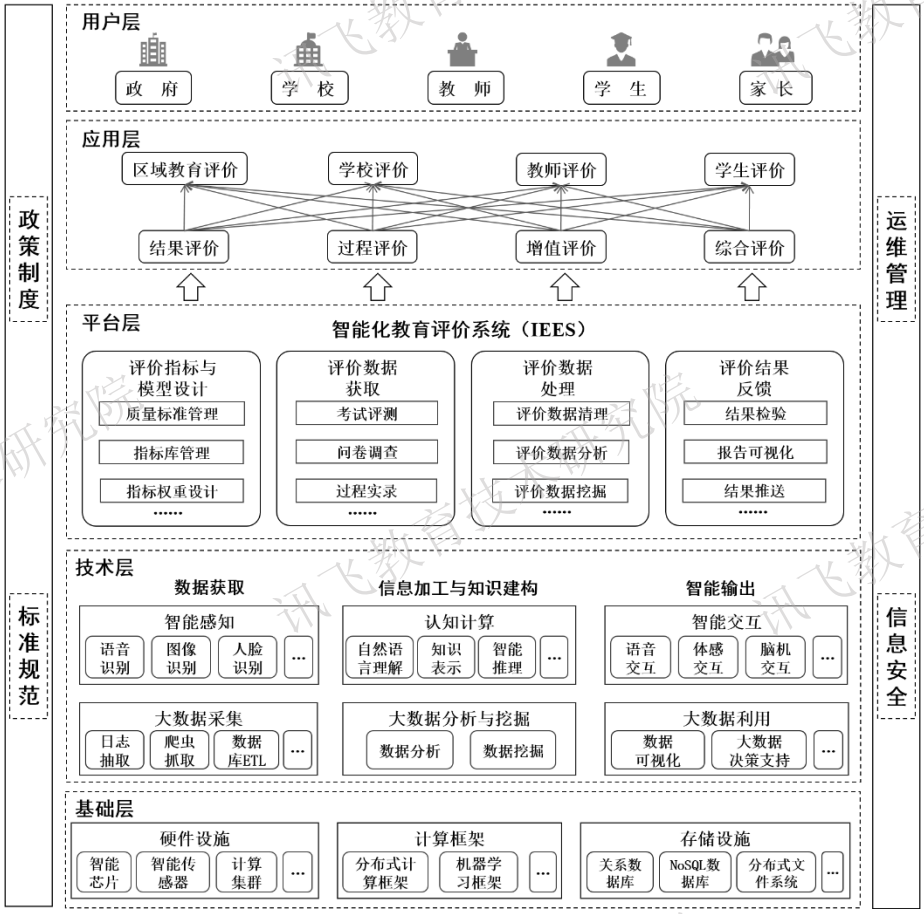


图 1-1 智能技术赋能教育评价生态体系总体框架

智能技术赋能教育评价生态体系由“5+1”的总体框架构成，形成了开放、多元、科学、高效的智能化教育评价生态体系，具体包括基础层、技术层、平台

① 蒋东兴,袁芳,刘启新,沈力强,付小龙. 信息化顶层设计[M]. 北京：清华大学出版社，2015:17-34.
② 刘邦奇,王亚飞. 智能教育:体系框架、核心技术平台构建与实施策略[J].中国电化教育,2019,(10):23-31.

层、应用层、用户层 5 个层次结构和 1 个保障体系。

（1）基础层。技术赋能教育评价的基础环境，由硬件设施、计算框架、存储设施等构成。硬件设施为整个教育评价过程提供硬件支撑环境；计算框架为系统提供统一的计算服务、统一的机器特征提取和算法模型建模服务；存储设施为整个评价过程提供统一的数据存储、读写和管理等数据服务，保障数据的安全等。

（2）技术层。通过利用各种智能技术实现数据的智能化采集、加工分析和可视化输出等。主要包括数据获取中的语音识别、图像识别、日志抽取、数据库 ETL 等技术；信息加工与知识构建中的自然语言处理、知识表示、数据分析、数据挖掘等技术；智能输出中的语音交互、体感交互、数据可视化等技术。

（3）平台层。为智能技术与教育评价深度融合提供关键服务，基于人工智能、大数据等技术构建智能化教育评价系统，为应用提供智能化数据采集、分析、诊断、反馈和改进指导能力[①]。主要包括评价指标和模型设计、评价数据获取、评价数据处理、评价结果反馈等服务。

（4）应用层。各类教育评价场景技术应用的价值体现，为区域、学校、教师、学生等提供各类评价。如利用智能技术优化智能阅卷与评测、智能考试分析等，对教学结果进行评定；通过伴随式数据采集、多模态数据诊断分析发现过程中的问题和即时反馈调控；通过综合评价模型构建与优化、数据深度挖掘分析等实现教育质量综合评价、学生综合素质评价等。

（5）用户层。用智能技术赋能教育评价，以实现评价的目的，是教育评价的归宿所在。具体来说，政府、学校、教师、学生、家长等用户通过智能终端设备实现一站式登录智能化教育评价系统，为不同用户提供不同的评价反馈，其评价反馈结果对用户具有导向、诊断、鉴定、调控、改进等功能。

（6）保障体系。智能技术赋能教育评价有序实施的重要保证，使整个教育评价生态体系有效运转，包括政策制度、标准规范、运维管理、信息安全等保障。

① 新华社. 我国将建“智能化基础教育评价系统”助力科学教育评价[EB/OL]. [2021-03-11]. http://www.gov.cn/xinwen/2020-12/13/content_5569206.htm.

第 2 章 智能技术赋能结果评价

《总体方案》明确提出要“改进结果评价”，强调肯定结果评价的价值和作用，同时应克服结果评价在具体实践中过于强调结果、唯分数论等弊端，在评价工具的开发和量化结果的运用上深度挖潜。智能技术赋能结果评价，主要应用包括智能命题与组卷、智能考场管理、智能阅卷、智能考试分析、智能语言测评、智能艺术素质测评等场景。

2.1 结果评价发展现状及面临挑战

结果评价通常是在教育活动完成后，对最终结果做出的价值判断，即结果与目标间一致程度的评价，在一定程度上可以反映学生的学习情况、教师的教学质量。随着教育评价改革的深化发展，结果评价需要逐步建立明确的测评标准，提升评价的准确性和科学性，改变简单以分数呈现评价结果的方式，增加对评价结果的分析，提高结果评价指导发展的价值。结果评价强调用客观、具体的数据对复杂的教育现象进行定量描述，具有目标性、规范性、客观性、可操作性等特点。

当前我国最重要的结果评价就是中高考[①]，而以中高考为代表的正式考试其面临的主要问题与挑战体现在：如何提高命题与组卷的科学性，如何保障考试环境公平公正，如何实现试卷评阅的准确高效，如何深度分析利用考试数据，如何响应语言、艺术类考试加入后带来的评价挑战等。

2.2 智能技术如何赋能结果评价

智能技术的发展与应用为改进结果评价提供了有效的解决路径，主要表现在以下四个方面。

① 辛涛. 挖掘过程性数据中的教育价值[EB/OL]. [2021-08-15]. http://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/jyzt_2020n/2020_zt21/zhuanjiawenzhang/202012/t20201214_505246.html.

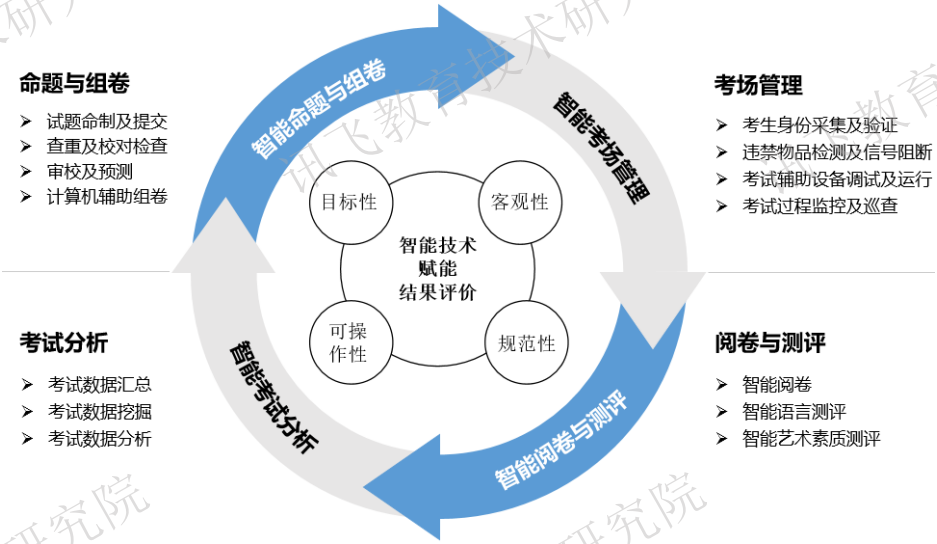


图 2-1　智能技术赋能结果评价

2.2.1　智能命题与组卷

智能命题与组卷是破解传统命题与组卷难题的有效解决路径，可实现命题与组卷工作的信息化、自动化、精准化、智能化，提高试卷编制的质量与效率，精准体现命题人的出题意图，间接保障考试结果的真实有效性。其作用主要体现在：信息化命题管理助力降本增效；自动化查重降低命题重复风险；精准化难度预测提前预知命题效果；智能化组卷为试卷质量提供稳定保障等。

在智能技术的支撑下，智能命题与组卷推动传统线下命题与组卷工作的信息化、自动化。例如命题人在计算机上借助试题呈现技术完成各类题型的命制，再导入 UKey 后提交。通过对试题进行自动查重，程序自动校对命题人填写的知识点、学科核心素养等属性是否达标，同时自动补全某些属性，完成对试题内容的完整性、合规性检查等。经过多轮修改审校后形成多个试题版本，线上评议选定最终版本，通过难度自动预测提前预测命题效果。应用智能组卷完成试卷编制，并对组卷质量进行分析，供命题人判断试卷是否达到考试要求。

案例 2-1　高利害考试中的智能命题与组卷*

教育部考试中心为新高考命题而建设的国家题库 2.0，依托该题库开展智能命题与组卷，在保障国家教育考试命题能力建设上发挥了积极的作用。国家题库 2.0 以基于高考评价体系的“一核四层四翼”为内核进行试题属性建构，通过线上模拟传统的线下入闱命题流程，为命题人员提供了全流程、全方位、浸入式的

* 本案例根据教育部考试中心国家题库 2.0 相关论文、官网等资料整理总结。

在线命题及组卷体验，同时根据不同学科的个性化需求，增加了多种个性化功能（如数学公式编辑器、英语单词检测）。此外，国家题库 2.0 探索利用试题查重、难度自动预测、智能组卷等智能技术解决各种复杂命题场景下的难题，实现传统考试命题全流程的信息化、智能化重塑。

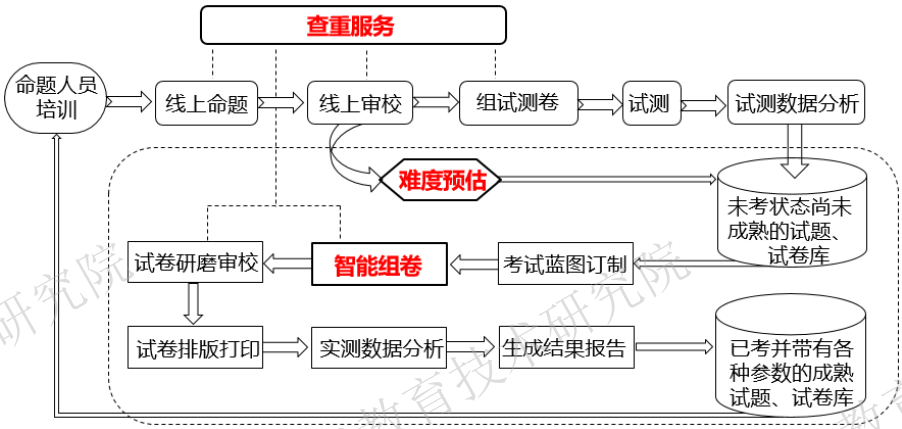


图 2-2 智能技术在智能命题与组卷标准化流程中的作用

2.2.2 智能考场管理

智能考场作为标准化考场的高级发展阶段，是考试管理模式在信息化时代下的演化方向。智能考场包含考生身份多重验证、考试辅助设备、智能视频监控等先进技术设备，为考试过程提供智能化、信息化的管理和服务，有效保障了考试环境的公平公正。其作用主要体现在：考生身份验证效率和准确性显著提升；智能硬件可支撑多种考试形式；智能监考助力防范作弊等。

在智能技术的支撑下，传统考场实现智能化、信息化全面升级，并在多次实践应用中形成了智能考场管理的标准规范。例如在考生身份采集及验证方面，采集考生多种生物特征并进行验证，确保考生身份信息准确唯一；违禁物品检测及信号阻断方面，使用金属探测仪对考生进行周身检查，开启信号屏蔽器阻断作弊设备与外界互通的信道；智能监控系统可自动发现考试过程中的疑似作弊行为，网上巡查系统可有效协助监考人员加强对考试过程的掌控和管理。

案例 2-2 智能考场实现考场异常行为智能分析*

北京市海淀区作为国家教育考试综合管理平台试点区，积极建设“区-考点-考场”三级共享的国家教育考试管理云平台及智能考场设施，全区已有 18 个考点、608 个考场实现了考情数据全入库、高清视频全覆盖、指挥指令实时达、考

* 本案例根据北京市海淀区教育局 2020 年高考试点智能监控相关公开报道、官网等资料整理总结。

务管理电子化。在 2020 年新高考实践中，海淀区智能考场的成功应用受到了广泛关注，多项智能技术有力保障了人脸识别及身份认证、违禁物品检测、考试过程巡查等环节的平稳高效实施。智能考场创新性的采用 AI 智能分析技术，对考场视频图像进行智能分析，让计算机“读懂”海量的视频，推送“异常行为”到监控屏幕，帮助巡考、监考人员工作，彻底改变“盯屏”的监考模式，最终实现了全周期监管、全行程跟踪、全方位监控、全过程留痕的预期效果。



图 2-3 海淀区智能考场网上巡考实况

注：本图片来自北京市海淀区人民政府官网 2020 年 7 月 10 日公开报道。

2.2.3 智能阅卷

智能阅卷为实现更加高效、公平、客观的评阅试卷提供了有效路径，通过图像检测、手写识别、自然语言理解等智能技术的深度应用，可实现主观题智能评分、无效作答检测、雷同卷判定、自动统分等智能化操作，使得“具备专家评分水平”的计算机标准化地执行和实施对主观题评分这一设想初步变成了现实[①]。其作用主要体现在：阅卷过程实现降本增效；智能甄别抄袭、套题等虚假结果；智能评分更加客观稳定；评分过程有迹可循等。

在智能技术的支撑下，智能阅卷已经逐步取代传统以人工方式进行的阅卷，并形成了标准化的阅卷规范流程。首先通过智能算法筛选出定标集数据，然后由专家对数据进行定标，计算机学习定标规则后完成评分模型的构建。阅卷人员使用高速扫描仪对纸质试卷进行扫描，计算机自动切割作答区域检测空白卷、完成客观题评分。接下来进行主观题数据预处理，在该环节通过手写识别技术完成数据转换，同时进行相似度检测，完成雷同卷、套题的甄别。预处理完成后通过基于自然语言理解的评分引擎进行主观题评分。最后结果汇总阶段，智能质检技术

① 汪张龙. 人工智能技术在考试中的应用[J].中国考试,2017(11):37.

的应用能有效确保评分结果公正公平。

案例 2-3 智能阅卷在高利害考试中的应用*

安徽省教育招生考试院在高考中探索应用智能阅卷取得了丰硕的成果。在 2017 至 2019 年连续三年的高考网上评卷期间，考试院将智能评分作为网上评卷辅助质量监控的重要方法，积极实施了相关应用，取得良好预期。从对高考英语、语文作文题的智能评分；再到对高考语文简答题、高考数学证明/计算题、高考文科综合能力测试政治简答题的智能评分，一步步探索出了智能阅卷在高考多科目、多题型上应用的可行性和人机结合评分的新模式。



图 2-4 智能阅卷主要流程示意图

2.2.4 智能语言测评

智能语言测评是利用智能技术赋能语言类考试评价的有效探索，借助专用耳麦、语音识别等智能技术设备，可实现智能语音采集、智能语音识别、智能语言评分等智能化操作，是一种采用人机对话、智能评分方式进行的新型语言考试①。其作用主要体现在：智能语言测评助力降本增效；完整清晰采集答题语音；语言能力评价客观全面；测评结果更有益于进步提升等。

在智能技术的支撑下，智能语言测评可按规范流程全方位考察考生的听说读写能力，避免陷入传统语言类考试只关注听力或口语某一方面能力的误区。对语言考试结果评价的改进体现在通过专用耳麦完整清晰地采集考生的答题语音，结合智能音质检测技术，保障采集考生语音的质量。通过智能评分引擎解析考生语

* 本案例根据安徽省教育招生考试院采用人工智能辅助高考阅卷相关论文、公开报道、官网等资料整理总结。

① 汪张龙,潘泽亮. 人工智能技术及其在语言产业中的应用[J].语言产业研究,2018(7):115.

音数据，从多个维度对考生答题语音进行智能评分。测评系统支持将多维评分结果导入报告模板，自动生成语言综合能力测评报告，指导改进语言学习中的问题。

案例 2-4 智能语言测评应用于语言类考试*

长沙市作为湖南省英语口语首个改革示范区，于 2020 年率先启动“人+机”英语口语与听力考试试点工作。2020 年 11 月 13 日，长沙雅礼洋湖实验中学的 400 多名学生首次采用“人机对话”进行英语口语测试，依托科大讯飞的智能口语评测技术，实现从“流畅度”“完整度”“连续性”“语音语调”“准确度”等多个维度直观全面地评价考生的口语和听力水平，受到前来考察的省、市教育考试院专家组的肯定。国家语言文字工作委员会于 2004 年在其重点科研项目“智能语音技术在普通话辅助学习中的应用”中提出将计算机辅助测试和语言学习系统结合起来，建立比较完善的普通话智能口语学习系统。截至 2018 年，普通话智能考试已覆盖国内全部省份，经过各省测试站的连续观测，全国考生整体普通话水平保持逐年提升态势。

2.2.5 智能艺术素质测评

智能艺术素质测评是利用智能技术赋能学生艺术素质考评的积极探索。依托专门的艺术考试系统、音乐唱奏题智能评分、操作题评分等智能技术，可实现大规模上机考试、计算机智能评分、自动汇总统计艺考成绩等智能化操作，是一种采用人机对话方式进行的学生艺术素质终结性评价方式。其作用主要体现在：智能艺术素质测评助力降本增效；评价标准统一可量化；评价过程公平公正；智能评分准确高效等。

在智能技术的支撑下，智能艺术素质测评可同时开展音乐和美术两门主要艺术课程的考试，响应了各地初高中艺术素质测评相关政策中将两科合并考试的需求。其对艺术素质测评结果评价的改进体现在通过艺考定制化题型和考试作答设备完成音乐作答数据和美术作答数据的采集，并同步完成数据校验工作，确保考试结果真实有效。应用智能评分技术，依托音乐测评模型与唱奏题评分技术从多维度对音乐作答数据进行评分，依托操作题评分技术按步骤对美术作答数据进行评分。计算机自动汇总成绩，将分数结果按转换规则计为等级，供教学效果评价和录取学校参考。

* 本案例根据长沙市教育考试院“人+机”英语口语与听力考试相关公开报道、官网等资料整理总结。

案例 2-5 音乐、美术考试中的智能测评*

东莞市教育局于 2019 年发布了《东莞市初中学业水平考试音乐、美术考试实施方案》（以下简称“方案”），方案要求音乐学科及美术基础知识、欣赏评述均通过引入相关考试系统，采用计算机辅助考试方式进行考核，同时应用机器智能评分方式进行评卷。东莞市于 2020 年 12 月 30 日首次开展初中学业水平考试音乐、美术上机考试，运用先进的测评工具和音乐、美术评分技术手段，高效完成对全市初中生艺术素养的测评工作，为今后全市应用智能技术大规模开展艺术测评积累了经验与信心。



图 2-5 东莞市 2021 年初中学业水平考试音乐、美术上机考试实况

2.2.6 智能考试分析

智能考试分析是全面分析考试数据，实现考试多元评价和精准教育决策的有效手段，依据教育测量理论为指导，自动汇总考试上下游数据、深度挖掘数据间信息及规律、科学展示可视化分析结果，让教育主管部门、教师、学生等更好地理解并运用考试分析结果，并实现全面、科学、精准的多元评价效果。其作用主要体现在：高效处理考试数据；深度挖掘影响考试结果的因素；分析报告有助于改进指导等。

在智能技术的支撑下，智能考试分析改变了以往考试分析过于关注考试分数和排名的倾向，引导教师、学生、家长透视每个分数背后的问题并发现提升改进办法。如通过深度挖掘分析考试数据中隐藏的信息及规律，对总分构成因素、个体特征与学业成绩关联进行分析。分析结果以试卷分析报告和成绩分析报告呈现，

* 本案例相关资料由广东讯飞启明科技发展有限公司提供。

向不同对象揭示命题、教学、学习中的问题并予以改进指导。

案例 2-6　智能考试分析发挥学业水平考试数据价值*

青海省考试管理中心自 2017 年 12 月始，每年发布《青海省学业水平考试成绩分析报告》（以下简称“报告”）。报告依托青海教育云将全省学业考试数据汇总后深度挖掘分析，建立多维度分析指标体系，有效帮助教育行政部门准确掌握教育教学现状，为科学决策提供支撑；为高中学校开展针对性的教研活动和教学指导提供依据，促进教育教学质量的提高；为考生提供全科成绩综合诊断，定位薄弱学科和知识点，助力学业发展规划①。



图 2-6　智能考试分析报告效果示意

*　本案例根据青海省学业水平考试成绩分析报告相关公开报道、官网等资料整理总结。

① 青海省人民政府. 我省发布首个普通高中学业水平考试成绩评价分析报告[EB/OL]. [2021-04-28]. http://www.qh.gov.cn/zwgk/system/2018/04/10/010299192.shtml.

第 3 章 智能技术赋能过程评价

《总体方案》明确提出要“强化过程评价”。过程评价是在教育实施过程中对评价对象进行评价，是与教育教学实践紧密结合的价值发现过程。智能技术赋能过程评价，主要应用包括全过程伴随式数据采集、多模态数据诊断分析和实时反馈调控等场景。

3.1 过程评价发展现状及面临挑战

过程评价是通过强调过程的价值，突出过程与目标并重的双重取向，注重对综合能力的培养和评价对象成长所应具备能力发展的考察，关注被评价对象习得过程的评估与测量，将评价诊断、即时反馈贯穿教学全过程，从而实现以评促学、以评促教目标的评价方式[①]，具有动态性、诊断性、即时性的特点。

针对传统评价过分注重静态、可量化和浅层次学习效果的弊端，过程评价进行了修正，将师生在教与学活动进程中的全部信息纳入评价范围。但在信息纳入评价的过程中依然存在不少问题阻碍着过程评价的顺利开展，比如过程数据采集较为单一分散且粗放、在评价开展过程中实施难度大、过程评价与教学过程融合的要求较高等。

3.2 智能技术如何赋能过程评价

智能技术的发展与应用为探索过程评价提供了有效的解决路径，主要表现在以下三个方面。

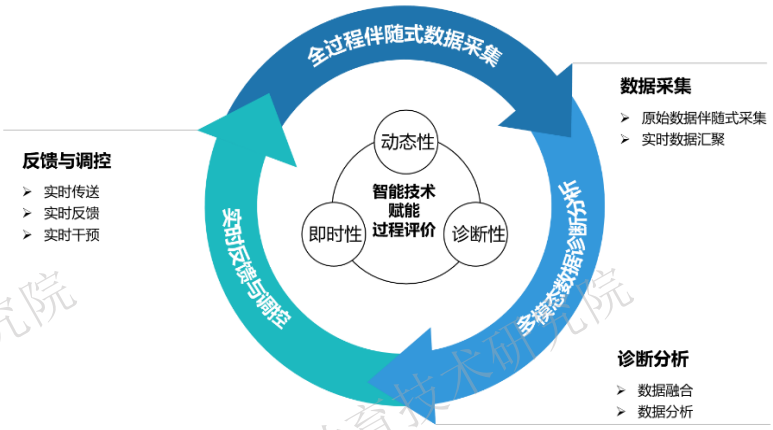


图 3-1 智能技术赋能过程评价

① 许华琼,胡中锋. 形成性评价及其反馈策略[J].教育测量与评价(理论版),2010(01):23-26.

3.2.1 全过程伴随式数据采集

全过程伴随式数据采集是过程评价的首要环节。全过程伴随式数据采集能够实时采集教师课堂教学、学生在线学习、学生成长记录等全过程数据，为学校的可持续发展、教师的教学、学生的学习等提供全过程数据服务。

在智能技术的赋能下，数据采集的整个过程大多是全自动、智能化的，使得采集的数据具有全过程、连续性等特征；而智能技术支撑下的数据采集也能够有效降低数据失真率、信息采集故障率等情况，提升过程数据的真实性、准确性；同时，通过智能技术采集的数据没有特定的目标性，也没有将数据进行特殊筛选，数据内容通常类型多样、结构异质。

案例 3-1 智慧体育教学模式中的全过程伴随式数据采集*

重庆两江新区星辰初级中学在七年级体育课教学中，利用全过程伴随式学生心率数据采集辅助教师精准教学，具体做法是：①学生佩戴对应编号的运动心率臂带，运动心率臂带采用动态医疗心率传感器融合技术实时采集学生的心率，精准监测动态心率；②通过低功耗蓝牙（Bluetooth Low Energy，BLE）与微型感知基站通信，实时采集学生运动心率数据并存储，进行实时心率监测和预警；③系统将班级和学生个人的运动强度、运动密度、有效运动时长、心率预警等各项数据可视化呈现给教师，辅助体育教师识别学生个体差异，有针对性地指导学生进行体育锻炼，如图 3-2 所示和图 3-3 所示。从图 3-2 来看，系统提示学生 10 运动安全预警，其靶心运动时间为 13 分 50 秒，平均心率、最高心率、运动密度和心率指数偏高，需要教师特别关注其身体素质，适当调整运动量，并提供针对性的运动指导。从图 3-3 来看，班级学生整节课心率曲线记录完整，整体变化平稳，准备阶段心率上升较快，最大心率出现在基本阶段，结束阶段逐渐下降，并始终维持在运动区间范围，可见，班级全体学生体育锻炼充分，内容安排合理。

* 本案例相关资料由重庆两江新区星辰初级中学提供。



图 3-2 学生 10 课堂分析情况

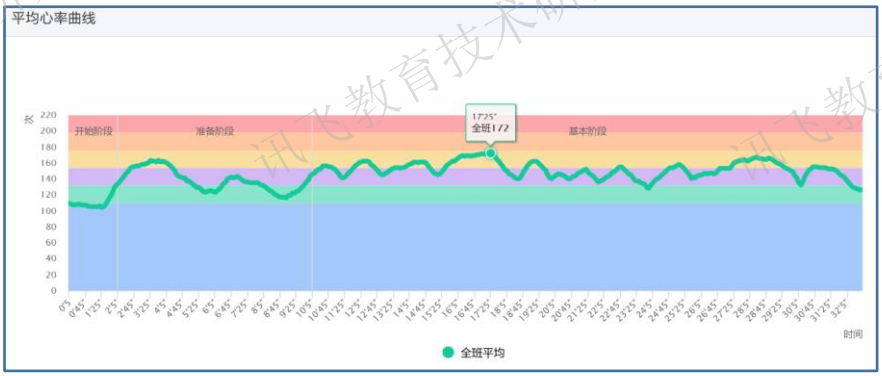


图 3-3 班级平均心率曲线

3.2.2 多模态数据诊断分析

多模态数据诊断分析通过借助技术手段对多模态数据进行有机整合，提取相关符号表征状态中潜藏的语义信息，以衡量诊断学生或教师的发展过程，进而做出判断或预测。

在智能技术的赋能下，多模态数据诊断分析通过数据挖掘算法、内容分析、预测性分析、系统建模等技术对采集的多模态数据进行融合分析，挖掘其存在的相关性与因果性，在为决策提供更多信息的同时，也实现了对学习者特征要素的准确表征[①]，提升了决策总体结果的准确率。

案例 3-2 多模态诊断分析助力教师改善学生计算机协作学习*

丹麦奥尔堡大学通过多模态诊断分析对学生的协作学习活动展开相关研究，以视频片段、摄影、共享网页上记录学习活动的时间轴、教师在博客上分享的项

* 本案例根据丹麦奥尔堡大学多模态诊断分析相关论文等资料整理总结。

① 黄涛,王一岩,张浩,杨华利. 智能教育场域中的学习者建模研究趋向[J].远程教育杂志,2020,38(01):50-60.

目经验等多渠道进行全过程伴随式采集，创建丰富的纵向数据集，对文本、语音、动作等进行数据分析。基于 ELAN 转录平台可将转录的数据微型化精确至帧，符号资源转变为单独的“层”，且标注层没有限制，标注编码表可自行定义，标注与标注间互相链接，标注结果可根据研究需求进行多种格式的输出以满足多模态数据分析需求。基于视频的多模态数据分析有助于教师了解儿童如何协作学习，给予教师更加微观的视角获得儿童协作学习的过程性数据，由此协助教师改善协作学习的教学设计。

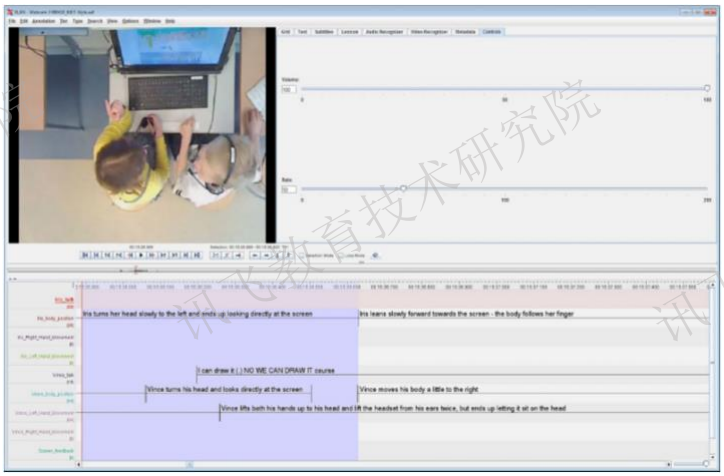


图 3-4 ELAN 转录层

3.2.3 实时反馈与调控

在教育评价领域中，实时反馈与调控是发挥教学评价“诊断”与“促进”的必要环节，是过程评价的重要组成部分。通过实时反馈与调控可以及时关注问题或错误，从而有效纠正不恰当的教学策略或学生的错误认知，帮助教师和学生最终达到预期的教学目标[①]。

在智能技术的赋能下，通过利用无线通信技术、数据可视化技术、大数据技术等对诊断分析的结果数据进行实时反馈与调控。如以网络传输协议为基础，借助互联网通信技术，实现各种设备之间的无线通信，能够打破时空界限，实时接收到评价反馈的信息。还可通过数据可视化技术，将数据进行图示化以便于信息的传播。在过程评价中利用知识节点路径规划、知识图谱等技术根据学习者的自身学习数据还可创建自定义的学习内容和学习路径，将最优化的学习方案推荐给学习者。

① 黄显涵,李子建. 建构有效教学的策略:反馈理论[J].教育发展研究,2011,31(04):43-47.

案例 3-3 基于过程评价反馈生成个性化知识图谱补齐学习短板*

江苏省连云港市赣榆华杰双语学校学生在学习苏科版数学“探索三角形全等的条件”过程中需要先完成系统推荐的 12 个题目，之后系统会基于大数据、学科知识图谱构建技术，实时反馈学生的知识图谱画像，知识节点规划引擎会同时引入知识图谱节点之间的前驱后继拓扑关系，并用不同颜色呈现学习评价结果，如图 3-5 所示，系统会结合本省知识点的考察频度，推荐学生首先要攻克的知识点，并按照知识点先后次序有针对性的为该学生推送数学微课视频，学生完成推荐的资源后，系统会再次生成学情知识图谱，根据学生的学习情况，及时为学生重新推荐最佳学习路径，学生完成一轮智能学习机和教师推送的习题后，知识图谱发生了相应的变化，如图 3-6 所示，图谱中绿色部分变多说明有一定的提升，学生可以循序渐进、逐步掌握，避免了盲目学习，真正做到个性化学习。

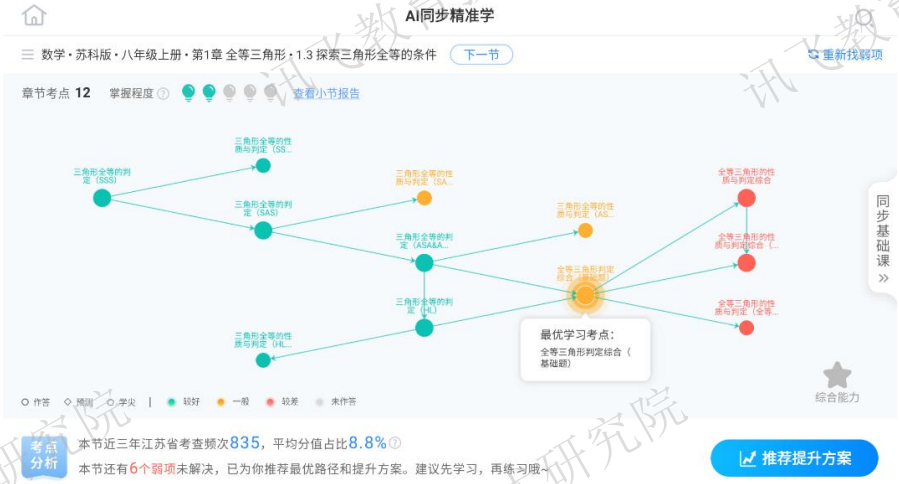


图 3-5 “探索三角形全等的条件”学习前知识图谱

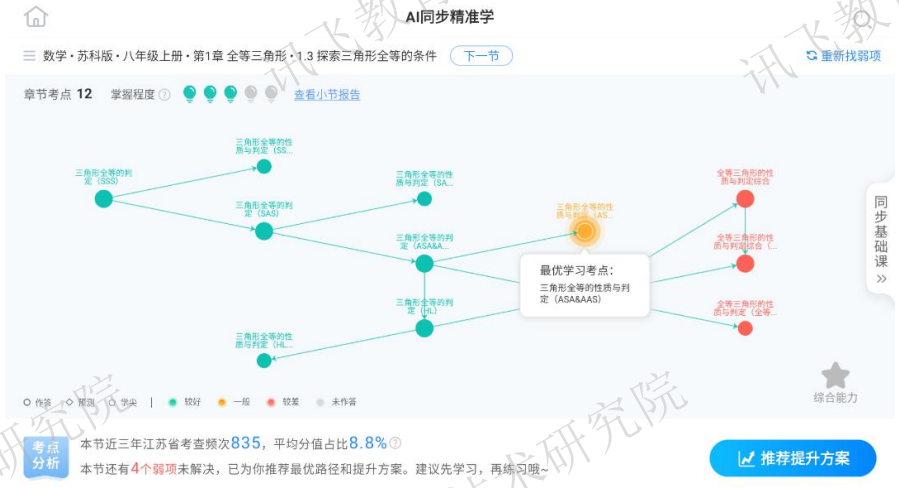


图 3-6 “探索三角形全等的条件”学习后知识图

* 本案例相关资料由江苏省连云港市赣榆华杰双语学校提供。

第 4 章　智能技术赋能增值评价

《总体方案》明确提出要“探索增值评价”。增值评价是国际上的一种前沿教育评价方式，通过对比学生在教育活动后获得的进步和成效来科学、客观地衡量学校和教师的工作成效。智能技术赋能增值评价，主要应用包括建立和维护追踪数据库、选择和应用增值评价模型、发展水平可视化呈现与交互等场景。

4.1 增值评价发展现状及面临挑战

增值评价是一种关注学生的学习起点和过程，通过采集学生在一段时间之前和之后的学业水平、理解能力、分析综合能力、表达和欣赏美的能力、情感态度等方面的表现数据，并对这些数据进行纵向比较，从而测量出此段时间内学生的成长程度，并以此为基础来评定学校、教师等教育主体的绩效，获得对学生的“净增值”的评价方法。从增值评价的概念内涵及其实际应用来看，其具有成长性、阶段性、激励性等鲜明特点。

在增值评价的实践中，由于其特殊性以及刚刚起步探索，客观存在一些问题与挑战，比如增值评价对数据的稳定性与可靠性要求高，传统数据存储与管理方法难以支撑；增值评价模型的选择与应用需要较强的专业性，一般教师与管理者难以上手；增值评价结果较为抽象复杂，普通用户难以准确理解和运用等，一定程度上阻碍了增值评价的落地与发展。

4.2 智能技术如何赋能增值评价

智能技术的发展与应用为探索增值评价提供了有效的解决路径，主要表现在以下三个方面。

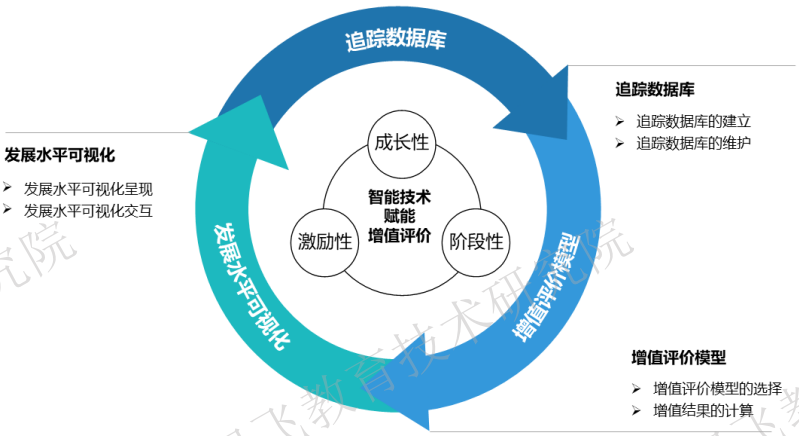


图 4-1　智能技术赋能增值评价

4.2.1 建立与维护追踪数据库

追踪数据库是增值评价的重要支撑部分，不但可以实现大规模追踪数据的可靠稳定存储，并对追踪数据进行链接与整合，便于数据的调用；而且可以在日常维护中处理追踪数据中的问题数据，并确保追踪数据中的敏感信息不被泄露，保护隐私安全。对于提高增值评价的科学性具有十分重要的意义。

在智能技术的赋能下，通过在系统架构、硬件选择上的特殊设计和智能化的管理与维护，追踪数据库相比于传统数据库而言可以有效提高追踪数据的稳定性与可靠性，从而为增值评价提供更好的支撑，使增值评价结果更加客观科学。如服务器虚拟化等技术的应用提高了追踪数据库对于硬件的利用效率，硬件方面的进步提高了追踪数据库的承载上限，分布式结构等数据库架构方面的优化提高了追踪数据库的可靠性与易拓展性。同时，人工智能算法的应用使得追踪数据库中大量的残缺数据、错误数据和重复数据的检测与清洗工作更加高效准确；数据库安全技术的进步能够更好地确保追踪数据中的敏感信息不被泄露，保护隐私安全。

案例 4-1 大型追踪数据库的建设与维护*

自上世纪 90 年代起，美国田纳西州就逐步建立了包含学生个体信息、教师信息和学生标准化测验成绩的一个庞大的、纵向合并的数据库，并通过学生的学号和地区、学校的编号等，将学生的学习成果与他们注册的学校以及在他们年级过渡时分配的老师对应起来，实现了信息的相互链接。同时对于因为种种原因导致的不完整的学生数据，每年田纳西州增值研究和评价中心（Value-Added Research And Assessment Center，VARAC）的分析师们都会在数据匹配的基础上编辑新数据，并通过加权来减少缺失数据的影响。截至 2010 年，田纳西州增值评价系统（Tennessee Value-Added Assessment System，TVAAS）数据库已经包含超过八十八万多名学生的追踪数据，记录总量超过二千六百万条，为 TVAAS 开展全州范围的增值评价提供了坚实的数据基础。

* 本案例根据美国田纳西州 TVVAS 相关论文、报道、专著、官网等资料整理总结。



图 4-2 教师发放 TVAAS 评价报告

注：图片来源于美国田纳西州州政府官网。

4.2.2 选择与应用增值评价模型

增值评价模型是增值评价的核心所在，可以对追踪数据进行分析，从而界定与计算“增值”。不同的增值模型有着不同的理论假设前提、数据适用条件和优劣势，基于不同种类的模型计算出的增值评价结果也各不相同，如何选择与应用适切的增值评价模型直接关系到增值评价结果的科学性、正确性。

在智能技术的赋能下，通过自适应推荐技术和智能的模型处理算法，增值评价模型可以根据用户的实际情况进行相应的调整与运算，提高增值评价的科学性和效率。如以自适应推荐等智能技术为基础，可以准确地对模型和实际数据的匹配程度进行计算，并推荐合适的增值评价模型，从而避免模型选择不当对增值评价结果的不利影响。同时，随着人工智能技术的进步，复杂的增值结果计算流程可以由计算机实现，有效提高运算的效率和准确性。

案例 4-2 百分等级增值评价模型的构建与应用*

苏州教育质量监测中心在每年面向全市初中开展的义务教育学业质量监测实践中构建并应用了“百分等级增值评价模型”。监测中心根据项目反应理论模型（Item Response Theory，IRT），采用锚题技术，根据学生作答情况，估计出学生能力分数，并将其转化为均分为 500、标准差为 100 的量尺分数，从而实现试卷等值。学生、学校乃至区域的百分等级将由量尺分数决定。在模型构建之初，为了更加客观科学地反映初中学校的增值情况，监测中心采集了苏州市从 2016-2018 年均参加监测且可追踪的学校数据，以这些学校各科上一年的百分等级和

* 本案例相关资料由苏州市教育质量监测中心提供。

下一年的百分等级画出散点图，进行函数的拟合，得出不同的函数模型，并在综合考虑模型的科学性、合理性、简洁性和一线老师的可接受性后选定其中一种，完成增值评价模型的构建。截至 2020 年，苏州监测累计参测学生超过 130 万人次，积累监测数据超过 1.8 亿条，累计发放 5000 余份针对区域和学校的监测结果报告，20 余份综合分析报告及 20 余份决策咨询报告，已经形成了成熟的应用。

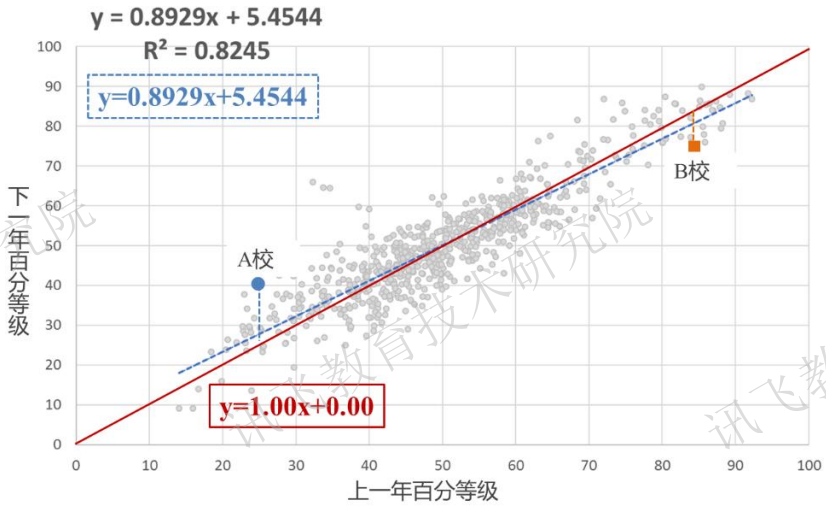


图 4-3 两种百分等级增值模型的比较

4.2.3 发展水平可视化呈现与交互

发展水平可视化是对增值评价结果的直观表现，能够准确直观地展示发展水平，将复杂的分析结果以更加直观、更易于理解的方式呈现给用户；并能通过不同的图表及说明文字，从不同的角度与侧重点进行发展水平的呈现，满足不同层次的用户群体对发展水平信息的个性化需求，是增值评价落地与推广的有力抓手。

在智能技术的赋能下，通过可视化技术以及可视化交互系统进行发展水平的可视化呈现与交互，可以高效直观地向不同的用户展示增值评价的结果，从而有效降低增值评价落地的阻力。如通过计算机图形学、计算机动画、计算机仿真等技术的进步使得可视化图表的种类和容量得以提升，能够展示更多类型和数量的数据，并可以用更丰富、更精美的表现手段将发展水平呈现出来，使其更加直观、美观、清晰，提升呈现的效果。同时，随着信息技术的进步，用户可以在可视化交互系统中通过更丰富、更快捷的人机交互方式实现对数据库内容的快速访问，更高效地获取信息，提高了信息传递的效率。

案例 4-3 发展水平可视化报告助力学校自我评估*

我国香港地区从 2003 年开始通过学校增值资料系统（Schools value added information System，SVAIS）为中学一年级至中学五年级和中学六年级至中学七年级的学校群组提供学校报告，其所提供的关于学校发展的各项数据是香港学校质量保证机制的重要指标，是学校外部评价、自我评价及促进学校教育教学质量提高的重要依据与保障。报告中使用了大量适切的可视化图形来对数据进行表征，极大地增强了报告的表现力与可读性，如在“学校学年报告”中，为准确表征学校历年间学生发展水平的增值情况，报告以箱型图为载体，通过不同置信区间对学校的增值得分进行刻画，一方面体现增值评价是一种相对性评价的本质；另一方面可以直观展现学生发展水平的进步幅度。同时辅以详尽的说明文字，助力一线教学人员理解运用。

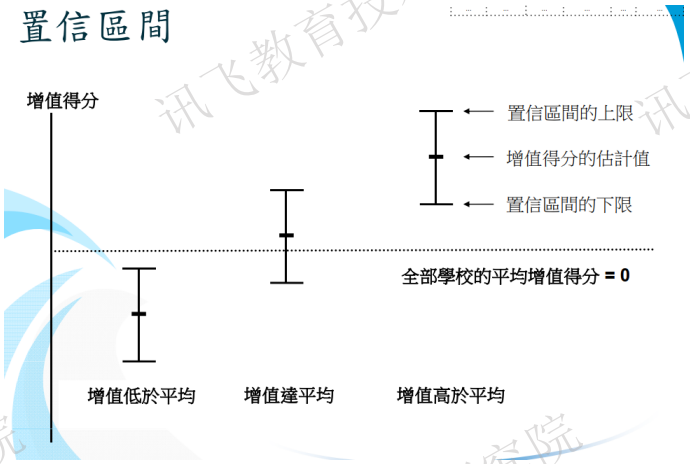


图 4-4 学校学年增值得分情况的可视化图形

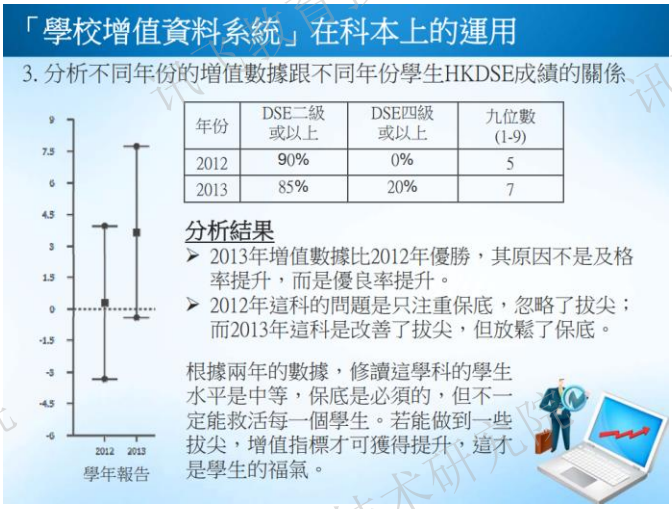


图 4-5 香港学业增值计算原理示意图

* 本案例根据中国香港地区增值评价实践相关论文、报道、专著、官网等资料整理总结。

第 5 章 智能技术赋能综合评价

《总体方案》明确提出要“健全综合评价”。综合评价注重对评价对象进行全面、综合、整体的评价，对于扭转片面化的评价倾向具有重要意义。智能技术赋能综合评价，主要应用包括评价指标模型构建与优化、全域评价数据采集汇聚、数据深度挖掘分析、数字画像建立与应用等场景。

5.1 综合评价发展现状及面临挑战

综合评价的本质在于综合，关注评价对象的整体性、合成性。综合评价是一种从全方位关照被评价对象的思维角度，通过设计科学的综合评价指标体系，对评价对象进行系统、全面和完整的综合评判，综合反映评价对象的总体特征，全面考量和判断评价对象的价值。综合评价能够综合多方意见、采取多种方法和视角对评价对象进行整体性评价，在评价实践中呈现系统性、多样性、复杂性、全面性等特点。

目前我国综合评价应用实践最广泛的主要是学生综合素质评价、教师综合评价、中小学教育质量综合评价等，进行了积极探索并取得了一定进展，但是在评价指标构建的科学性、评价数据采集和处理、评价结果客观公正等方面仍然面临诸多问题与挑战。

5.2 智能技术如何赋能综合评价

智能技术的发展与应用为健全综合评价提供了有效的解决路径，主要表现在以下四个方面。

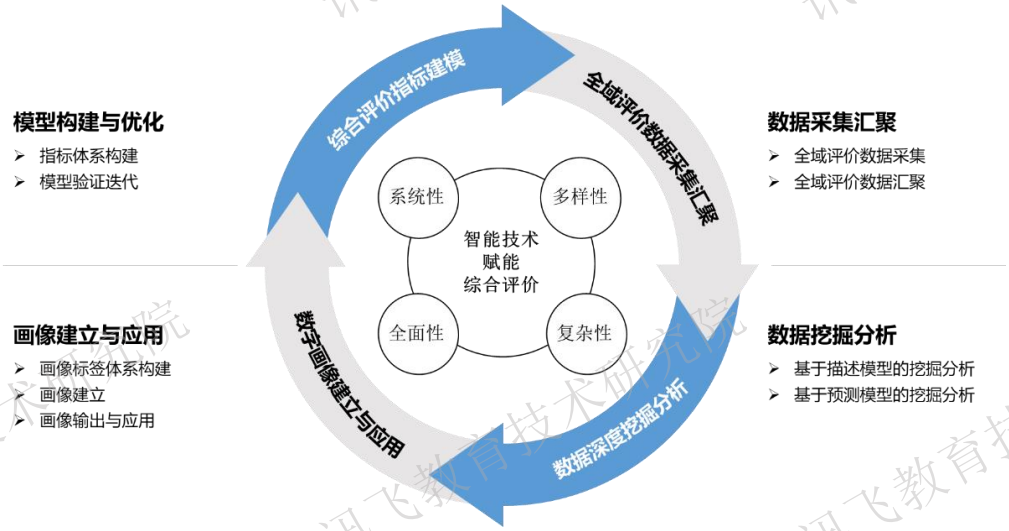


图 5-1 智能技术赋能综合评价

5.2.1 综合评价指标模型构建与优化

在教育评价领域，综合评价指标模型是评价指标体系内在逻辑关系和数学关系的具体体现，是获取综合评价结论的重要途径和工具[①②]。智能技术在综合评价指标建模中的运用，有助于以科学、高效、简化的指标来反映各地教育综合评价的发展状况，进而系统、全面、有效地监测各地学生成长与发展的整体现状，为地方和国家制定促进学生综合素质、提升教育质量的有关教育政策、教学对策提供科学的测评依据。对于全面提升综合评价效果和评价效率是十分有必要的。

构建智能化的综合评价指标模型，能够系统地、科学地分析和处理评价指标的各类复杂问题。较为常用的智能化方法包括专家系统、机器学习、人工神经网络、遗传算法等，对于科学筛选指标和计算指标权重，确保用精简的关键指标反映评价对象系统的主要特征、判断该指标对评价结果的贡献度发挥着至关重要的作用[③]。此外，智能技术的引入能够从模型的应用目的出发，考查模型在其应用过程中是否准确地代表了原型系统，并基于一定的验证方法和验证工具，能够对得到的结果进行验证，通过不断的修正补充，直到出现科学合理的评价结果，实现对模型的更新与优化。

案例 5-1 学习者综合评价参考模型构建*

北京师范大学远程教育研究中心通过理论演绎和专家访谈构建了以投入度、完成度、调控度、联通度和主动性为核心的综合评价参考理论模型；并通过学习行为数据聚合特征变量，构建了相应的计算模型。在此基础上，通过在线学习平台汇聚的学习数据，进行了从行为数据到特征数据，再到维度计算，设计和研发了相应的工具，并通过华中科技大学远程与继续教育学院的实践应用验证了学生综合评价参考模型的科学性和可用性，实现将定性与定量建模应用于在线教学评价模型的构建，打通了教育学、心理学、数据科学之间的阻隔，从教育基本规律、认知发展规律以及数据科学角度建立了关于学生综合评价等系统性的模型体系，一定程度上解决了学习者、教师以及课程评价科学性与客观性不足的问题，从而实现了提升学习质量的最终目的。

*　本案例相关资料由北京师范大学远程教育研究中心提供。

① 杨勇. 智能化综合评价理论与方法研究[D].浙江工商大学,2014.

② 陈玉祥,朱桂龙. 科学选择的理论、方法及应用[M].北京:机械工业出版社.1994:161-162.

③ 彭张林,张爱萍,王素凤,白羽.综合评价指标体系的设计原则与构建流程[J].科研管理,2017,38(S1):209-215.

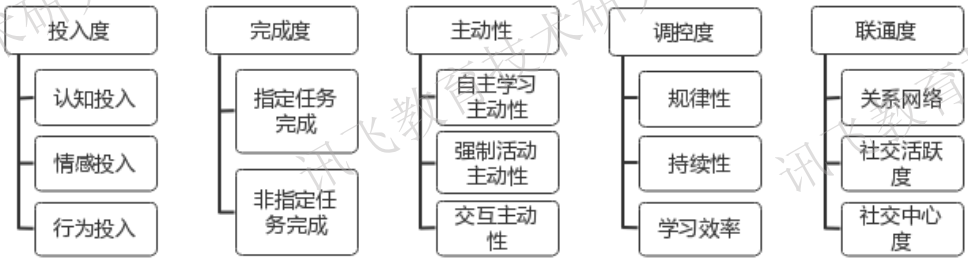


图 5-2 综合评价参考理论模型

5.2.2 全域评价数据采集汇聚

全域评价数据采集汇聚[①]是有效提升综合评价科学性、专业性、客观性的基础支撑，可在全时空对评价对象进行数据采集，采集的全样本、混杂、海量数据再汇聚成一个立体化的智能“网络”，最终实现对评价对象整体发展情况的全域关注与监测[②]，为客观真实的综合评价提供可靠数据支撑。

全域评价数据采集汇聚需要因地制宜、多源归一，应根据数据采集场所的变化灵活的选择一种或多种采集技术，将不同来源的海量数据统一集成并有序存储。通过物联感知、平台采集、图像识别、视频录制等技术，可以采集评价对象在各个时空下的数据，如学生在学习、生活、运动、社会实践中的相关数据，以及道德品质与公民素养、学习品质与文化素养、身体品质与健康素养、艺术品质与审美素养、实践品质与创新素养等支撑综合评价的相关数据。同时，通过数据集成技术高效处理不同数据类型的海量数据，并快速汇聚到数据中心，实现评价数据从“碎片”到“集约”。

案例 5-2 基于学生综合素质评价平台实现数据采集汇聚*

2019 年 10 月，上海市普陀区基于人工智能和大数据打造学生综合素质评价平台，开始在 7 个学校探索德智体美劳的评价指标体系。该平台构建了大数据采集链，通过系统对接和信息填报等方式，实现包括学业质量、艺术素质、科学素养、品德发展、劳动教育、体育素养等多方面评价数据的采集。平台在数据采集过程中遵循统一标准的多种异构数据源接入，支持来源不同学校多源数据的汇聚。平台目前采集了 2 万条数据，打通可用于学生成长记录和评价的底层数据，实现“五库联动”，凸显“五育并举”，助力实现学生学习现状、教师教学现状、学校发展现状的智能诊断，辅助管理者概览学生综合素质评价总体水平，指导区域综评工作开展和资源投入。

* 本案例根据上海市综合素质评价相关论文、报道、官网等资料整理总结。

① 田爱丽. 综合素质评价：智能化时代学习评价的变革与实施[J].中国电化教育,2020(01):109-113+121.

② 杨鸿,朱德全,宋乃庆,周永平. 大数据时代学生综合素质评价：方法论、价值与实践导向[J].中国电化教育,2018,372(01):27-33.

案例 5-3 基于立德树人监测与评价平台实现全链路监测*

自 2019 年起，合肥市教育局开始探索构建服务中考改革的学生综合素质评价系统，采集学生过程性数据。2021 年 6 月，合肥市教育局明确以平台为载体，结合智能物联设备，推动合肥市初中学生综合素质评价系统改造升级，构建面向初中学生多维度的综合评价，能够对学生思想品德、学业水平、艺术素养、身心健康、社会实践等过程性数据进行写实记录和思想品德测评、艺术素养评估、体质健康监测、心理健康测评、学业水平评估等结果性数据进行评估监测，从数据采集、指标合成、数据拟合、评价报告生成与发布等方面实现学生综合素质评价的全链路监测。

5.2.3 数据深度挖掘分析

数据深度挖掘分析是获得更科学、更有价值综合评价结论的必要路径。通过对不同来源、不同类型、不同结构的评价数据进行深度挖掘分析，深刻揭示评价数据在特定时空中的变化过程与特征，挖掘变量间隐含的关系和深层规律，从而整体把握评价对象的发展过程、发展特点与发展趋势，助力对评价对象作出整体性、科学性的综合评价。

数据深度挖掘分析可以从大量的数据中提取有效信息，将其转化为易于理解的结构以供进一步使用，选择适当的挖掘技术与方法才能取得较好的应用效果。如基于描述模型的挖掘分析，通过分类聚类算法、关联规则挖掘、过程挖掘、文本挖掘、图构建与挖掘等方法来分析评价数据，发现新的模式或结构；按性格类型、学习风格等维度对学生进行分类，对师生关系和学业成绩间进行关联分析，对学生同时选修多门课程间的关联分析等。如基于预测模型的挖掘分析，通过多元回归分析、马尔科夫预测等方法，深刻揭示评价对象的未来发展趋势；为综合评价学生潜能、心理成长等提供科学依据，根据学生的初中各科历史成绩预测学生高中选考的科目。

案例 5-4 行为数据挖掘助力教师综合评价*

为助力教师综合评价，讯飞教育技术研究院以教师信息化教学工具应用能力评价为例，随机选取“因材施教”人工智能+教育创新应用行动项目中的两所中小学共计 313 名教师的平台交互行为数据进行挖掘分析，不再简单的以频次、时长统计、分数或等第制对教师信息技术应用能力进行评价，而是深度挖掘数据中的潜在关系与价值，为教师综合评价及管理决策制定提供基于数据的证据支持。

---

* 本案例相关资料由合肥市教育局提供。
* 本案例相关资料由“因材施教”人工智能+教育创新应用行动项目组提供。

教师信息化教学工具使用的行为数据分析过程具体如图5-3所示，利用人工智能、机器学习、模式识别、可视化等数据挖掘技术助力教师信息化教学工具应用能力评价，能够通过分析教师多方位、多角度、跨时段的行为数据，被评教师不仅能获知自己的相对位置，更能掌握信息化教学工具应用积极性、风格习惯等多维度的信息，帮助教师找到自身问题所在，为管理者打通“教、评、培”完整闭环、建立定制化培训方案起到了关键性的指导作用。

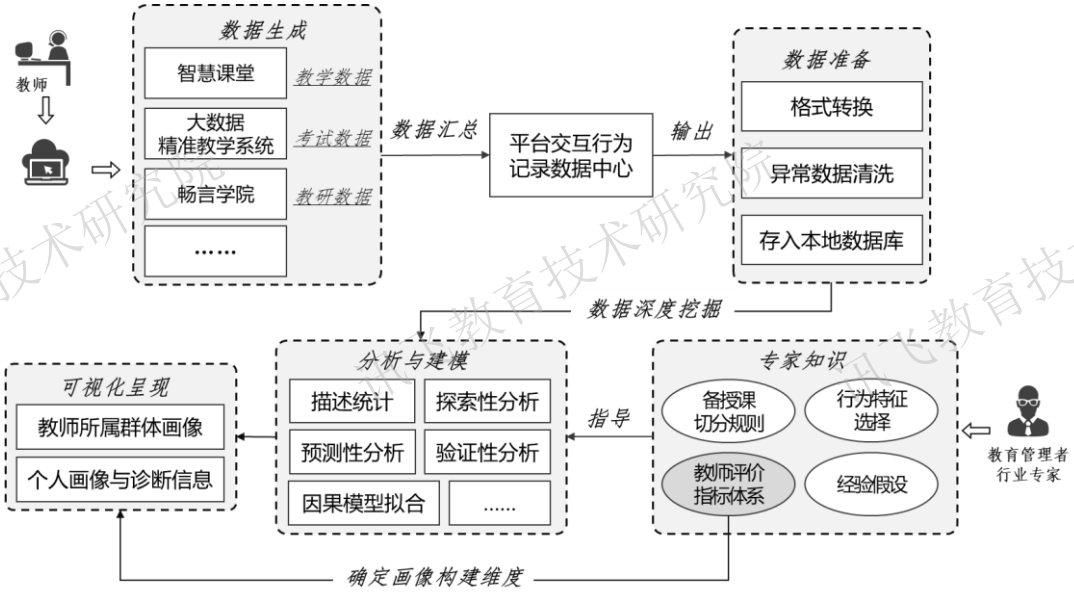


图 5-3 教师信息化教学工具使用的行为数据分析过程

5.2.4 数字画像的建立与应用

数字画像是以大数据、学习分析等技术为基础，通过基本信息、行为数据、心理数据等多源数据构成的代理原型，用以支持教育决策、改进教育管理和优化教育服务[①]，是综合评价实现精准诊断、及时干预和个性化支持的基础，能够为综合评价呈现更精确的评价过程分析和更精准的内容推送，并助力评价对象进行资源共享、行为调整、体验改善、效率提升。

在智能技术的赋能下，通过对综合评价数据的记录、分析、挖掘与呈现，并以自动化方式处理分析，提炼出可描述其特征和行为的标签集，最终可从各维度完整刻画数字画像，为综合评价满足个性化需求、实现精准服务提供了更多可能。如在构建画像标签体系的基础上，通过各类数据挖掘的方法分析数据，以画像的形式呈现计算结果，实现标签画像库的建立。在画像输出与应用的过程中，根据画像目标和综合评价数据，利用数据可视化、虚拟现实、区块链等技术的相互配

① 艾兴,张玉. 从数字画像到数字孪生体：数智融合驱动下数字孪生学习者构建新探[J].远程教育杂志,2021,39(01):41-50.

合，将数字画像以图示的方式对抽象的内部结构加以可视化处理，呈现学习者画像，并保障画像的安全与应用。

案例 5-5　中小学智慧教育质量画像构建与应用*

2019 年蚌埠市依托大数据和人工智能技术构建大数据平台，以数据的管理、可视化、开发利用为牵引，生成蚌埠市中小学智慧教育质量的画像，为区域的智慧教育综合评价提供数据支撑。大数据平台能提供丰富的可视化模板，采用“搭积木式”的开发手段，配置图形分析工具的应用，使数据分析结果以简洁的方式进行展现，如下图 5-4 所示：以饼图表示教师智慧教学中备课、授课和互动的参与率；以折线图表示资源建设和使用随时间变化的趋势；利用 Cascading Style Sheets 或者 Javascript 来实现学科资源分布情况的滚动呈现效果。通过对蚌埠市中小学智慧教育质量画像的构建和输出，能够为管理者的科学管理提供数据支撑和全方位的宏观概览，全面、动态、实时地掌握蚌埠市智慧教育建设情况，从辅助教育决策、改进教育质量、促进区域教育均衡等方面提供高质量、专业的教育数据创新应用。

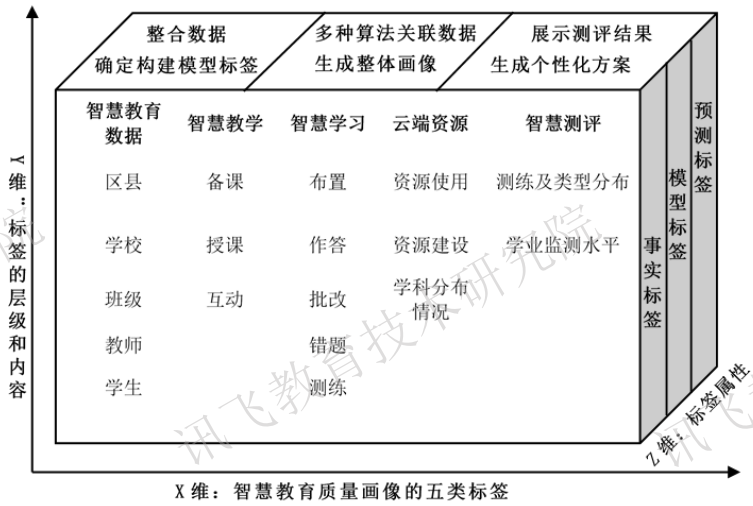


图 5-4　蚌埠市中小学智慧教育质量画像标签体系

*　本案例相关资料由蚌埠市智慧学校建设项目组提供。

第 6 章 智能技术赋能教育评价典型实践案例

智能技术对教育评价的影响不是单方面的，它能够从整体上对教育教学评价活动的多个层面产生深刻影响，助力构建新型教育评价体系。在教育评价实践中，智能技术在基础教育质量监测、学校发展评价、在线课程评价、教师课堂教学评价、学生综合素质评价等方面具有重要的支撑作用。

6.1 智能技术赋能教育评价实践概述

在教育评价实践中，由于教育评价本身的复杂性，因此在利用智能技术促进教育评价改革时，可以从多个角度切入研究和阐释。从当前研究来看，智能技术赋能教育评价的相关实践大致可以分为两个视角，如图 6-1 所示。

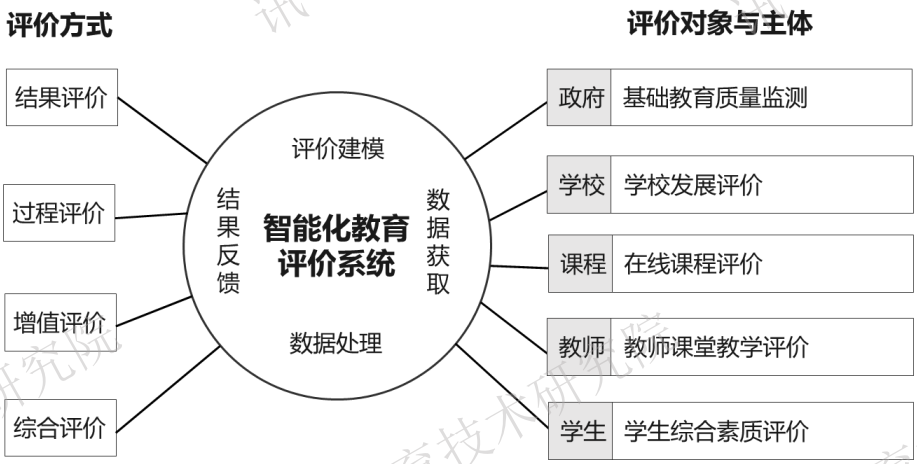


图 6-1 智能技术赋能教育评价实践的两个视角

①从评价方式视角看智能技术赋能教育评价。该视角侧重于探讨智能技术在不同评价方式中的赋能作用。随着国家考试评价改革文件的发布和新一轮课程标准的修订，国家对学生的全过程纵向评价和全要素横向评价的要求越来越高，导致了教育评价方式发生了改变，与以往相比呈现出较大不同，从基于纸笔测验、观察、计算机等传统手段的单一评价方式转向人工智能、大数据等智能技术支撑的多样化评价方式，具体包括智能化结果评价、过程评价、增值评价、综合评价等方式，分别对智能技术的作用发挥提出了不同的要求。该视角的研究旨在突出某种类型的教育评价在智能技术支持下的优化和完善，强调赋能教育评价的智能技术及应用场景。

②从评价对象与主体视角看智能技术赋能教育评价。该视角侧重于基于不同评价对象的现状和问题，借助智能技术的支持，通过精心设计评价方案，认真组

织实施评价过程，更科学和高效地达到评价目的，强调技术支持下教育评价本身的变革与创新。政府、学校、课程、教师、学生是五个典型的评价对象与主体，从基础教育质量监测、学校发展评价、在线课程评价、教师课堂教学评价、学生综合素质评价五种类型的综合实践案例来展开描述，有助于从实践主体角度完整的呈现技术支持下的不同评价改革与发展情况。

6.2 智能技术赋能基础教育质量监测

基础教育质量监测是根据国家制定的有关方针、政策、法规等，定期对基础教育质量进行评估与判断，利用数据了解并改进基础教育的外部评估活动。我国基础教育质量监测的实验与探索已经取得显著成果，每年实施国家义务教育质量监测成为基本国策，但整体也存在着监测内容侧重于学业成绩、监测手段传统单一、监测实施与应用规范性不足等问题。随着人工智能、大数据、可视化分析、智能评测等技术的发展与应用，我国正在逐步实现智能技术赋能的教育质量监测，完善国家和区域义务教育质量监测体系。

（1）国家义务教育质量监测案例*

国家义务教育质量监测是基于中国国情形成的一种基础教育质量监测形式。通过多年周期性监测的开展，国家义务教育质量监测从工作布局、组织实施到结果发布等一系列环节的建设之路逐步成型，实现了监测工具、监测内容、监测范围和监测手段等多方面的突破，智能技术逐步在监测实施、结果反馈与应用等多个环节中发挥重要作用。

①智能技术支撑监测数据的多样化采集。随着人工智能、大数据等技术的不断发展，部分学科开始使用技术支撑监测过程，如通过“国家义务教育质量监测问卷调查系统”完成问卷调查、测评进度可视化呈现等；通过“演唱测试系统”完成电子化测试、实时数据采集等。以“演唱测试系统”为例进行详细介绍。该系统在 2016 年和 2019 年的艺术学科现场测试项目中被用于测查学生的演唱能力[①]，系统依托科大讯飞人工智能技术，通过语音合成、语音评测、虚拟现实、模拟人机对话等，实现了语音数据的采集、储存、评分一体化和智能化。

②智能技术支撑监测结果的可视化反馈与应用。在实际反馈过程中，联合专业机构将报告以易理解或可视化形式呈现，为教育行政部门和学校提供结果的可视化反馈和解读服务[②]，是当前监测结果反馈与应用的一个重要趋势。如，某区

* 本案例相关资料由中国基础教育质量监测协同创新中心提供。

① 严谨. 用好“体检报告”助力质量提升——2019 年国家义务教育语文、艺术学习质量监测结果报告解读[J].人民教育,2020(17):34-37.

域联合科大讯飞建立了区域义务教育质量监测结果可视化平台，实现了对区域义务教育质量的有效监测。目前，该可视化平台主要包含国家监测结果指标框架与数据服务、管理决策服务、分类诊断服务、监测图谱服务等功能。以管理决策服务为例，区域教育管理者可以借助该平台的可视化数据分析结果，科学采用学校软硬件管理改进等措施，及时修正教育决策和行为，促进区域教育科学、全面、健康发展。

此外，随着智能技术的发展，国家义务教育质量监测的整项监测样本抽取工作主要通过国家义务教育质量监测平台中的信息上报及抽样系统自动化完成；监测数据与检测报告也通过相关专业软件和系统完成自动化分析、智能化生成，系统体现了国家义务教育质量监测手段、监测实施多元智能的特点。

（2）苏州市义务教育学业质量监测案例*

苏州市教育质量监测中心根据项目本土化的个性需求，通过自主研发、合作共研和购买成熟服务等方式，全面构建了学业质量监测全环节的智能技术支持系统。基于该系统的监测结果，苏州市开展了系列督导评估工作，为苏州市初中教育质量的提升打下了坚实的基础，并逐步实现了标准化数据治理、智能化数据分析、可视化数据呈现等技术突破。

①智能技术支撑标准化的监测数据采集。依托考务管理系统完成学校信息上报、学生抽样、考务监控、考务编排等组织准备工作。确认考务数据后，也制定了规范严谨的数据采集标准化流程，如利用扫描阅卷系统提供纸质工具的数据采集服务，依托问卷星问卷逻辑判断系统自动转跳、隐藏、显示题目，简化了后期数据清理工作，大幅提升工作效率与数据质量。

②智能技术支撑智能化的监测数据分析。通过将监测各环节的数据处理流程不断标准化，并将算法和运行程序固化成软件模块，构建了苏州监测数据分析系统。通过该系统可以实现监测数据的自动清理、质量分析、计算分析、数据推送等。

③智能技术支撑自动化的监测报告和可视化的监测结果。通过报告生成和大数据展示两大系统支撑监测报告的自动生成和监测大数据的可视化展现。监测报告生成系统：在报告生成前端，开发了图表集成系统对各类数据图表进行审校，随后通过报告生成系统批量生成市、区、校各级报告，包括主报告、学科数据报告、相关因素报告等。监测大数据展示系统：该系统可以呈现市、区和学校层面的参测基本信息的导航地图，以及学业水平、学业增值、学生相关因素（含阈值

* 本案例相关资料由苏州市教育质量监测中心提供。

预警）、教师相关因素四个维度、24 个关键指标，构成了区域和学校的多维度、立体化、多层次的数字画像，如图 6-2 所示。



图 6-2 苏州市义务教育学业质量监测大数据展示平台示意图

6.3 智能技术赋能学校发展评价

学校发展评价，也称发展性学校评价、学校发展性评价等，是指评价主体以促进学校良好发展为目的而对学校的活动及状态所进行的价值判断[①]。学校发展评价注重在分析学校现状及其主要优势、劣势的基础上，发现和发掘学校的发展潜力，找准学校的最佳发展区和生长点，指导和帮助学校发扬优势、弥补不足，逐步形成学校特色[②]。2020 年 5 月，上海市浦东教育发展研究院督导事务中心建设了浦东新区教育督导系统（以下简称“系统”），基于系统开展学校督导，实施学校发展评价，推动浦东新区义务教育的优质均衡发展和素质教育的全面实施。

浦东新区学校督导评价分为小学督导评价、初中督导评价和普通高中督导评价三个层次，不同层次学校的督导评价内容有所差异，主要是考虑到学校自身发展的特点和需求。以浦东新区第二中心小学为例，上海市浦东教育发展研究院督导事务中心于 2020 年对该学校进行了督导评价*。其中，智能技术的赋能主要体现在以下几个方面。

①问卷在线调查与智能分析。督导评价小组通过系统面向教师、学生、家长发放问卷，收集不同角色对学校管理、学生德育、课程教学、师资队伍等各个指标的评价。在问卷提交后，督导评价小组可以通过系统实时查看问卷调查的进度

* 本案例相关资料由上海市浦东教育发展研究院督导事务中心提供。
① 史晓燕. 发展性学校评价模式探索[J].教育探索,2004(10):21-23.
② 卢立涛. 试论发展性学校评价的内涵与特点[J].教育测量与评价(理论版),2010(04):12-17.

和统计分析结果。对于客观题，系统可以基于数据分析模型对数据进行自动统计分析；对于主观题，系统可以利用自然语言处理引擎等技术对文本信息进行结构化处理，降低识别主观评价内容的复杂性。同时，系统中的内容安全监管引擎还可以对数据进行自动化监管，保障数据的科学性、合规性。

②可视化归类呈现课堂观察结果。督导评价小组根据排课表，通过线上、线下观课等多种方式开展课堂观察，并从教师行为、学生行为、教学成效等维度对课堂教学进行打分并点评。系统对所有专家的评课分数进行统计分析，形成可视化的统计图表，并将专家在班级文化、主要优势和改进之处等方面的质性评价内容可视化归类呈现。

③自动生成学校督导评价档案。对于督导评价小组的赋分分值，系统基于数据分析模型，生成学校在各模块的得分；对于督导评价小组填写的评价内容，系统基于自然语言理解等人工智能技术，将专家填写的学校发展中的问题和亮点等意见进行关键词文本分析。经过以上分析，系统自动生成学校督导评价档案，如图 6-3 所示。

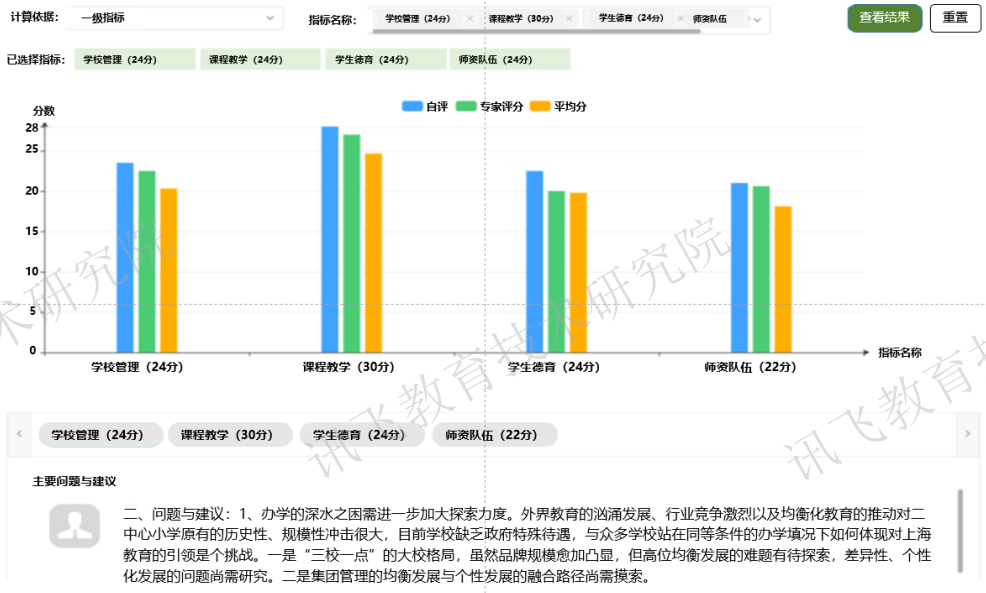


图 6-3 学校综合督导结果分析

6.4 智能技术赋能在线课程评价

智能技术赋能在线课程评价主要体现在：借助在线学习平台自动采集、存储、分析和呈现学习者的在线行为数据[①]，建立预测模型或综合评价模型，提供学习

① Mazza R, Bettoni M, M Faré, et al. MOCLog–Monitoring Online Courses with log data[C]// Moodle Research Conference. 2012.

分析报告，促进反思与策略调整。以下两个典型案例体现了智能技术如何赋能在线课程评价。

（1）犹他州立大学在线课程评价案例*

美国犹他州立大学基于 Canvas 平台的课程评价，包括对课前（课程内容质量）、课中（互动活跃度）和课后（课程通过率）全过程的评价。技术的应用主要体现在课中对在线学习活动中的互动行为进行评价，通过师生互动、生生互动、以及学生和内容互动等多个指标体现课程互动的活跃度，具体涉及行为数据采集与处理、特征提取与指标构建等步骤。

①互动指标的构建。在线课程的互动包括师生互动，生生互动，学习者和教学内容的互动，以及教师和教学内容的互动四种类型[①]，与之对应的是四种互动指标，其指标构建方法如表 6-1 所示。

表 6-1 四类互动指标的构建

互动指标	行为数据		指标构建
教师-内容（IC）	ic_atta	教师上传附件	(ic_atta+ic_disc+ic_wiki+ic_quiz+ic_assi)/5
	ic_disc	教师发布讨论主题	
	ic_wiki	教师发布 wiki 主题	
	ic_quiz	教师布置随堂练习	
	ic_assi	教师布置家庭作业	
学生-内容（SC）	sc_atta	（平均）学生查看附件	(sc_atta+sc_disc+sc_wiki+sc_quiz+sc_assi)/5
	sc_disc	（平均）学生查看讨论	
	sc_wiki	（平均）学生查看 wiki 主题	
	sc_quiz	（平均）学生完成随堂练习比例	
	sc_assi	（平均）学生完成家庭作业比例	
学生-学生（SS）	ss_disc	（平均）学生发表的讨论消息（初始消息和回复消息）	ss_disc
学生-教师（SI）	si_disc	教师发布的讨论消息（初始消息和回复消息）	si_disc

②数据采集与处理。该校的在线课程都依托于 Canvas 学习管理平台，Canvas 提供各种标准化 API 接口，自动记录用户的行为数据，项目团队通过对行为数据进行特征提取和指标构建，结合统计分析方法计算出用户互动的复合指标，由此表征课程互动的活跃度。项目团队通过 Canvas 的系统埋点自动采集教师和学生的行为数据，按照表 6-1 的定义计算四类互动指标。在最初选择的 202 门课程中，

* 本案例根据犹他州立大学在线课程评价相关论文等资料整理总结。

① Anderson T, Garrison D R. Learning in a Networked World: New Roles and Responsibilties. Distance Learners in Higher Education: Institutional Responses for Quality Outcomes, 1998.

有 81 门课程因为数据缺失而无法得出运算结果，项目团队最终得到了 121 门课程的互动活跃度指标。

（2）北京师范大学《互联网+教育》课程评价案例*

北京师范大学的《互联网+教育》cMOOC 课程评价，在评价方式上使用了学习者个人学习报告，包含了对学习者的学习状态、个体活跃度和内容贡献度等多个维度的综合性评价。该课程的学习平台主要通过 WordPress 实现内容管理，通过 Odeon 平台[①]实现数据汇聚。WordPress 包含模版和插件两个子系统。模版系统支持 cMOOC 平台的框架和主要功能（教学功能），插件系统支持 cMOOC 平台的功能迭代，包括教学数据分析、新增的评价报告等。Odeon 大数据平台主要负责数据采集、数据管理、数据分析、数据推送等。数据计算的结果经过可视化处理后输出成不同的图表，如反映学习者画像的雷达图、反映个体活跃度的社会关系图、以及反映内容贡献度的概念网络关系图等，这些图表在个人学习报告中呈现。

①个体活跃度评价。个体活跃度评价对应三类学习行为的可视化分析，分别是交互行为统计、交互时段统计和社会网络关系，通过这三个指标全面直观地反映学习者的个体活跃度。个体活跃度重点评估学习者在平台的互动情况。社会交互可以用社会网络关系来表征。在 cMOOC 平台中，社会网络关系是对平台参与者们关注、点赞、评论等交互关系的表征。

②内容贡献度评价。课程团队认为学习者对内容的贡献度可以用概念网络关系来表征，提出了描述概念网络特征水平的三个评价维度——概念数量贡献度、概念热度贡献度和认知参与度。学习者在积极发表观点和评论的过程中，系统能够自动捕捉、分析和提炼学习者的热点关键词并绘制成词云图，便于学习者回顾课程内容。此外，系统支持“相似关键词”用户推荐功能，帮助学习者发现与自己志同道合的学习伙伴，有助于学习社区互动交流。

③学习状态评估。cMOOC 平台会计算每位学习者在每个维度上的得分，通过雷达图对学习者的学习状态进行可视化。雷达图能直观的反映学习者的风格偏好，例如在图 6-4 中的学习者在反思维度上得分最高，说明学习者的学习行为倾向于反思型。“当前平台最高水平”和“证书标准”可以为学习者提供参照物，帮助学习者清晰客观的了解自身的学习状态，并促进学习者进行改进。

---

* 本案例相关资料由北京师范大学《互联网+教育》课程团队提供。

① Odeon 是由科大讯飞推出的基于图形化 Web 操作界面的一站式大数据平台，具备数据采集、任务编排、调度处理、数据展现、权限管理、日志追踪、集群监控等功能。



图 6-4 学习者学习状态评估雷达图

6.5 智能技术赋能教师课堂教学评价

教师课堂教学评价是在课堂教学中，根据课堂教学理论和教学规律，运用观察的手段收集课堂教学过程的事实性材料，对教师行为及其行为所导致的结果作出价值判断的过程[①]，具有诊断反馈、激励向上、改进教学等功能。课堂教学评价是教育评价的研究重点和重要组成部分，也是促进教师专业发展、提升课堂教学质量、深化教育改革的重要手段和关键环节。下面以蚌埠高新教育集团实验中学（以下简称为“蚌埠实验中学”）为例，说明智能技术赋能课堂教学评价的实践情况*。2021 年，蚌埠实验中学依托智慧课堂信息化平台，利用认知智能国家重点实验室智能教育研究中心开发的“智慧课堂教学评价指标体系”开展课堂教学评价。蚌埠实验中学基于智能技术开展的课堂教学评价中的主要应用有：

①自动采集和分析课堂教学过程数据。根据学校的智慧课堂教学评价方案和“智慧课堂教学评价指标体系”，明确了需要采集的数据，如课堂实录数据、课堂互动数据等，通过智慧课堂信息化平台可以伴随式采集这些课堂教学的过程性数据。此外，针对采集到的数据，平台可以自动进行数据分析和建模。图 6-5 展示了蚌埠实验中学罗老师《函数》一课的微课录制情况，包括八年级七班该节数学课的课堂教学实录视频和实录文本。

* 本案例相关资料由蚌埠实验中学提供。

① Gardner H E. Intelligence reframed: Multiple intelligences for the 21st century[M].Hachette UK, 2000.



图 6-5 八年级七班数学课课堂实录

②智能生成课堂教学报告。智慧课堂信息化平台可以对采集到的课堂教与学数据进行分析，并自动生成可视化课堂教学报告。学校管理者、教研人员及教师自己可以查看这些报告，辅助开展课堂教学评价和反思。课堂教学报告主要包括课前学情诊断报告、实时测评和互动报告、课堂教学风格报告等几种类型。例如罗老师《函数》一课的课堂风格报告，如图 6-6 所示。

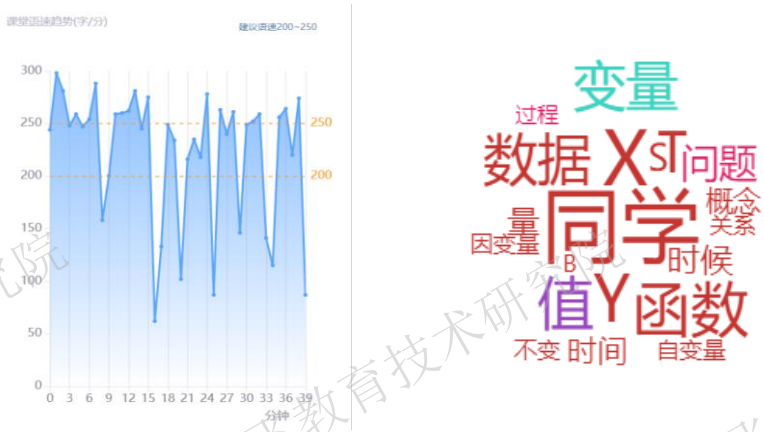


图 6-6 课堂教学风格报告（部分）

③基于报告进行评价与反馈。数学学科智慧课堂评价小组成员根据教师的教学设计、教学视频和系统提供的教学分析报告，使用智慧课堂教学评价表对学校数学教师的课堂教学进行评价，同时，每位被评教师也使用评价表进行自我反思。一轮评价活动后，评价小组对评价数据进行汇总统计和详细分析，并将评价结果和改进建议反馈给教师。通过此种方式可以帮助教师发现、总结智慧课堂教学中的问题，进一步改进和优化教学。

6.6 智能技术赋能学生综合素质评价

综合素质评价是对学生全面发展状况的观察、记录和分析，是发现和培育学

生良好个性的重要手段，可以改变教师评价观念，破除“唯分数论”，是深入推进素质教育的一项重要制度。近年，政策文件纷纷倡导积极利用现代化技术手段赋能学生综合素质评价。例如南京师范大学附属中学*（以下简称“南师附中”）应用智能技术促进学生综合素质评价，取得了良好的成效。南师附中根据教育部和江苏省对学生综合素质评价的要求，结合学校培养目标和办学特色，建立了南师附中学生综合素质评价指标体系，利用智能技术开展学生综合素质评价取得了良好的成效。南师附中利用智能技术促进学生综合素质评价的主要做法有：

①多种方式全面采集学生综合素质评价数据。一是写实记录。记录主体以学生为主，教师和管理员等其他主体为辅，及时、客观记录学生参加各项活动的相关情况和成果。二是数据导入。通过平台对接、接口导入等方式采集学生综合素质评价数据，既保证了数据的真实性和及时性，又减少了填报程序，避免师生重复填报。

②学生综合素质评价内容监管和数据审核。对于学生上传的内容，南师附中采用了多种方式进行内容监管和数据审核，为内容的安全性和真实性层层把关。第一，利用智能技术，自动进行内容审核。快速识别含有色情、暴力恐怖、涉政谣言的内容。第二，教师进行审核，确认信息真实。第三，公示信息，保障评价过程透明。第四，建立诚信体系，作为进一步保障学生综合素质评价数据真实、结果公正的重要措施。

③基于数据统计分析进行过程性调控。评价过程中，学校通过系统可实时掌握评价任务进度、学生发展情况等方面的数据，从而基于数据统计分析进行过程性调控，及时精准发现问题并干预。首先，南师附中基于系统统计的评价任务进度进行调控，其次，基于系统记录的学生发展情况进行调控。例如，学校管理者通过系统查看 2020-2021 学年上学期高二年级各指标的记录次数，结果如图 6-7 所示。

* 本案例相关资料由南京师范大学附属中学提供。

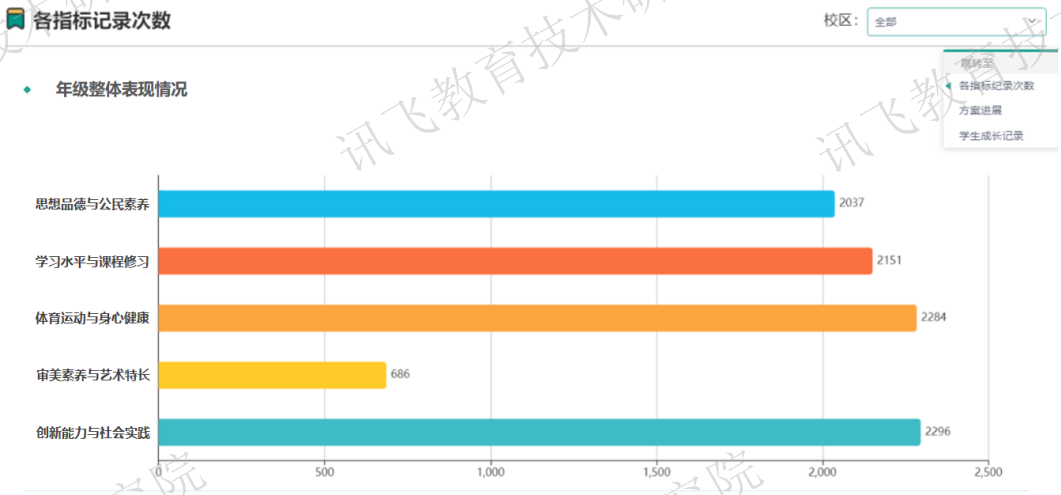


图 6-7 2020-2021 学年上学期高二年级的各指标记录情况

④自动生成与应用学生综合素质评价报告。系统整合经过采集、审核、分析的信息，自动生成学生的综合素质评价报告。学生、教师、家长都可以查看报告，全面了解学生的发展情况。首先，报告展示了学生的整体表现情况，包括学生的基本信息，班主任、科任教师对学生的评语和学生的自我评价，以及学生在各项评价指标中的上传情况。其次，报告展示了学生课内、课外在各个指标方面的具体发展情况，记录了重要过程性表现和主要成果。

讯飞教育技术研究院

第 7 章 行业现状及发展趋势

智能技术赋能教育评价最终要落地到具体的产品与服务中，即以技术为支撑，以产品与服务为载体，驱动“传统教育评价”向“智能化教育评价”转变。在政府宏观政策的引导下，高校、专业研究机构、企业等都对“智能化教育评价”领域进行了积极研究与探索，催生了多样化、智能化的教育评价工具、产品、服务等，形成了相关行业蓬勃发展的态势。

7.1 行业现状分析

通过对国内开展智能化教育评价研究与实践的政府机构、高校、专业研究机构、企业、用户进行深入调研，以教育部教育移动互联网应用程序备案管理平台上备案的企业与第 76、77、78、79 届中国教育装备展的参展企业及其发布的产品为总样本，从中剔除已注销、信息不全以及与智能化教育评价不相关的无效样本后，最终筛选出 525 家企业、889 个产品作为智能化教育评价相关行业的有效样本，从多个维度分析企业竞争格局与产品布局情况。

7.1.1 “政产学研用”协同创新的产业生态体系初步形成

智能化教育评价是一项技术性、专业性非常强的复杂工作，涉及政策、理论、技术、产品等各方面的内容，需要政府、高校、专业研究机构、企业、用户等多主体协同攻关、形成合力。目前，我国智能化教育评价产业已初步形成“政产学研用”协同创新的产业生态体系：一是政府在产业生态体系中扮演着引导者、支持者、组织者的角色，主要通过制定相关政策、提供相关财政保障、设立相关监管机构、设置相关课题等方面来推动产业协同创新与发展；二是高校在产业生态体系中承担着知识再生产与智力再生产的双重任务，主要通过开展科学研究与人才培养工作来参与到产业生态中；三是专业研究机构与高校共同扮演了科技创新成果的研发者与生产者的角色，在产业生态体系中主要提供专业化的教育评价服务与科学研究支持，推动科研成果转化；四是企业作为技术创新的主体，在产业生态体系中主要通过相关技术与产品的研发与应用，将科研成果转化为现实生产力。五是区域、学校、教师、学生、家长等用户群体基于自身的使用体验，向企业反馈相关改进意见，推动相关技术、产品、服务的创新升级。

讯飞教育技术研究院

7.1.2 企业多方入局，多元化市场竞争格局逐渐显现

（1）企业类型分析

目前整个市场中教育评价相关企业以教育服务提供商为主。同时评价服务提供商、互联网服务提供商、人工智能技术提供商、大数据服务提供商等多类商家均抓住了智能技术赋能教育评价的战略机遇，纷纷入局该领域布局相关业务，整个市场逐渐呈现多元化的竞争格局。具体如图 7-1 所示。

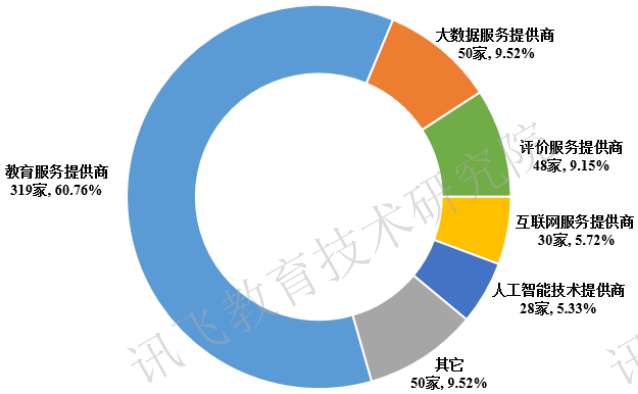


图 7-1 智能化教育评价相关企业类型分布情况

注：“其它”表示除教育服务提供商、评价服务提供商、互联网服务提供商、人工智能技术提供商、大数据服务提供商以外的其它企业。

（2）企业区域分布分析

教育评价相关企业的区域分布呈现出梯队化、集聚化特点。北京、广东与上海是智能化教育评价相关企业的主要集聚地。具体如图 7-2 所示。

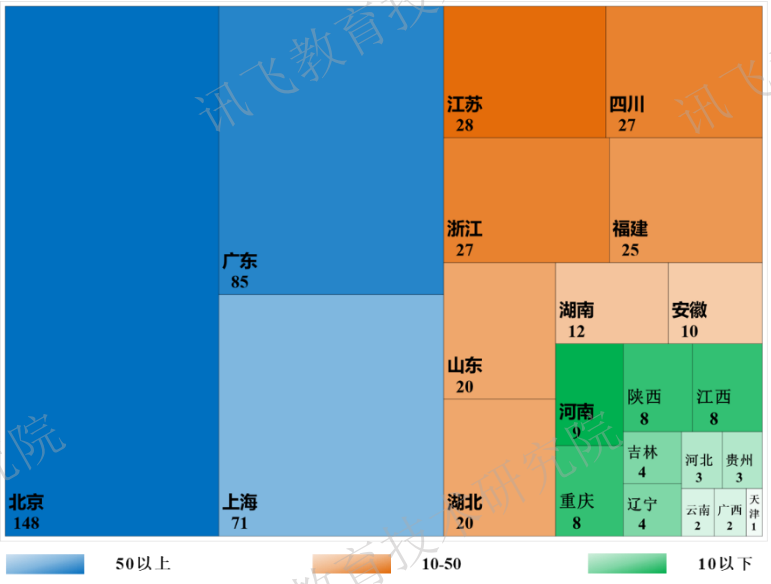


图 7-2 智能化教育评价相关企业的区域分布情况

注：省、直辖市上标记的颜色越深则代表该地区汇集的企业数量越多。

（3）企业规模呈现“金字塔式”分布，小微企业是行业主力军

智能化教育评价相关企业的企业规模呈现“金字塔式”分布，大型企业处于顶部且数量较少，小微企业是智能化教育评价相关行业的主力军。具体如图 7-3 所示。

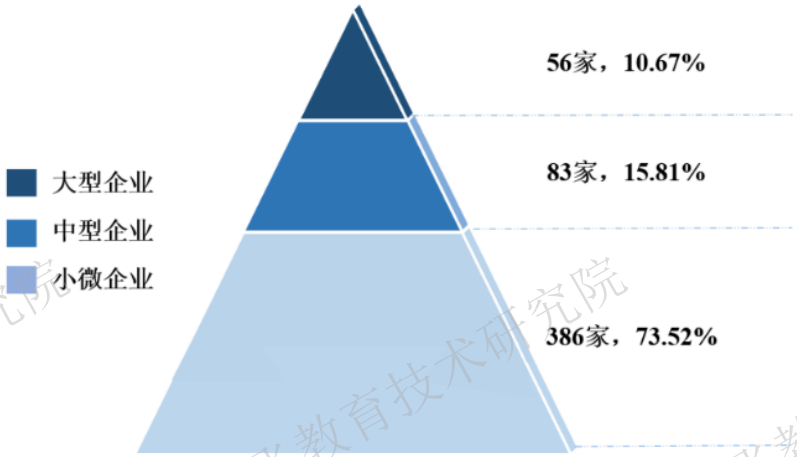


图 7-3 智能化教育评价相关企业规模分布情况

7.1.3 产品形态不断丰富，技术赋能教育评价初显成效

将 889 个有效产品样本按产品类型进行统计、聚类，梳理出了智能化教育评价相关产品的主要产品形态，绘制出相关产品形态词频图，如图 7-4 所示。



图 7-4 智能化教育评价相关产品形态词频图

（1）评价主体维度

目前智能化教育评价相关产品主要集中在评价主体为学校、教师和学生的校内评价，同时教育评价的主体呈现出多元化、多样化的特点。具体如图 7-5 所示。

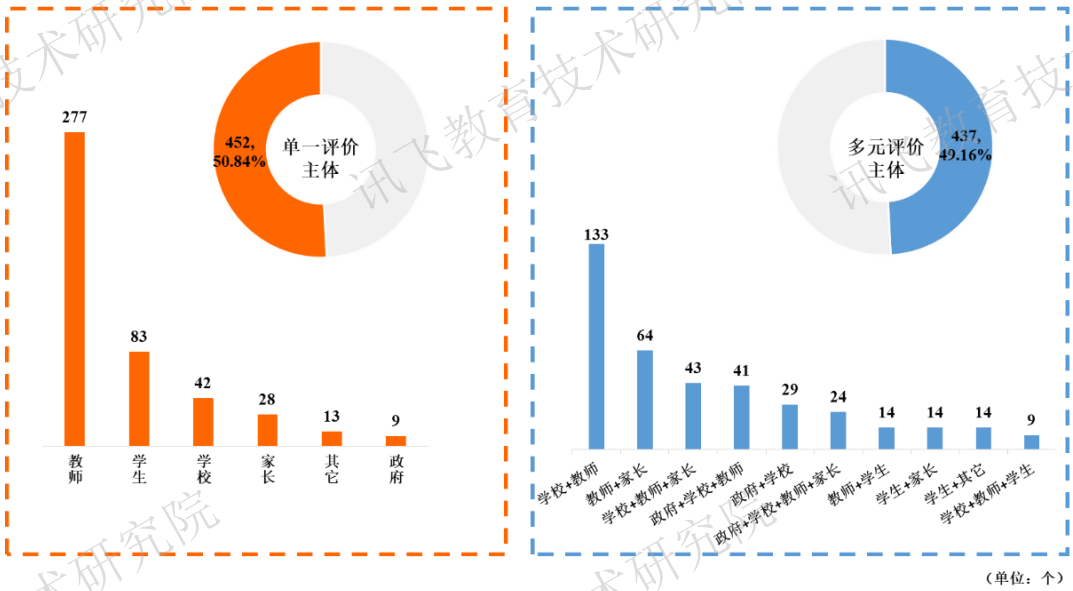


图 7-5 智能化教育评价相关产品中评价主体的详细分布情况

注：“多元评价主体”中仅列举了数量排名前十的组合。

（2）评价对象与评价内容维度

目前市场上的智能化教育评价相关产品以学生为最主要的评价对象。其中，大部分产品的评价内容较为单一，学业水平仍是学生评价最主要的评价内容。同时，学生素质评价的内容逐渐丰富，逐渐涵盖思想品德、身心健康、艺术素养、劳动与社会实践、职业兴趣等多个方面。具体如图 7-6、7-7 所示。

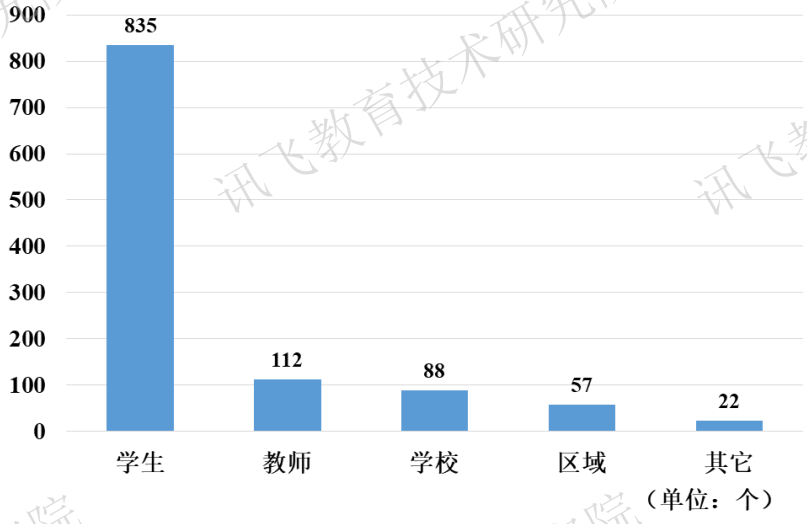


图 7-6 智能化教育评价相关产品中各类评价对象的数量分布

注：图中各项数据有交叉，原因是部分教育评价产品涉及多个评价对象。

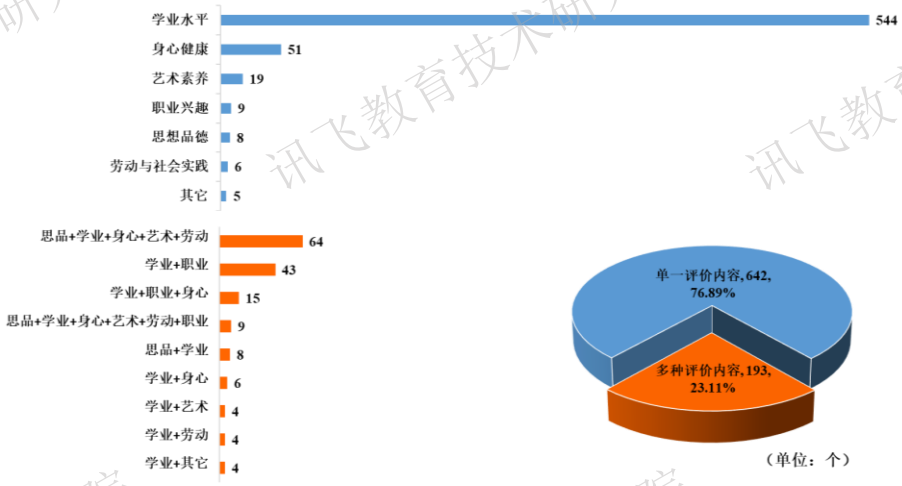


图 7-7 以学生为评价对象的智能化教育评价相关产品的评价内容详细分布情况

注：“多种评价内容”中仅列举了数量排名前十的组合。

（3）评价方式维度

智能化教育评价相关产品中，结果评价与过程评价是最常用的评价方式；增值评价方式的实践应用相对较少，并且往往与其它几类评价方式相结合应用；各类评价方式的结合应用开始在市场中探索实践。具体如图 7-8 所示。

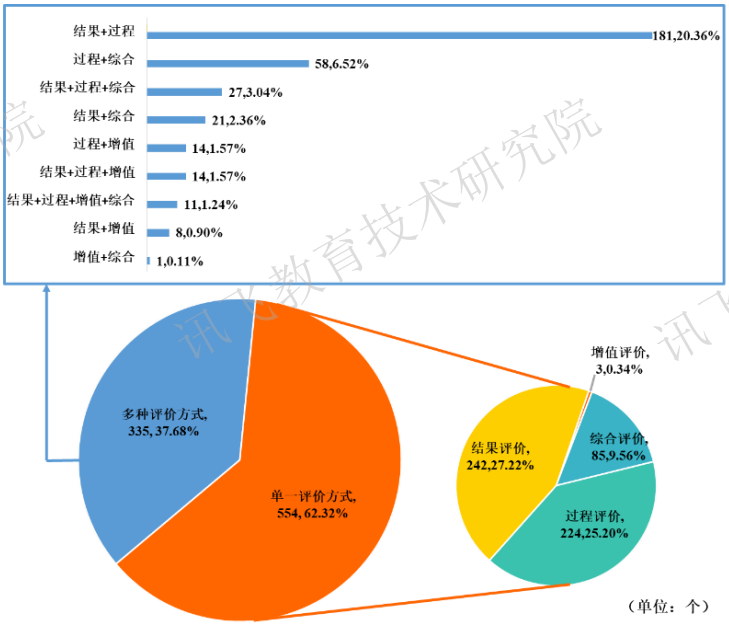


图 7-8 智能化教育评价相关产品采用的评价方式分布情况

（4）数据采集手段维度

智能化教育评价相关产品中使用单类采集手段的居多，其中平台数据采集是目前最常用的数据采集手段。同时，综合应用多类数据采集手段的智能化教育评

价相关产品开始崭露头角，尤以平台数据采集与语音数据采集相结合的产品最多。具体如图 7-9 所示。

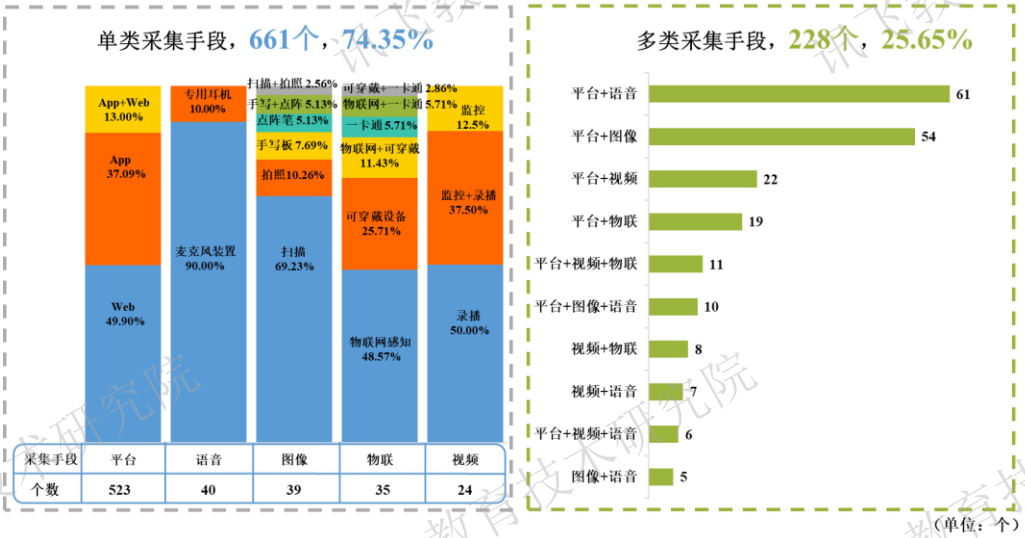


图 7-9 教育评价产品涉及的数据采集手段的分布情况

注：“多类采集手段”中仅列举了数量排名前十的组合。

（5）数据分析技术维度

目前可视化分析是智能化教育评价相关产品采用的最主要的数据分析技术，市场上的智能化教育评价相关产品在数据分析环节采用预测性分析与语义引擎技术的相对较少。具体如图 7-10 所示。

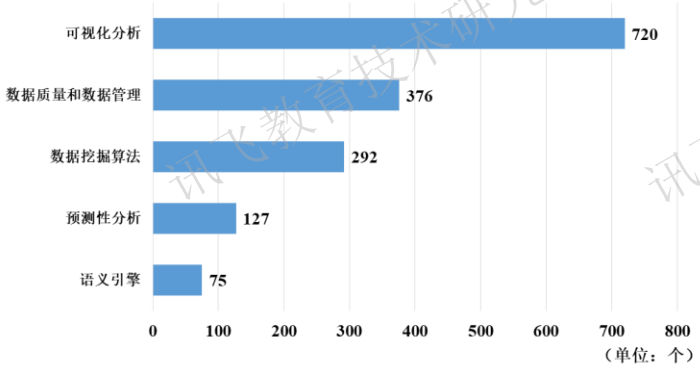


图 7-10 智能化教育评价相关产品涉及的数据分析技术分布情况

注：图中各项数据有交叉，原因是部分教育评价产品涉及多个数据分析方法。

7.2 技术应用趋势

7.2.1 区块链支撑更开放与更有公信力的教育评价

区块链技术在教育评价中的探索运用，或可解决评价工作中存在的数据孤岛、

信任危机等问题。由于其具有去中心化、不可篡改等特点，可在区块链成绩单、综合素质档案、评价结果永久记录与存储、电子证书管理、学分认证等教育评价场景下探索运用[①]。

7.2.2 边缘计算开启泛连接与低时延的教育评价

实时反馈在教育评价中具有重要意义，边缘计算可助力实时评价的实现。边缘计算在靠近用户或数据源的位置提供网络、计算、存储服务，不仅能够实现流量的本地化处理，降低对远端数据中心的流量冲击，而且能够提供低时延和高稳定的应用运行环境，解决评价反馈滞后的问题，满足实时计算反馈的需求。

7.2.3 数字孪生探索虚实融合的教育评价

数字孪生技术因其虚实共生、数据驱动等特点引起教育工作者们的关注。在教育领域中，通过构建学习者孪生体，通过虚实结合的方式，动态呈现学习者的相关学习数据，助力学习者状态评估、学习者自我认知与改进、教师个性化干预与指导、学习过程全方位综合评价等教育评价活动[②]。

7.2.4 情感计算体现更具人文关怀的教育评价

情感计算是当前人工智能的重要研究方向之一。在教学中，情感计算可对学习者学习过程中的情感状态进行检测与识别，助力学习状态评估、学习过程干预、教学策略调整，为学习者提供个性化的学习服务[③]。

7.2.5 脑科学推动见微知著的教育评价

脑科学与教育相结合，或将变革传统教育评价方式。在教育评价中，通过检测学习者在学习过程中脑电信号变化，可助力学生心理健康状态监测与评估、学习障碍检测与评估。当前“因脑施教”的理念正渐渐被教育行业从业者所接受[④]，但仍需警惕与慎重的使用学生脑信号数据。

7.2.6 类脑计算提供更快速更高效的教育评价

类脑计算是脑科学的重要研究领域之一，可为实现更加快速、高效的教育评价提供可能。在教学中，可通过类脑传感器精准采集学生数据，脉冲神经网络为教育评价提供强大算力支撑，对于收集到的教学数据进行快速的计算与分析，将

① 沈忠华. 新技术视域下的教育大数据与教育评估新探——兼论区块链技术对在线教育评估的影响[J].远程教育杂志,2017,35(03):31-39.
② 郑浩,王娟,王书瑶,顾雯.认知数字孪生体教育应用:内涵、困境与对策[J].现代远距离教育,2021(01):13-23.
③ 杨金朋,薛耀锋,李佳璇,等. 基于人脸表情识别的在线学习情感计算研究[J].中国教育技术装备,2017(18):35-36.
④ 科大讯飞 AI 研究院脑科学组.脑科学产业研究报告[R].2020.

计算结果与反馈建议实时地提供给学生，实现更快速更高效的教育评价。

以上分析表明，智能技术的发展为智能化教育评价实现提供了可行的解决路径，但智能时代的教育评价应始终谨记与围绕教育的本质来实施。促进人的全面发展是人类一切教育的出发点和归宿[①]。要求智能技术相关者能够树立信息社会的生态伦理观、建立教育评价中技术实践约束与数据应用伦理的法律法规、尊重教育活动中师生的自主性以及其自我决定权，让智能教育评价应用真正成为有教育意义的、合乎道德的评价。

7.3 行业发展趋势

7.3.1 多方合作攻关需求强烈，“政产学研用”一体化成为产业创新发展必然趋势

一是多元力量合作一体化，即充分发挥政府的宏观调控功能，有效整合政府、高校、专业机构、企业与用户等创新主体在技术创新、人才培养、科研成果转化、新产品研发等方面的资源优势，五位一体协同创新；二是技术创新与产业创新一体化，强化技术与产业对接，以先进技术、适用技术引导市场需求，以市场调节机制推动产学研用相关创新主体的合作向纵深发展，提高创新效率；三是技术链和产业链一体化延长，即同步延长技术链与产业链，以政学研产用各方互促共进机制和创新继起机制促进产业集群向高端发展。

7.3.2 业务范围逐渐聚焦，专业性高的教育评价企业受到市场青睐

一方面，随着智能化教育评价市场的进一步发展与规范，相关企业的业务范围将进一步聚焦，评价业务在企业整体业务布局中的比重将进一步上升，除部分行业巨头之外，其它企业的评价业务将从大而全转向小而精，从而能够在各自细分赛道内深耕。另一方面，随着人工智能、大数据、区块链、云计算等智能技术的飞速发展，教育评价服务对先进技术的依赖程度将会更高，拥有核心技术的企业将在市场中占据较大优势。

7.3.3 教育评价产品不断创新发展，智能化、精准化评价服务成为未来风向

虽然目前智能技术已经在智慧课堂、自适应学习、智能阅卷、智能口语评测、学生综合素质评价、区域教育质量监测等教育评价相关产品中实践应用，但大部分产品在技术赋能方面仍较为初级，普遍存在智能化程度不足、评价结果不够精准等问题。利用智能技术的突破，创新智能化的教育评价工具、系统等，成为推进教育评价精准智慧运行的有效解决路径。智能化、精准化的教育评价服务将成

① 冯锐，孙佳晶，孙发勤. 人工智能在教育应用中的伦理风险与理性抉择[J]. 远程教育杂志，2020,38(3):47-54.

为未来风向。

7.3.4 技术壁垒将被打破，多种数据采集与分析方法的集成应用将成为升级方向

随着大数据的深入驱动，教学评价的途径将不再是传统意义上通过经验获取，而是依靠大量数据的归纳与整理[①]。教育评价业务将从传统的经验驱动转为数据驱动，打破信息孤岛与技术壁垒，实现多系统融合与数据集成，集成应用多种数据采集与分析方法实现多模态、全内容、全过程、全时空、全场景数据的自动采集与评分，将成为智能化教育评价相关产品的未来升级方向。

7.3.5 学生素质评价的内容逐渐丰富，素质评价市场或将迎来爆发期

学生综合素质测评成为教育评价的重要关注点，除学业水平外，对学生各方面素质、素养的评价越来越受到各方关注，评价的内容也不断丰富，包括思想品德、身心健康、艺术素养、劳动与社会实践、职业兴趣、科学创造素养等。同时，受“双减”政策影响，当前各类教培机构纷纷聚焦素质教育，借助技术的赋能进行业务转型，这也必将推动学生素质评价市场的发展。在政策、市场、技术的共同驱动下，学生素质评价市场或将迎来爆发期。

① 吴晓威,陈旭远. “大数据”理念的教育应用与中国教育改革——从数据分类到证据转化的机遇识别[J]. 内蒙古社会科学(汉文版),2014,35(06):168-171.

工作组与致谢名单

专家委员会

主任　陈丽（北京师范大学教授、博导）

副主任　吴晓如（科大讯飞股份有限公司总裁）

高校专家　（按姓氏笔画排序）

方海光（首都师范大学教育技术系数字化学习实验室主任、教授、博导）
孙波（北京师范大学未来教育学院副院长、教授、博导）
杨现民（江苏师范大学智慧教育学院院长、教授、博导）
吴砥（华中师范大学国家数字化学习工程技术研究中心副主任、教授、博导）
汪琼（北京大学数字化学习研究中心主任、教授、博导）
张海（东北师范大学传媒科学学院副院长、教授、博导）
胡小勇（华南师范大学教育信息技术学院副院长、教授、博导）
钟晓流（清华大学信息技术中心副总工程师、中国教育技术协会副会长）
贾积有（北京大学教育学院教育技术系主任、教授、博导）
顾小清（华东师范大学教育信息技术学系主任、教授、博导）
郭绍青（西北师范大学教育技术学院院长、教授、博导）
黎加厚（上海师范大学教育技术学系教授）

行业专家　聂小林、王　政、汪张龙、周佳峰、王士进

编写组

主编　刘邦奇（讯飞教育技术研究院院长、教授）

副主编　王亚飞、袁婷婷、张金霞、刘俊生

编写者　（按编写章节排序）

刘邦奇、袁婷婷、代怿鑫、胡　健、喻彦琨、朱广袤
刘碧莹、刘俊生、陈　乐、董晶晶、凌婧婧、王冰洁
崔东泽、张金霞、胡婷婷、许佳慧、许星伟、刘　悦

参与者　卓　晗、邬诗韵、唐佳慧、冯子芳、纪玉超、马丹丹、李　岭

致谢

实践案例评审专家　许　林、刘邦奇、鲍潜巢、袁婷婷、刘俊生

案例编写技术支持者　鲍潜巢、李　鑫、朱　香、郭　鲲、陈玉珏
程泽宇、杜　浩、李　晨等

实践案例提供者　王　勇、段元丽、李新义、张　生、郑勤华
罗　强、王　蕾、何屹松、李晓娟、袁玉伟
谭中胜、吴英明、张家成、徐亚倩、王　伟
白丽波、匡昌林、居　艳等