



教育评价中的信息技术应用： 赋能、挑战与对策

□ 张志祯 齐文鑫

【摘要】

教育评价事关教育事业发展方向。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》对新时代教育评价改革做出了全面部署，指出应充分利用信息技术，创新评价工具。作为教育活动的“指挥棒”，教育评价中信息技术应用不能仅处于自发水平，需要达到自觉理性的水平。通过分析教育评价一般过程与信息技术特点可以看到，信息技术赋能教育评价主要体现在“增强”评价工具、“创新”评价工具、优化评价管理、提升评价质量和拓展评价结果等方面。同时，在教育评价中利用信息技术也存在挑战与风险，主要是评价主体应用能力不足、信息技术自身不稳定不可靠、易流于“数据驱动”、易形成“算法霸权”及易滥用过程性数据等。为此，可采取的对策包括：提升评价主体教育评价信息技术应用能力、研制教育评价数据规范与建立公开测试数据集、建立评价迭代更新机制、发展评价对象自评能力、提高全社会教育评价素养等。

【关键词】 教育评价；信息技术应用；评价过程；评价质量；数据驱动；算法霸权；再评价；信息化

【中图分类号】 G40-058.1

【文献标识码】 A

【文章编号】 1009-458 x (2021)3-0001-11

DOI:10.13541/j.cnki.chinade.2021.03.001

2020年10月，中共中央、国务院印发了《深化新时代教育评价改革总体方案》（以下简称“《方案》”），对新时代教育评价改革做出了全面部署。《方案》打开了教育系统“黑箱”，明确了教育系统内、外的重点评价对象，即学校、教师与学生三类教育系统主体，党委与政府、选人用人社会机构两类与教育系统直接相关的社会子系统；明确了结果评价、过程评价、增值评价、综合评价、综合素质评价、德育评价、选拔评价等评价任务。

《方案》指出教育评价应充分利用信息技术，创新评价工具。教育评价活动类型多样，都是数据与知识“密集型”实践，高度依赖客观数据与对数据的分析处理，因此教育评价领域历来对信息技术的新进展十分敏感，往往新技术一经出现，很快就有相关探索其在教育评价中的运用，如柯林斯（2013, p.114）、芦咏莉等（2012, p.27）。

但是，这并不意味着不需要对信息技术在教育评价中的应用做系统探讨。理由如下：一是教育评价中信息技术应用存在不同水平，可以用新技术按照已

有方式行事，也可以是创造新的行事方式；二是正式教育评价多为自上而下的任务，学校、教师被“裹挟”到教育评价活动中，或许对信息技术应用有体验，但难以有系统理性的认识；三是教育评价是教育活动的“指挥棒”，对教育活动的广泛、深远、持久，信息技术应用不能只停留在自发水平，需要达到自觉理性的水平，只有明确信息技术的优势与问题，才能充分发挥信息技术的预期作用，避免滥用误用；四是信息技术在其他领域的广泛应用已产生了未预料到的社会后果，与“信息迷”的乐观声称背道而驰。例如，充满不同声音的互联网让更多人变得更加狭隘偏激（尼科尔斯，2019, p.128），解放劳动力的自动化改变了工作性质（卡尔，2015, p.78），互联网新闻推送让传统新闻业哀鸿遍野、独立思想日渐式微（福尔，2019, pp.156-157），掠夺式广告让穷人陷入更深的窘境（奥尼尔，2018, pp.72-73），MOOC不但并未颠覆传统高等机构，反而被逐渐吸收整合（Reich, 2020, p.43），等等。教育评价事关国家教育事业发展方向与个体命运，因此有必要对信息技术应用的可能

影响做系统分析。

一、作为特殊知识生产实践的教育评价

教育评价是对教育主体、活动或现象的价值评定，是一种特殊的“知识生产”实践。Gronlund提出“评价=测量（定量研究）或非测量（定性描述）+价值判断”，这抓住了评价活动的本质（陈玉琨，2019，p.13）。在实践中教育评价人员大量运用研究设计、测量工具、数据分析技术等教育研究方法，因此也常常将评价活动称为评价研究。教育评价与教育研究的区别不在于方法与方法论，而在于目的与追求（高尔等，2002，p.558）：评价的目的通常是改进或支持决策、评定绩效，而教育研究的目的则是增进对教育现象的理解。这导致研究结果与评价结果的概括程度不同：教育研究结果更概括，追求对“某类现象”（that）或者关系的普遍理解，希望产生的知识是超越情境的；评价结果则非常具体，明确指向“某个对象”（that one），希望产生的知识是特定情境中准确、可靠和有用的，即基于某对象的知识能够做出有效可信的价值判断。

正如生产工具只能作用于具体的生产过程，信息技术也只能在教育评价的微观活动过程中起作用。为了更深入地探讨教育评价中如何充分利用信息技术，需要明确教育评价活动的微观过程。基于陈玉琨（2019，p.38）、吴钢（2018，p.46）、希尔伦斯等（2017，p.21）和高尔等（2002，pp.559-566）的教育评价过程模式，可概括出教育评价的一般过程（见图1）。

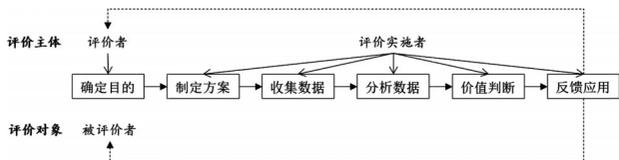


图1 教育评价的一般过程

教育评价的评价主体包括评价者和评价实施者，前者是评价的发起者，如政府、用人单位等；后者是具有实施评价能力的专业人员或机构。评价对象是学校、教师、学生等教育系统主体。教育评价的一般过程为：①确定目的，即评价者确定评价目的、被评价者、评价实施者，组织安排评价活动资源；②制定方案，即确定评价的具体目标、指标体系与标准、评价工具（数据收集和分析方法工具）、数据来源、评价

结果使用方式与呈现形式、评价过程安排等，大规模评价还需要明确分工合作机制、数据标准、数据管理平台等；③收集数据，包括定量数据和定性数据、专门收集的数据和日常运作产生的数据（伴随式收集数据）；④分析数据，即整合不同来源的数据，开展定量与定性分析；⑤价值判断，即汇总分析结果，对照评价目的、指标体系和标准，做出价值判断，形成评价结果；⑥反馈应用，即向评价者、被评价者和公众反馈评价结果，利用评价结果进行选拔鉴定、问责或者改进教学。总之，在评价中“收集数据”“分析数据”是为了获得有关评价对象的知识，是情境性很强的本地知识；价值判断以前面的步骤获得的知识为基础；评价结果的“反馈应用”特别重要，且具有直接的社会后果，例如高校专业的级别、考生是否被录取、应聘者是否被录用等。

有研究者（吴钢，2018，p.179；陶西平，2012，pp.164-166）强调应对评价进行评价，即“再评价”。鉴于“再评价”以评价活动为评价对象，是新的、独立的“元”评价活动，其过程也符合图1中的一般过程，故未在图中体现。

二、信息技术赋能教育评价

信息技术已经成为日常工作、生活、学习中的普遍技术。为了使分析基础更加牢靠，仍有必要进一步明确对于信息技术的认识。

（一）对信息技术特点的再认识

这里的信息技术特点指现代信息技术，即以电子数字计算机系统为核心的信息获取、处理、存储、传输、呈现与管理的系列技术，是一个多层次多样化的技术体系。图2是与教育评价应用直接相关的信息技术三层次结构与特点示意图。

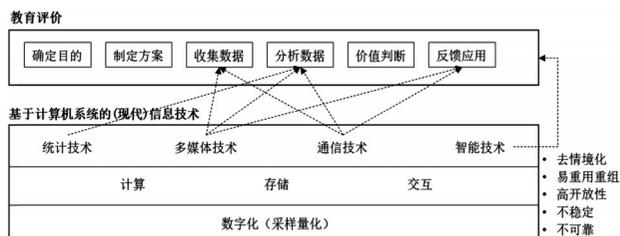


图2 与教育评价相关的信息技术及其特点

在最“底层”，数字化是数据与信息进入信息技术营造的“比特世界”的前提条件。数字化后信息就



脱离了原子的束缚,可以自由穿梭于数字世界(尼葛洛庞帝,1997,pp.22-23)。自由是有代价的:经过采样量化,信息通常会失去“上下文(context)”,导致其意义难以自明,完整性难以保证,易被曲解、篡改(布朗等,2003,p.116)。在早期,数字化需要人来完成,即人工用纸带打孔、键盘等方式为计算机输入数据;声卡、数码相机、二维码扫描器等多媒体设备,能够自动实现比特层次的数字化;“传感器+人工智能技术”则实现了具有人类感知特点的数字化,如人脸识别、拍照翻译、语言识别等。

中间层是通用的计算、存储与交互基本服务。计算不仅是数值计算,还包括文本查找、逻辑判断等非数值计算,高速、准确是计算机计算的突出特点。计算机存储的基本特点是容量大、读写速度快,且数据与处理程序的存储机制相同。交互指通过键盘、鼠标、显示器等输入输出设备实现与计算机的“人机交互”。

最上层是专门技术,和教育评价密切相关的是统计技术、多媒体技术、通信技术和智能技术。“统计技术”指计算机实现定量与定性数据统计分析的软硬件技术,集中体现在R、SPSS、Excel等软件包中。统计技术是信息技术用于评价的切入点,极大提高了数据分析的效率与准确性,在“分析数据”中普遍应用。“多媒体技术”指语音、图像等视听信息的输入、处理与输出技术,是真实社会场景与活动的捕获、分析与呈现的有效手段。多媒体技术包括虚拟现实技术,通过它可以创建沉浸式体验(张志祯,2016),在“收集数据”“分析数据”“反馈应用”中应用广泛。“通信技术”包括实现人际通信的互联网服务与实现“物体间通信”的传感器与物联网技术。前者使大规模的教育评价合作成为可能,也为数据管理、传递提供了新的平台;后者为新的测评工具研发开辟了道路。通信技术在“收集数据”“分析数据”“反馈应用”中应用广泛。“智能技术”指与实现人类个体特有的感知、推理与行动的“智能行为”相关的计算机技术,如图像识别、自然语言处理、知识推理等(周志华,2018;Russell,等,2013),大致上可以分为自上而下赋予计算机知识的知识工程技术(如专家系统)和自下而上从大量数据中自动发现“模式”以做出预

测的机器学习(如图像识别深度神经网络)。有经济学家认为智能技术的基本特点是使预测变得廉价(阿格拉沃尔,等,2018,p.19)。智能技术对于教育评价的影响正在逐步显现,这一影响将是全过程、全方位的。

总体上看,上述多层次的信息技术表现出如下特点^①:①去情境化,这是“采样量化”数字化过程的结果,这一处理有得有失,具体请参见前文“数字化”相关分析;②易重用重组,数字世界的资料与软件程序功能多姿多彩,但由于其底层都是比特,因此易于重新利用,借助超文本超链接技术,不同来源的信息很容易就能并置、整合在一起(重组),方便了跨越时空、情境的比较分析;③高开放性,主要表现在可进一步做编码、加工、转换和通过网络进行不受时空限制的访问;④不稳定(instability),指信息技术软硬件总是处于发展过程中,硬件设备的使用寿命不长,软件则受制于操作系统的频繁更新,一款软件可能几年之后就难以找到适合的运行环境,这导致信息技术应用需要持续的经费与人力投入;⑤不可靠(unreliability),指计算机软硬件可能包含错误,纠正错误的“补丁”常常会引入新的错误,若功能依赖互联网在线服务,则互联网连接、网络速度与服务器的不可控与不确定,进一步增加了信息技术的不可靠。

(二) 信息技术“增强”教育评价工具

教育评价工具通常指对评价对象进行测定时所采取的方式和手段,如论文考试、标准化测验、行动观察记录、作品分析等(顾明远,1998,p.767)。由于数据分析处理与数据收集方法密切关联,本文扩展“教育评价工具”的外延,将数据分析处理方式与手段也包括进来。

本部分的“增强”指穿新鞋走“老路”,即用信息技术完成了之前需人工完成的工作。信息技术增强使教育评价工作更快速准确、更规范、更大规模。

“更快速准确”突出表现在数据收集与分析环节,尤其是定量数据的统计分析上。以考试评分与成绩统计为例,借助数据处理软件,教师可以实时对成百上千学生的成绩做交互式统计分析、可视化呈现、多角度对比,快捷、准确。自动阅卷技术提高了阅卷

^① 严格地说,这些特点有的是关于信息技术的(如不稳定、不可靠),有的是关于信息技术所产生、处理的信息或数据的(如去情境化),为行文简便起见在表述时不做区分。

速度和准确率,极大节省了人力。在线调查网站使采用问卷、量表、测验等方式收集数据变得更加高效。对于教育评价而言,数据收集与分析的“快速准确”可缩短评价周期,降低评价成本,使反馈应用更及时,有利于提高实践改进所需的时效性,更好地落实评价促进发展的功能。

“更规范”指在利用计算机收集与分析数据时,能够标准化数据测量、存储、传输与分析方法,严格控制数据收集流程与时间,使数据处理方法标准化、规范化。例如,在2018年宁夏普通高等学校招生体育专业术科测试中采用信息技术测量仪器与设备,测试过程更加科学、规范、准确(贾海生,2019)。

“更大规模”指即使采用同样的评价工具,由于信息技术增强了数据收集能力和数据处理能力,使得可以在更大规模开展评价。如利用计算机网络汇总分析数据,5个人的小团队就可以在有限的时间内高质量地完成涉及1,816个高职专业的全国项目的绩效评价与报告撰写(张志祯,等,2015,p.473)。

信息技术工具将部分数据收集、统计工作自动化,大大简化了某些评价活动,使这些评价活动能够整合到日常教学中。例如,用Excel做成绩分析,用“问卷星”收集数据,使用课堂教学互动系统、QQ群的投票功能等,这些都有助于教师在教学中开展形成性与诊断性教学评价。也就是说,信息技术的“增强”作用,有时体现在降低评价活动技术门槛与时间成本的“简单化”上,这对于评价活动的“日常化”至关重要,对于日常教育实践非常有价值。

(三) 信息技术“创新”教育评价工具

信息技术不仅能使已有工具与方法更加高效,还能够创造出新的工具与方法,实现穿新鞋走“新路”。信息技术的创新应用使教育评价有了新数据、新工具、新模式。

“新数据”,信息技术大大拓宽了评价数据来源。例如,有关真实社会场景与活动的多媒体信息(如视频)、活动过程中伴随式采集的实时生理与心理数据(如脑电波、情绪、注意力等)、在线学习行为数据、机构运作过程中自然产生的数据记录等。新型数据已运用于各种教育评价研究与实践,例如课堂教学视频已在美国NBPTS专家教师的认证中长期使用(袁锐镔,等,2004)。

“新工具”,信息技术催生了很多新型“评价工

具”,典型的如适应性测试、游戏测评、虚拟现实仿真场景测试以及合作问题解决测试等。适应性测试能够根据考生的表现,动态选择下一道题目,“每道题都被看作是一项独立的迷你测试……(考生)回答会不断改变我们对其真实分值的推测”(Thorndike,等,2018,pp.165-166)。基于项目反应理论的适应性测试能够缩短测试时间,提高测试信度和效度,已大规模用于GRE考试,也用于Knewton等大型适应性学习系统(Reich,2020,p.53)。适应性测试的实现依赖高强度的实时计算,只能在计算机上实现。电子游戏测评提高了测试任务真实性,能够动态检测任务监控等复杂能力,是更为真实的表现性测试(孙鑫,等,2018)。PISA测验采用的合作问题解决也在计算机、互联网环境中开展(檀慧玲,等,2018)。此外,《方案》强调的综合素质评价,需要处理大量的非结构化数据,其分析与综合评价有赖于新的智能分析技术。

“新模式”,简单而言,对于信息化比较成熟的机构,由于日常运作能够产生大量的过程性数据,为更好地利用这些数据,提高评价效果与效益,需转变评价思路,探索采用“伴随式数据收集—确定目的一制定方案—数据选择与分析—价值判断—反馈应用”这一新的过程模式。

(四) 信息技术优化教育评价管理

在计算机时代开启之际,维纳就预言了计算机的生产、社会控制潜力(维纳,1978,pp.138-139)。大规模教育评价是涉及面很广的复杂活动,需要有效管理才能有序可控,保证质量。对于这类评价而言,信息技术已经不是可选项,而是不可或缺的。

信息技术增强了大规模、高利害教育评价的管理与控制。我国的高考是个典型例子。高考的考务管理包括报名、投档、录取、身份识别等,早已实现全程信息化,信息技术为高考保驾护航,推动了高考现代化改革(游忠惠,2013)。

信息技术使大型教育评价项目合作开发与实施成为可能。例如,经合组织(OECD)的PISA测试、国际教育成就评价协会(IEA)的TIMSS研究、我国的基础教育质量监测、美国的国家教育进步评价(NEAP)等都依托互联网实现大范围合作式评价开发、施测、过程与数据管理。

信息技术使对于教育机构的增值评价具备可操作性。增值评价涉及复杂的数据收集与统计,尤其是分



层次、分块以及历史数据的对比分析(辛涛, 2020), 若没有信息技术数据管理与统计分析的支撑, 增值评价的时间与人力成本会极其高昂。

信息技术还开辟了课程教学资源有效性评价的新方式。基于互联网的智能教学系统可以作为被试分配、资源传递与展示、学习过程开展、评价数据收集、评价结果报告的新型教学实验平台, 实现了对大规模、高生态效度的教学实验的全过程自动化管理(Heffernan, et al., 2014)。

(五) 信息技术提升教育评价质量

陶西平(2012, p. 166)认为评价具有双重性, 好的评价促使教学工作变得更好, 不好的评价可能使教学工作的方向更偏。这个判断也适用于其他类型的教育评价。《方案》将“充分利用信息技术, 提高教育评价的科学性、客观性与专业性”作为改革的原则之一, 明确了教育评价的质量标准、信息技术应用的目标和价值。

对于教育评价而言, “科学性”主要表现为评价的高信度、高效度^①(吴钢, 2016, p. 180); “客观性”是信度的一种, 即评分者一致性信度(梅耶, 2016, pp. 96-97); “专业性”主要表现在评价实践的伦理性、高效性和规范性, 即高质量的教育评价应该是高信度、高效度、合乎伦理、高效、规范。对于运用信息技术为何能够提高评价质量, “高效、规范”前文已多有涉及, 含义也较为明确, 限于篇幅, 不再讨论。下面分析信息技术对于信度、效度、伦理性的促进作用。

“信度”指教育评价指标和工具的准确性和精确度, 表示通过某一特定测量程序得出的分值的一致性和可复制性的程度(Thorndike, 等, 2018, p. 130)。在评价过程中产生误差、影响信度的因素很多, 如评价对象与评价者状态、评价工具、评价实施环境、评价过程等(吴钢, 2015, p. 195)。信息技术通过测试过程与任务呈现的标准化、数据统计分析的准确性以及利用项目反应理论模型动态计算等方式提高评价信度。在互联网在线数据收集中, 如在线问卷填写或者在线测验时, 计算机可以通过强制作答、自动呈现/

跳过题目、检测回答有效性等方式, 减少答题误差, 提高评价信度。此外, 配备特定的硬件后, 计算机还可以实时检测被试的注意力、情绪状态是否异常等, 从而进一步提高评价信度。

“效度”指教育评价在多大程度上测量了所要测量的。教育评价的效度主要取决于指标体系的合理性和所用测量工具(如测验)的效度。测验效度指在运用测验时对于测验分数的解读可以得到证据和理论支持的程度(Thorndike, 等, 2018, p. 172)。具有一定效度的评价结果才是有决策参考价值的。具有一定的信度是高效度的前提, 但高信度不能保证高效度。例如, 学生头围的测量信度很高, 但用其测量学生阅读能力则效度很低, 而用来决定买多大的帽子则效度很高。效度取决于解释与使用评价结果达到特定目的的程度。教育评价至少应具备内容效度、结构效度与效标关联效度。内容效度的鉴定主要采用逻辑分析的方法, 常用专家判断的方式(吴钢, 2015, p. 184), 信息技术的作用不大。结构效度的鉴定多采用因子分析法, 计算量很大, 需要借助计算机完成(吴钢, 2015, p. 184)。效标关联效度是看评价结果与外部效标(如: 在研发新型测试时, 其他成熟的等价测验; 在招聘中, 被录取人员在工作中的实际绩效可作为外部效标), 评价系统若能与其他系统对接, 关联相关数据, 对于未来考察评价的效标关联效度将是很有益的。影响评价度的因素和影响信度的因素很多是重合的(吴钢, 2015, pp. 187-188), 因此“信度”部分所提到的信息技术作用也有利于效度的提高。此外, 由于计算机多媒体技术, 尤其是虚拟现实技术能够构建交互式虚拟仿真环境, 且能够跟踪环境中问题解决步骤, 对于问题解决能力、认知与动作技能、认知策略的测评可能是效度更高的测评方法(孙鑫, 等, 2018)。

“伦理性”指教育评价不应歧视任何评价对象, 应公平公正, 重视隐私保护, 维护评价对象利益。对于教育评价而言这应该是第一位的。信息技术在这方面的应用空间很大。例如, 通过合理设计测试系统, 保证学生匿名及评分阅卷过程的身份隐藏; 借助视

^① 对于选拔性评价(如高考)而言, 评价结果必须能够用于排序、筛选被评价者, 这要求评价结果必须具有区分度, 但对于标准参照评价、项目评价等区分度不是必要的。鉴于本文关注的是一般意义的教育评价, 而且在高信度、高效度的前提下信息技术应用不会影响区分度, 因此这里不把区分度列为评价质量的指标, 也不讨论信息技术应用对于区分度的影响。

觉、听觉辅助技术,使特殊需要学生也能够参加测试;根据学生情况,生成文化公平的测试,等等。

(六) 信息技术拓展教育评价结果应用

信息技术的信息重组重用能力以及开放性,为评价过程与结果信息的重用提供了更多的可能,主要用于改进教学、改进评价、促进相关者实践反思。

改进教学指教师或者计算机根据对学生表现做出的形成性评价,调整学习内容或方法,实现教学个性化。智能辅导系统(如Cognitive Tutor)是用“算法”调整教学的典型,其调整教学所依据的学生模型就是在监控学生问题解决的过程中实时评价数据的汇总(Anderson, et al., 1995)。对于学生的在线测试,系统能够实现自动统计与可视化呈现,帮助教师客观、全面地把握学生学习情况,辅助教师的教学决策,比如高中物理精准教学(祝令健,等,2020)和ASSISTments系统在教学中的使用方式(Heffernan, et al., 2014)。

改进评价指利用评价过程与结果的信息,对评价全过程进行“再评价”,以提升评价质量。若能在评价活动中注意利用信息技术收集整理过程性数据,就可以完整、真实地再现评价过程,使再评价过程“基于真实数据”,提高客观性与准确性。

促进相关者实践反思指信息技术为评价结果提供了新的传播展示途径。互联网与移动互联网是评价结果传播的有效平台。例如,视频网站、微信公众号、微信群等,不但可以把评价结果精准推送给特定人群,还可以提供阅读量、转发量等统计数据,而且具有互动功能,有利于基于评价结果形成实践改进共同体,更好地发挥评价的实践改进作用。

三、教育评价中利用信息技术的挑战与风险

教育评价是价值导向的活动,信息技术应该只在工具层面起作用,但是工具会改变主体的实践方式,会潜移默化地改变主体的认知方式与主体间的相互作用方式。对组织机构来说,信息系统有时像打入机构内部的“木马”,可能会打破组织的边界与平衡。工

具性信息技术应用也会产生社会后果。鉴于教育评价影响的滞后性和自我强化性^①,必须警惕信息技术应用带来的挑战与可能风险。

(一) 评价主体评价信息应用能力不足,影响信息技术应用效果

李政涛(2020)认为《方案》的落实关键在于“落实主体”,即教育局局长、校长和教师对方案具有足够的“理解能力、运用能力,以及运用之后的反思能力与重建能力”,落实过程充满挑战。

在教育评价中有效运用信息技术,对评价主体提出了更高要求。在认识层面,需要对教育评价改革理念和运作方式有深刻理解,需要明确信息技术的微观作用方式,中观层面上对于组织行为的影响,以及对组织和个体信息技术条件、能力的要求;能够开阔视野,了解社会系统中可以凭借的信息化力量;能够充分估计到信息技术应用的预期作用和副作用。在实践层面,能够设计开放灵活的机制,整合资源,充分利用社会专业力量补齐能力短板,创造条件,保障评价顺利实施,确保善用评价结果,控制总体成本与潜在风险。这些要求与前面的挑战相叠加,一方面凸显信息技术赋能评价的潜力,另一方面让评价过程更加复杂动态,若支持与管理跟不上、方法设计不合理,可能反而会增加评价成本,降低评价质量。

(二) 信息技术自身不稳定、不可靠,增加评价成本

对于评价者而言,信息技术的不稳定,尤其是硬件、平台类软件技术(如操作系统、浏览器、Flash^②、Java、Python、办公套件等)的更新,可能会导致已有“投资”成为沉没成本。所谓“投资”,不仅包括硬件和操作系统,可能还包括评价应用系统、人员的知识经验与技能、培训资料等。若软硬件升级导致历史数据无法使用,可能会造成更严重的后果。因此,对于常规性的评价工作,信息技术的不稳定所造成的成本增加,是必须要充分考虑的。

信息技术的不可靠性很隐蔽,给教育评价质量造成的影响可能很大,可能会导致数据收集失败,或者数据分析出现难以察觉的错误。为应对这种不可靠、

① 教育评价的自我强化性指一种评价一旦制度化就会不断寻找证据证明其自身的合法性与有效性,并伺机扩展其影响面。信息技术应用一旦制度化,也存在类似倾向。

② 近期一个典型案例是Flash。Adobe公司在2020年12月停止更新Flash播放器,主流浏览器也不再支持,对于很多交互式评价资源库、视频库的影响很大。



不确定,在正式数据收集之前应进行全流程的充分测试,而且要制定备用计划,如利用本地服务器、预留备用数据收集时间、保持备份服务器同时运行等,这都会增加组织、时间、人力成本。

(三)易流于“数据驱动”,不利于评价的价值导向与问题驱动

“数据驱动”是当前教育信息化领域的“日常”用语。和所有教育术语一样,数据驱动并没有其字面意思所暗示的那样客观、中立、直截了当。就“数据”一词而言,尽管人们仔细厘清了现象、数据、信息、知识等相关术语的含义,也达成了一些共识,但在实践或研究中数据指代的东西非常多样化。例如,以下都被称为“数据”:学生的身高、一节微课的观看次数、一节数学课的视频、学生期末考试的排名、用回归方程预测的学生高考分数、学生志愿活动的记录单……,但它们的确定性以及所包含的加工处理、推断程度,代表的意义、价值(信息量)存在很大差异。也就是说,“数据”自身就是多样多层次的。最基础的原始数据只是“原油”,量大、庞杂、价值稀疏,需要“精炼”才能使用。在“数据驱动”过程中,选择什么样的数据,如何分析数据,如何呈现分析结果,分析结果和预期有出入,尤其是数据对于目标产生怀疑时该如何处理冲突,这些问题的解决有赖于隐含在背后的价值导向与问题定位。若望文生义,仅从字面上看容易忽略这些问题。具体到教育评价,在利用信息技术获取大量数据后,一味强调“数据驱动”存在如下风险:

第一,过分关注指标操作定义的便利性,降低评价效度,影响价值导向。在确定评价指标体系与指标的操作定义时,不应仅考虑数据易获取性,更要注重价值导向与问题逻辑,否则会降低评价效度。例如,在高校学科排名中,指标“课程教学质量”的操作定义就不尽合理(石中英,2019)。

第二,满足于相关关系,不去探究因果机制,降低评价结果的发展作用。评价结果要改变实践,就需要提供明确的操作建议,在不清楚因果机制和作用路径的情况下,仅靠相关(调查同一现象的多侧面数据,数据之间很容易相关,样本量足够大,也很容易在统计上显著)实践建议是否有效是随机的,长此以往会弱化评价的发展作用。

第三,在教育研究领域,“数据驱动”研究取向

的结果令人失望,这给同样是知识生产的评价领域敲响了警钟。MOOC让教育研究领域兴奋不已的原因之一是可以获取大规模真实的学习行为数据,有望极大推动教育规律的探索。但是,2013年充满希望的预期(李曼丽,等,2013,pp.176-177)并没有成为现实;七年后,Reich梳理了相关研究,发出“海量大数据,微量新洞见(terabytes of data, little new insight)”的慨叹;即便有一些发现,也只是类似“学习行为越多,学习成绩越好”的常识性结论(Reich,2020,p.37)。

(四)易形成“算法霸权”,降低评价结果的可解释性

《算法霸权》是奥尼尔(2018)著作的中文译名。书名直译应为《杀伤性数学武器——大数据是如何扩大不平等和威胁民主的》。“算法霸权”生动地表达出某些控制数据和算法的机构对个体泰山压顶式、不容置疑的威权性,故在此直接借用。

数据驱动这一术语简短有力,但其背后有复杂的数据处理过程。如前文所述,原始数据复杂多样、意义不彰,难以驱动决策;由于其量大且形式多样,处理复杂,个体难以在有限的时间内进行有效分析处理进而把握其意义,因此需要借助计算机计算分析。借助算法从数据中提炼出“模型”,基于算法和模型所得出的“结论”才是真正能够驱动决策与行动的。

在商业、体育、金融、娱乐、招聘等领域,算法模型已得到广泛应用。算法在为资本赚取高额利润的同时,恶化了处境不利群体的生存状况(奥尼尔,2018,p.172)。奥尼尔(2018,p.17)认为信息技术不仅能够造就好的模型,更容易造就不好的模型(杀伤性数学武器)。前者如棒球队模型,使用者(教练和球员)清楚模型的原理、能够用最新的真实数据调整模型、模型使用直接变量(如队员的比赛得分),而不是间接、替代变量(如队员的胡子长短)做预测;后者如营利性大学招生广告投放算法、招聘简历筛选模型,与棒球队模型相反,不透明(学生不了解其运作机制)、大量使用替代变量(种族、居住地等与学术表现无直接关系的变量)、模型更新不及时(模型训练出来后持续运行,不会根据最新数据更新),使迫切需要教育机会的不利人群为教育花费更多,背负上沉重的贷款负担,处境更糟。

教育评价应具备权威性。因某些评价的高利害性,公平公正的权威性有时会给评价对象造成很大的心理压

力。信息技术尤其是互联网和以数据为基础的智能技术,会大规模扩大专业权威的作用面,可能加剧其不透明性倾向,使其成为更具社会压迫性的力量。

可以预期,在未来评价中算法自动做出的判断会越来越多。“算法霸权”因判断过程的不同特点,存在很大差异。具体来说,第一类是人类个体能够达成高度一致、能够解释判断过程的,在这种情况下算法的危害不大,即使有问题也能很快调整。比如,阅卷时的题目切分、单词拼写、自动推荐的学习材料的适切性等。第二类是人类个体能够基本达成共识,但不容易解释清楚评价过程和依据的,在这种情况下算法如果有问题会造成混乱和不公平,虽然模型能调整,但是代价很高。比如,作文质量判断、一节课的教学效果、英语口语水平测试等。以英语口语自动评价为例,用东南沿海地区初中生的英语口语料训练出来的模型,在全国使用或者用于西部地区就可能存在公平性问题。第三类是对于个体未来表现的预测性评价,其影响因素众多,且效果验证具有滞后性,人类专家有时也难以达成共识,在这种情况下算法霸权的危害可能会极大。例如招生考试、企业招聘等。预测正是智能技术的特长,但这也是智能算法最受诟病的地方。目前企业招聘部门普遍利用智能算法筛选简历,其结果是,很多应聘者因为相关变量,如性别、居住地、第一学历学校级别等就被刷掉,根本没有展示与岗位相关能力的机会(奥尼尔,2018,p.126)。这样一来,短期是对特定群体中的个体利益的侵害,长期将降低社会的垂流动性,助长阶层固化。这是评价缺乏效度的典型案例。这警示我们利用信息技术并不一定能够提高评价质量。

当智能技术大规模应用于评价,尤其要警惕这种现象的出现,要警惕不可解释、不透明的算法所做出的评价判断。这里的不可解释指技术上的不可解释,如深度神经网络训练出的模型;不透明,指人为设置制度障碍来阻止利益相关者理解算法决策原理,比如以商业机密为由拒绝公开算法细节,不透明算法的伦理风险更高。不管哪种情况,都需要逐步纠正改进。教育评价发展功能的实现,需要评价对象理解评价标准和过程,需要算法具有可解释性、透明性。

(五)易滥用过程性数据,混淆形成性评价与过程评价

《方案》将“强化过程评价”作为改革的主要原则之一。鉴于信息技术极其擅长过程跟踪、数据整合

(重组重用)、数据管理,再加上目前对教育大数据、学习分析的普遍关注,这些因素与对过程评价创新的热望与压力重叠,可以想象在未来几年,用信息技术创新过程评价一定是研究与实践的热点。信息技术会在过程评价变革中起到重要作用,甚至会成为过程评价创新的基础。但也需要注意一些可以预见的挑战与风险:

第一,对于过程评价的认识不统一,容易导致实践偏差。过程评价不是新鲜事物,在研究与实践中已长期存在。但即使在学界,对其也存在多种认识。雷新勇(2020)列举了三种理解,即对学生发展的过程性评价(如档案袋)、对学生知识技能习得过程的评价以及对教师教学过程的评价。边新灿(2016)探讨了高考中的过程评价,认为它是为了缓解结果评价独大、“一考定终身”的不良影响而设计的,其实现路径包括把高中学业水平考试纳入高考评价体系、部分科目实行一年多考等。就学生评价而言,雷新勇(2020)的第一种理解和边新灿(2016)的理解是一致的,也是合理的,即过程评价与结果评价相对,强调在为选拔、问责而评价时,要综合某一时间段的多次表现信息(过程评价)和最后某一时间点的表现信息(结果评价),而不仅仅靠最后一次结果进行评价决策。

第二,易混淆形成性评价与过程评价,滥用过程性数据,造成外部总结性评价泛化。形成性评价以为正在进行的教育活动提供反馈为目的,旨在改进课程设计、教学实施和学生学习;它与总结性评价相对(陈玉琨,2019,p.18)。形成性评价是课程与教学的有机组成部分,是教师教学工作的重要方面,应该在教师专业自主权范围之内。例如,单元测试是很常用的形成性评价手段,帮助教师了解教学效果,帮助学生查漏补缺。这类测试成绩一旦变为总结性评价(如初中学业水平考试、中考、高考)的一部分,它就不再是形成性评价,而是过程评价(陶西平,2012,p.173)。由于信息技术在学校教学中的广泛应用,使学习过程数据极易收集,尽管过程评价改革的压力很大,教育局、学校一定要抑制将形成性评价转变为过程评价的冲动,这有很高的伦理风险,同时可能会锁定学生成长,增加学生学业负担,还可能导致家庭对学生学习产生更大的影响力。

第三,区分评价对象与评价类型,合理设定过程评价与结果评价的比重。对处于成长中的儿童和青少



年个体,过程评价存在锁定效应与可信性风险,应充分尊重个体生命成长的不确定性、非线性与独特性,慎用过程评价,防止过程评价限制、锁定学生发展,导致“浪子”无法“回头”。对于教育机构(如学校),过程评价有助于提升教育供给质量,提高日常教学、服务、管理质量,故宜增加过程评价占比。

四、教育评价中信息技术可持续应用的对策建议

基于以上分析,为促进教育评价中信息技术的可持续应用,使之切实为教育评价服务,进一步提高教育评价的科学性、客观性与专业性,建议如下:

(一)开展评价主体的教育评价信息技术应用能力建设,提升教育评价质量

学校和教育局是教育评价信息技术应用能力建设最重要的主体,其能力建设主要体现在人才储备和体制机制建设两个层面。信息技术只有通过专业技术人员方能实现“赋能”。《方案》也将“培养教育评价专门人才”作为组织实施的保障性举措。具体建议:第一,变革教育评价、教育测量等相关专业课程设置,在常规统计学、测量学课程之外适当增加线性代数、计算机编程和数据科学等相关课程,提高学生的信息技术素养、数据素养;第二,对于在职的教育评价专业人员,鉴于MOOC在在职/专业硕士教育中的有效应用(Reich, 2020, p.20),可以通过由专业机构(如统计与测量专业相关协会)开设网络直播讲座、MOOC课程等方式,提供课程学分认证,促进其专业发展;第三,在国家或区域层面,建设教育评价信息技术应用案例库,以传播先进经验;第四,倡导科教、产教融合,鼓励引导专业评价机构与信息技术行业企业探索整合力量的开放机制。

(二)研制教育评价数据规范,保证数据向后兼容,缓解信息技术不稳定带来的影响

数据规范的作用主要体现在以下几个方面:一是在信息技术快速更新(不稳定)仍会持续的情况下保

证数据的向后兼容性^①,保护最重要的资产“数据”在未来的可用性。二是在鼓励利用信息技术创新评价工具的情况下,为了降低创新成本,“数据孤岛”现象在所难免,关键是这些系统之间在必要的时候能够交换、集成数据,这时数据规范就能够起到作用。例如,目前各省的高中生综合素质评价数据不兼容、难以整合的现实问题,是招生高校在操作层面面临的难题,即评价数据规范已有现实需求。三是好的规范能够起到引导发展、帮助评价主体提高数据意识的作用。四是对于国家省部级立项的课题或行动计划,可以强制其遵循数据规范,逐步规范数据采集、存储,在这些项目的引导与示范下,数据规范有望更好地落地。

另外,在研制数据规范的同时,为促进规范落地,也为了更大规模的研究与人才培养,还应同时建立符合规范的、经过匿名化处理的、由真实数据构成的公共测试数据集。一方面是为了实际测试教育评价数据规范的可行性,另一方面也是希望起到ImageNet对于人工智能计算机视觉领域的作用^②,促进教育评价行业的健康发展。

(三)建立评价迭代更新机制,落实价值导向与问题驱动,规范“数据驱动”

教育评价改革不可能一蹴而就,教育评价中的信息技术应用也是如此。教育评价的“再评价”能够提高教育评价信度、效度和效益,但再评价的实践还不普遍(吴钢,2015, p.180)。可以以信息技术应用为契机,建立鼓励再评价的机制,逐步完善评价迭代更新的机制。在再评价过程中,为避免过分强调数据而忽略价值导向,应重点核查评价的效度,尤其是指标体系操作化涉及的数据与评价目的、内容的直接相关性。应实现评价指标和操作方法的公开透明,向评价对象说明评价各项指标的意义,以及指标数据对于实践活动的促进作用。以此逐渐杜绝为了方便而进入评价指标的数据,并规范、约束和引导“数据驱动”。

(四)发展评价对象自评能力,探索评价决策知情与算法透明机制,对抗“算法霸权”

为对抗“算法霸权”,提高算法与模型的可解

^① 向后兼容,在计算机领域可以理解为,软件更新到新版本后仍然能够使用旧版本软件创建的文档。例如,Office 365仍然能够编辑Word6的DOC文件。

^② ImageNet是美国斯坦福大学李飞飞教授发起的大型可视化数据库开源项目。现包含1,400多万张人工标注过的照片。ImageNet是计算机视觉物体识别领域最有影响力的测试数据集,已成为基准测试集,极大促进了该领域的发展。参见<http://image-net.org/>,2021年1月访问。

释,建议采取两方面的举措:

第一,建立关键算法与模型的公共审查机制。为防止“算法霸权”,应要求产品研发机构向公众或者利益相关者说明算法和模型原理,将其作为教育产品准入或者认证的基本要求。例如,在消费领域大数据杀熟等现象已引起普遍关注,中国消费者协会于2021年1月召开座谈会呼吁加强网络消费领域算法规制,保障消费者知情权、选择权和公平交易权^①。

第二,激发评价对象知情权意识,发展利用信息技术的自我评价能力。教育评价的“初心”是为了提高教育质量。正如王策三(2005, p.305)所指出的,在教学过程中的评价,最有效的是发展“自我评价”能力。在个体对抗“算法霸权”时,个体知情权意识的觉醒、对于评价本身的批判分析能力是最根本的。互联网与人工智能在这方面作用空间很大,一方面提供了评价对象共同体,可以汇聚发现的问题,另一方面也提供了公共数据、算力和工具(涂子沛,2014, p.272),因此有没有意识与能力运用这些工具就非常关键,而这有赖于个体评价素养的提升。

(五)提升全社会教育评价素养,确保教育评价的发展功能落实

《方案》落实是需要全社会参与的系统工程。由于教育评价的专业性与综合性,教育评价的直接利益相关者——教师、学生和家,都需要提升教育评价素养(评价意识、知识、技能与能力)。

教师是核心,而且要起到引导和辐射的作用。应为教师提供系统的培训,帮助教师认识到教育评价对于自身、学生以及课程教学的影响,尤其要明确形成性评价与过程评价的区别,尊重学生发展的规律,明确信息技术自动评价的问题与风险。

学生往往是大规模评价的直接对象,但因年龄段的差异,对评价的认识深浅不同。对于初中及以上的学生,需要教师在教学中明确培养学生的“自我评价”能力。同时,对于各种正式评价的目的、方式、后果,要帮助学生对此有日益清晰的认识。

对于学生的正式评价,尤其是在基础教育阶段,真正感兴趣的可能是家长,而不是学生。对于家长的教育评价素养提升,可借助两个途径:一是依托学区

学校的家长学校,通过专题讲座等方式系统地帮助家长改变观念、提升认识水平;二是依托我国的科学普及体系,通过科协的科普活动在科技场馆、网站、电视等各种平台上增加有关心理测量、考试、大数据与机器学习基本原理等知识。毕竟基于证据进行审慎判断与决策,也是科学素养的应有之义。

五、小结

本文一开始就强调教育评价是特殊的知识生产实践,是为了更突出评价也是对于未知的探索。对于教育评价,最根本的对象恰恰是充满可能性与不确定性的人类个体的身心成长与发展状态。评价者只能依靠数据去推测其真实状态。由于我们价值观念与认识能力的局限,推测是有可能与客观事实不相符的。因此,应对评价对象始终抱有认知上的谦卑与谨慎。评价的特殊性在于评价所产生的知识是针对特定对象的,情境性很强,而且具有直接的行动意蕴,会对特定的人类个体命运产生直接影响。在这个意义上,与其他教育研究相比,评价要求更高的知识与伦理标准。

在信息与智能时代,信息技术之于评价活动,正如技术之于人类的现代生活,“技术搭建起了我们居住的这所房屋……与早先的人类相比,我们几乎已没有生活在房屋之外的机会”(富兰克林,2019, p.9)。教育评价也已很难走到信息技术搭建的“房屋”之外。因此,在教育评价中用不用信息技术已经是伪问题,如何用好信息技术才是真问题。《方案》对教育评价应充分利用信息技术的强调和要求,能够推动信息技术应用的实践创新,推动其进一步规范化与规模化,也能够提高研究的自觉性、前瞻性与系统性。

信息技术,尤其是新兴的智能与网络技术,让教育评价者有了强大的数据收集、分析与应用工具。评价者对于数据收集过程的控制力,数据分析可用的数据规模、计算资源,评价可能影响的个体范围,都有了很大的扩展。智能技术的广泛采用更有可能大大加强评价的自动化、泛在性与影响力。但是,需要认识到,对于教育评价,即使全过程都利用了信息技术,有所谓全过程的“大数据”,也并不会改变评价的取

^① 新闻来自中国消费者报,参见《中消协呼吁加强网络消费领域算法规制 强化算法实施方的举证责任》, <http://www.ccn.com.cn/Content/2021/01-08/1643471552.html>, 2021年1月访问。



样与推测本质,即评价本质上依然是利用证据推测“未知”特质的研究性活动。信息技术支撑的评价活动,其推测推理成分并未减少,价值导向与问题驱动反而前所未有地重要。

[参考文献]

- [美]阿杰伊·阿格拉沃尔,乔舒亚·甘斯,阿维·戈德法布. 2018. AI极简经济学[M]. 长沙:湖南科学技术出版社.
- [美]阿兰·柯林斯,理查德·哈尔弗森. 2013. 技术时代重新思考教育:数字革命与美国的学校教育[M]. 陈家刚,程佳铭,译. 上海:华东师范大学出版社.
- 边新灿. 2016. 过程评价:高考改革和高中课程改革的共同探索——兼论高中统考、综合素质评价纳入高考评价体系的内在逻辑和现实途径[J]. 考试研究(5):37-43.
- 陈玉琨. 2019. 教育评价学[M]. 北京:人民教育出版社.
- [加]厄休拉·M. 富兰克林. 2019. 技术的真相[M]. 田奥,译. 南京:南京大学出版社.
- [美]富兰克林·福尔. 2019. 没有思想的世界:科技巨头对独立思考的威胁[M]. 舍其,译. 北京:中信出版集团.
- 顾明远. 1998. 教育大辞典[M]. 上海:上海教育出版社.
- 贾海生. 2019. 高考体育学科测试信息化的实践与思考[J]. 宁夏大学学报(人文社会科学版)(3):198-200.
- [美]凯西·奥尼尔. 2018. 算法霸权:数学杀伤性武器的威胁[M]. 马青玲,译. 北京:中信出版集团.
- 雷新勇. 2020. 关于教育评价改革的若干思考[J]. 中国考试(9):10-14.
- 李曼丽,张羽,叶赋桂. 2013. 解码MOOC:大规模在线开放课程的教育学考察[M]. 北京:清华大学出版社.
- 李政涛. 2020. 把新时代教育评价改革深化到“评价能力”的提升那里去[J]. 中国教育学报(12):8.
- [美]理查德·E. 梅耶. 2016. 应用学习科学:心理学大师给教师的建议[M]. 盛群力,丁旭,钟丽佳,译. 北京:中国轻工业出版社.
- 芦咏莉,申继亮. 2012. 教师评价[M]. 北京:北京师范大学出版社.
- [美]梅雷迪斯·D. 高尔,沃尔特·R. 博格,乔伊斯·P. 高尔. 2002. 教育研究方法导论[M]. 许庆豫,等,译. 南京:江苏教育出版社.
- [美]尼葛洛庞帝. 1997. 数字化生存[M]. 海口:海南出版社.
- [美]尼古拉斯·卡尔. 2015. 玻璃笼子:自动化时代和我们的未来[M]. 杨柳,译. 北京:中信出版社.
- 石中英. 2019. 学科排名的几点认识[M]//朱永新,袁振国,马国川. 重构教育评价体系. 太原:山西教育出版社,99-106.
- 孙鑫,黎坚,符植煜. 2018. 利用游戏 log-file 预测学生推理能力和数学成绩——机器学习的应用[J]. 心理学报(7):761-770.
- 檀慧玲,李文燕,万兴睿. 2018. 国际教育评价项目合作问题解决能力测评:指标框架、评价标准及技术分析[J]. 电化教育研究(9):123-128.
- 陶西平. 2012. 在反思中创新[M]. 北京:教育科学出版社.
- 涂子沛. 2014. 数据之巅——大数据革命,历史、现实与未来[M]. 北京:中信出版集团.
- [美]托马斯·M. 尼科尔斯. 2019. 专家之死:反智主义的盛行及其影响[M]. 舒琦,译. 北京:中信出版集团.
- 王策三. 2005. 教学论稿[M]. 第3版. 北京:人民教育出版社.
- 吴钢. 2015. 现代教育评价教程[M]. 北京:北京大学出版社.
- 辛涛. 2020. “探索增值评价”的几个关键问题[J]. 中小学管理(10):1.
- [荷]雅普·希尔伦斯,赛斯·格拉斯,萨利·M. 托马斯. 2017. 教育评价与监测——一种系统的方法[M]. 北京:教育科学出版社.
- 游忠惠. 2013. 高考全面进入信息化时代[N]. 中国教育报,06-10(7).
- 袁锐铿,易轶. 2004. 试析NBPTS优秀教师认定的标准与程序[J]. 比较教育研究(12):71-75.
- [美]约翰·希利·布朗,保罗·杜奎德. 2003. 信息的社会层面[M]. 北京:商务印书馆.
- 张志祯,童卫军,王博. 2015. “支持高等职业学校提升专业服务产业发展能力”项目——绩效评价报告[M]. 北京:机械工业出版社.
- 张志祯. 2016. 虚拟现实教育应用:追求身心一体的教育——从北京师范大学“智慧学习与VR教育应用学术周”说起[J]. 中国远程教育(6):5-15,79.
- 周洪宇. 2020. 以科学的教育评价推动新时代教育学发展[J]. 中国教育学报(12):1-2.
- 周志华. 2018. 创办一流大学人工智能教育的思考[J]. 中国高等教育(9):52-53.
- 祝令健,沈旭东. 2020. 大数据技术下基于精准预习的高中物理精准教学探讨——以“向心加速度”为例[J]. 物理通报(2):40-44.
- [美]N. 维纳. 1978. 人有人的用途:控制论和社会[M]. 陈步,译. 北京:商务印书馆.
- [美]Russell, S. J. & Norvig, P. 2013. 人工智能:一种现代的方法[M]. 第3版. 殷建平,祝恩,刘越,陈跃新,王挺,译. 北京:清华大学出版社.
- [美]Thorndike, R. M. & Thorndike-Christ, T. 2018. 教育评价:教育和心理学中的测量与评估[M]. 方群,等,译. 北京:商务印书馆.
- Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470-497.
- Reich, J. (2020). *Failure to disrupt: why technology alone can't transform education*. Cambridge: Harvard University Press. (Kindle edition).

收稿日期:2021-01-05

定稿日期:2021-01-15

作者简介:张志祯,博士,副教授,硕士生导师,北京师范大学教育学部教育技术学院(100875)。

齐文鑫,硕士,助理研究员,副部长,国家开放大学教务部(100039)。

责任编辑 刘莉

Abstracts

Information technology in educational assessment: empowerment, challenges and solutions

Zhizhen Zhang and Wenxin Qi

Educational assessment shapes the value orientation of education. Appropriate use of information technology (IT) may contribute to the guiding role of educational assessment. This article argues that IT can enhance and innovate educational assessment, optimize assessment management, improve assessment quality and use assessment outcomes for a wider range of purposes. Nevertheless, despite these advantages, IT in educational assessment also brings about challenges and risks. For example, evaluators may not be able to use IT skilfully; IT may not always be robust; assessment may be driven by data and manipulated by algorithm or misuse process data. Solutions proposed to address these challenges include capacity building for evaluators, standardization of assessment data, development of open test datasets, establishment of iterative assessment mechanisms, cultivation of students' self-assessment competence, and improvement of assessment literacy in the wider community.

Keywords: educational assessment; IT application; assessment process; assessment quality; data-driven; algorithm hegemony

Towards an occupational skills framework for online learning facilitator

Shuang Li, Hairong Wang, Hua'nian Cui and Qinhua Zheng

Online learning facilitator was officially recognized as a new profession by the Chinese Ministry of Human Resources and Social Security in July, 2020. Informed by learner support theories, this study attempts to investigate their status quo, professional development needs and professional identity as well as formulate a framework of online learning facilitator occupational standards. It is argued that this new profession reflects new developments in learner support as well as innovations in re-structuring instructional process and reforming educational service mechanism. The study also identifies four characteristics of online learning facilitation, i. e. knowledge-intensive, high-tech, data-driven and multi-dimension expertise. The study draws upon the National Occupational Skills Standards 2018 to establish a framework for this new profession comprising 13 occupational functions, 3 occupational directions and 5 occupational levels. Future directions of development are also discussed.

Keywords: online learning facilitator; perception of professional identity; occupational skills standard; standards framework; learner support; Internet Plus education; division of labor; educational service; reform of educational service system

Technology enhanced learning: getting the technology mix right

Michael Sankey

This paper explores the use of technology enhanced learning that is used to enhance student learning. It emphasises that technologies afford opportunities to enhance learning, but successful outcomes depend on the way in which technologies are embedded in subject and course design and the extent to which they facilitate alignment between expected learning outcomes, coursework and assessment. The use of the term "Technology Enhanced Learning" (TEL) is placed into a context within the paper to establish a clear focus on student learning as it guides you towards making effective decisions about how institutions establish, support and quality assure TEL.