

“互联网+”时代数据驱动的教育评价体系构架与实践进展*

杨现民¹, 顾佳妮¹, 邢蓓蓓²

(1. 江苏师范大学智慧教育学院, 江苏徐州 221116; 2. 江苏师范大学化学与材料科学学院, 江苏徐州 221116)

摘要:“互联网+”时代,教育评价从启蒙时期、开创时期、批判时期、专业化时期走向了精准化时期。与传统教育评价相比,数据驱动的教育评价在评价功能、评价对象、评价内容、评价方法、评价数据来源、评价数据分析、评价过程、评价结果等方面都存在较大转变,并呈现出六大特征:增值性、多元性、综合性、实时性、精准性和公平性。结合教育评价过程的准备阶段(评价目的、评价对象、评价目标、评价指标体系、评价标准、评价方法)、实施阶段(评价数据采集、评价数据处理、评价数据分析)以及决策阶段(评价结果生成、评价结果反馈)的各方面需求,构建出数据驱动的教育评价体系框架,并分别从国家、区域、学校三个层面分析当前数据驱动的教育评价实践进展。为促进数据驱动教育评价的发展,提出了四条建议:开展数据素养专项培训,提升评价者基于数据的评价能力;注重教育评价数据的开放与共享,确保教育评价的公正与精准;建立大数据智能测评分析系统,促进教育评价效率的提升;倡导增值性评价理念,完善数据驱动教育评价体系。

关键词:互联网+;教育评价;数据驱动的教育评价;体系构架;实践进展

中图分类号:G40-058.1 **文献标识码:**A **文章编号:**1001-5035(2019)04-0016-11

引言

2015年,李克强总理首次在《政府工作报告》中提出“互联网+”行动计划,正式揭开了“互联网+”时代的序幕。“互联网+”理念正深刻影响和重塑传统产业格局。同其他产业一样,教育的面貌也在不断改变,“互联网+教育评价”是“互联网+教育”的内容创新,教育事业的改革与发展离不开教育评价体系的支撑,合理的教育评价实践有利于保障国家教育方针的顺利实施。传统教育评价存在一些缺陷,如评价标准和内容过于片面、缺乏真实性和动态性评价、对数据挖掘和利用不充分、难以开展持续性和终身性评价等,一定

程度上阻碍了教育信息化和现代化的进程。

习近平总书记在全国教育大会上提出:要深化教育体制改革,健全立德树人落实机制,特别强调要扭转不科学的教育评价导向,从根本上解决教育评价指挥棒问题。^[1]得益于大数据、云计算、物联网等新一代信息技术的快速发展,大规模数据急速产生并流通,数据已成为当今时代重要的战略资源。^[2]随着数据密集型科学范式的出现,“量化一切”“让数据发声”等口号成为新时代的重要标志,以数据驱动为导向的研究方式为教育评价的研究提供了新思路,推动教育评价走向数字化、科学化、智能化和专业化。^[3]

收稿日期:2019-05-25

作者简介:杨现民(1982—),男,河北邢台人,江苏师范大学智慧教育学院教授,理学博士;顾佳妮(1995—),女,江苏常州人,江苏师范大学智慧教育学院硕士研究生;邢蓓蓓(1984—),女,河北邢台人,江苏师范大学化学与材料科学学院实验师。

基金项目:教育部人文社会科学研究青年基金项目“智慧课堂数据体系构建与应用研究”(18YTC880095);江苏省高校哲学社会科学重点研究基地重大项目“‘互联网+’时代高校智慧型课程建设与发展研究”(2015JDXM021)

一、教育评价的发展阶段

教育评价是人类对教育过程的一种价值判断活动。它起源于中国古代的教育考试制度,孕育于西方的教育测量批判思潮,形成于美国 20 世纪 30 年代的“八年研究”改革运动。纵观教育评价发展的历史和评价的重点差异,可以将教育

评价的发展划分为五个时期,分别为教育评价的启蒙时期、开创时期、批判时期、专业化时期和精准化时期(见图 1)。随着时代的发展和评价体系的不断完善,评价手段与评测技术的融合性和评价结果的客观性逐渐增强,评价主体、评价形式和评价方法也愈来愈多元化。

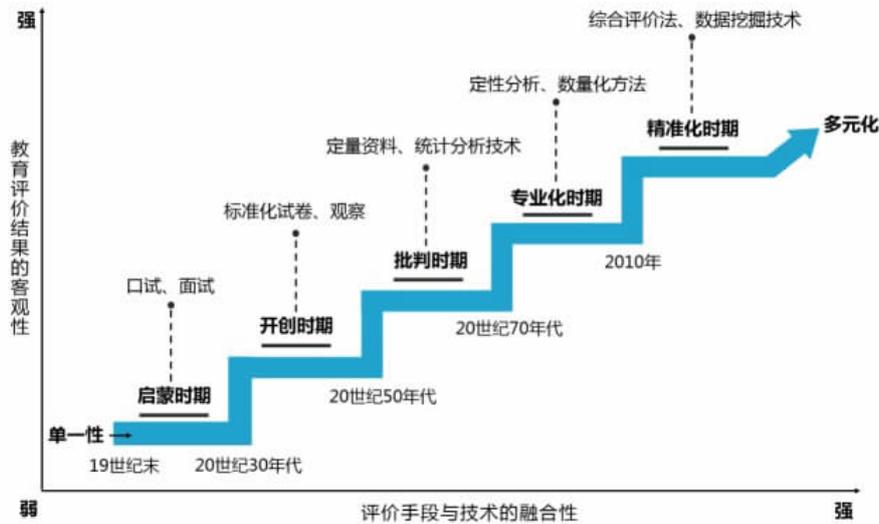


图 1 教育评价的发展阶段

(一) 教育评价的启蒙时期(19 世纪末至 20 世纪 30 年代)

1864 年,英国格林尼治医学校教师 George Fisher 公布了习字、拼字、数学、圣经和其他科目的标准对照表和实例,并规定了五分制评分的标准。1897 年,莱斯(Joseph Rice)发表了对 20 所学校 16 000 名学生所做的拼字测验结果,大大推动了对教育测量的研究。英国的差异心理学创始人高尔登(Francis Galton)是最早应用统计方法处理心理学研究资料的学者,他认为人的所有物质都可以定量叙述。高尔登的理论为教育评价工作的量化技术提供了重要基础。1904 年,美国教育心理学家桑代克(Edward Lee Thorndike)出版了《心理与社会测量导论》,该书介绍了心理统计方法和编制测量的基本原理,为教育测量的客观化、标准化奠定了理论基础。1905 年,法国人比纳(A. Binet)与其助手西蒙(T. Simon)编制了测量智力的“比纳-西蒙量表”。1923 年,美国发布了第一个标准化成绩测验——斯坦福成绩测验,标

志着教育测量的编制和运用已十分成熟。这一时期教育测量方法被广泛运用,但其仍存在一定弊端。首先,教育测量的研究对象有限,如学习兴趣、学习动机等人格因素就难以全面量化;其次,教育测量只关注学生学习结果的测量,而无法对学习过程做全面测验。为此,研究者和实践者不断探索更加科学的测量方法,于是教育评价应运而生。

(二) 教育评价的开创时期(20 世纪 30 年代至 50 年代初)

为了解决 1929—1933 年经济危机后出现的学生需要与学校课程间的尖锐冲突,泰勒(Ralph W. Tyler)受进步主义协会邀请,主持了为期八年的课程与评价研究,以帮助教师形成有效的评价方法去了解学生参与新课程之后的变化。^[4]泰勒认为学生学业成绩测验存在片面性,因此他提出了区别于“测量”的新概念——教育评价(Educational Evaluation)。泰勒的评价是以目标为导向,把评价内容分成具体可见、可操作的学生行为目

标,根据预期的目标有计划地收集资料,去判断实际活动是否达到预期目标或者达到了何种程度,其中收集资料的手段有多种(笔试、观察记录、问卷、访谈等),但任何一种手段都需要遵循客观性、信度和效度准则,否则评价都将无效。泰勒的评价实证化特点非常明显,他的工作也取得了巨大成就。这一时期基本可以称为“泰勒时期”,泰勒以目标为中心的评价极大提高了评价的实用性,^[5]但同时它也有其自身难以克服的局限。首先,并不是所有目标都可以转化成行为目标,如道德意识层面的目标就难以评价;其次,泰勒的评价过分关注如何实现预期目标,而忽略了对行为目标本身的评价。

(三) 教育评价的批判时期(20世纪50年代至20世纪70年代)

这一时期,教育评价方法的实证化仍占主要地位,但随着评价需求的快速增长,泰勒评价模式受到了严峻挑战,一些学者纷纷对其提出了质疑和批判,克隆巴赫(Lee J. Cronbach) 在一篇题为《通过评价改进课程》的论文中强调评价应发生在教育过程中,而不是教育过程结束之后,它不仅关注预期教育目标的实现,更应关注针对评价信息的教育决策。斯塔弗宾(Stufflebean) 同样认为教育评价就是为决策者提供服务的过程,为此,他提出了以决策为中心的 CIPP 评价模式,帮助决策者改进教学和管理。除此之外,斯克里文(Michael Scriven) 也认为教育评价不仅要教育目标的预期效果进行评价,一些非预期的效果也需要在评价中有所体现,因此,他提出了目标游离的评价模式,主张评价活动与教育目标分离,评价者应在不知道预期目标的情况下全面搜集评价信息,保证评价结果的真实性。

总之,这一阶段的评价都是评价者根据一定的标准去判断所得结果是否达到预期目标,从而作出价值判断。该时期的不足体现在过分强调科学实证主义的方法,忽视了定性方法的使用,严格、固定的评价流程使得评价活动缺乏灵活性和弹性。

(四) 教育评价的专业化时期(20世纪70年代至21世纪初期)

这一时期,大量教育评价刊物、书籍被出版,

各种研究机构、评价组织纷纷涌现,评价专业化活动也迅速兴起,教育评价进入了专业化深入发展时期。^[6]教育评价方法的人文化倾向得以强化,出现了一些新的评价模式,如斯塔克(Stake) 等人的应答评价模型,库巴(E. Cuba) 和林肯(Lincoln) 的自然主义评价模式等。这些评价模式不过分追求评价的客观性,更多关注评价参与者对评价对象的主观性认识。该时期的评价坚持“价值多元化”理念,因而评价结果是在评价过程中,以磋商的形式,不断消除分歧,最后形成一致的、公认的观点。在主张个人发展的质性分析的同时,实证化特征也同样得到了进一步发展,泰勒的目标导向评价模式在经过波帕姆(W. James Popham) 和布鲁姆(B. S. Bloom) 的改进后,仍在教学领域广泛应用。^[7]

总的来说,教育评价的价值取向从对学生的甄别选拔转变为关注个性化发展;评价方法从标准化测验转变为定性分析和定量分析相结合;评价方式从终结性评价转变为形成性评价、过程性评价和终结性评价相结合,教育评价体系正在逐步完善,也得到越来越多教育工作者的运用。教育评价走向客观是一种趋势,然而在大数据时代到来前,受限于数据采集技术,教育评价依赖于小样本数据,评价结果虽然能在一定程度上描述客观事实,但是仍然不够全面、精准。^[8]随着大数据的发展,小样本数据变为全样本的大数据,基于数据驱动的精准评价时代到来。

(五) 教育评价的精准化时期(2010年至今)

新一代信息技术(云计算、大数据、物联网、移动互联网、人工智能)与教育的深度融合促进了智慧教育的发展,催生出海量教育大数据,综合应用多种采集技术(物联感知、视频录制、图像识别、平台采集)对教育数据进行全面、自然、动态、持续采集,^[9]通过数据挖掘、内容分析、聚类、预测等技术方法透视数据背后隐藏的价值规律,精准分析评价对象的发展情况,针对现存问题进行及时有效的干预,并对其未来趋势做客观且科学的预测。总而言之,基于数据驱动的教育评价就是利用技术手段与方法,更加系统、科学、全面地收集、处理和分析教育信息和数据,对教育活动做出智慧判断的过程。相较传统的教育评价,数据

驱动的教育评价从经验主义转向数据主义,从模糊走向精准,在评价功能、评价对象、评价内容、评

价方法、评价过程、评价结果等方面存在较大转变(见表1),它具有如下特征:

表1 传统教育评价和数据驱动的教育评价对比分析

	传统教育评价	数据驱动的教育评价
评价功能	以决策为中心,重评价的鉴定和选拔功能。	以人为本的发展性理念,提倡评价的导向和激励功能。
评价对象	评价多是自上而下的单向评价。	评价主体多元化,提倡人人皆是评价者,人人皆是被评者,强调评价的双向性。
评价内容	大多是对学生成绩的评价,对道德意识、心理等方面难以评价。	教育的一切。
评价方法	单一性评价、终结性评价、静态评价。	全面综合评价、过程性评价、动态评价。
评价数据来源	小样本数据采集。	全样本数据搜集。
评价数据分析	定性分析通常为主观衡量,定量分析采用线性分析、回归分析等统计分析技术。	定性分析采用内容分析等技术,定量分析采用人工智能、大数据等新一代智能化技术。
评价过程	较为封闭。	评价活动透明、开放。
评价结果	评价结果较为客观,但不够全面。评价结果的直接受益者较少,有用性和有效性不够。	评价结果无限接近于真实状态,评价系统自动进行智能化诊断,并提供精准评价报告,同时针对存在的问题给予个性化反馈意见。不同的评价对象都可以是评价结果的受益者。

1. 增值性: 倡导以“改进”为主的评价目的

当前我国的教育评价以选拔为主,评价标准过于单一,评价内容仍停留在成绩的评定上,这在很大程度上阻碍了教育价值取向的创新发展。教育最大的价值在于发现个体兴趣,帮助其不断发展,最后创造出社会价值。大数据通过相关教育信息的关联挖掘、分析,能精准确定不同个体的认知起点,根据数据分析结果预测其未来发展趋势,帮助个体制定个性化的发展目标,且便于评价者实时监测个体的目标完成情况,及时提供针对性的改进建议。数据驱动的教育评价关注个体在一段时间内是否存在进步,以及具体的进步量是多少,例如通过当前学生质量与初始学生质量的增值帮助评价者快速洞察学习者在一段时间内的进步情况,同时进步量的差异也反映了不同干预措施的有效性。大数据“放大”了个体的细微变化,生动展示了教育“增值”的过程,同时也能为不同个体提供最合适的干预措施。

2. 多元性: 强调评价主体的多样化

管办评三位一体的教育管理格局使得教育评价的主体通常是政府和教育行政部门。这种评价主体相对单一的管理存在很多弊端,如长期以管理者主导评价容易导致教育评价行政化色彩浓重,美国评价学者派特(M. Q. Patton)早在1978年就提出了“多元主体参与”的概念。^[10]他认为使用评价信息的相关人员都应参与到评价过程中,但受限于技术环境,难以实现全员参与评价。新一代信息技术的发展为教师、学生、家长、管理者等不同主体提供了足够的空间和平台参与评价,不同评价主体根据评价需要有选择地处理数据,从而形成适合不同主体的评价报告。该种方式既拓宽了评价的视角,又增强了不同主体间的互动,还可以建立共同参与、相互影响的教育评价制度。

3. 综合性: 注重评价范围的全面性

过去,教育信息大多是“隐形”的,难以搜集、汇聚、分析和公开,然而大数据技术的出现促使教

育信息变得可量化。新一代信息技术的广泛应用为教育提供了海量的数据,使得教育评价的范围逐渐扩大到“一切教育和教育的一切”,^[11]它不仅涉及课堂教学质量评价、教师评价、学生评价、学校评价等微观、中观层面,还涉及区域教育发展水平评价、学科发展的国际比较评价、国家教育竞争力评价等宏观层面,任何教育发展和改革实践中存在的新问题、新进展都将纳入到教育评价的范围中。

4. 实时性: 注重评价过程的及时性

在传统教育评价中,评价者通常采用事后分析,导致评价结果具有一定的延时性。如果学习者在学习过程中无法及时获得关于自身的评价反馈,那么事后评价结果的作用也将大打折扣,因此教育评价要注意一定的时效性。大数据技术支持下的伴随式采集能实时获取学习者学习的过程数据和结果数据,形成评价数据源,根据评价指标建立数据分析体系,评价者在实时监测学习者学习效率时能将数据精准应用于其学习场景中,实现数据与学习者学习效率的融合转化,让学习者在学习过程中通过获得持续不断的评价结果反馈和精准干预而提高学习质量。

5. 精准性: 强调评价结果的有效性

以往学生学习的兴趣点、难点等都是凭借教师经验得以确定,但利用大数据技术能相对全面地记录学生的学习过程,如线上线下的学习表现、作业完成情况、师生互动情况、学习兴趣、学习态度、学习动机等,综合运用人工智能、图像分析、语音识别等方法分析学生学习和教师教学的过程,实现过程数据的智能化处理,利用决策树、机器学习、神经网络等技术掌握数据集中所隐藏的关系和价值,^[12]精准有效地指导学生反思学习过程,引导学生主动发现学习问题,并将教学过程中存在的问题准确反馈给师生,为促进学生的深度学习和教师的精准教学提供个性化指导。

6. 公平性: 注重评价程序的公开透明

原来的教育评价实施过程是一个“黑箱”,部分委托项目中,政府仍占据着评价程序的推进权和决策权,^[13]专家或第三方评估机构往往会因为种种利益关系而影响评价行为,因此很难保证评价程序的公正性。“互联网+”开启了一个高度

透明、具有更强公信力的时代,评价主体或评价利益相关体能看到教育评价过程中发生的一切(采集了哪些数据、如何聚合信息、怎样数据分析、做出什么决策)。伴随着对教育评价过程的公开,保障民众与教育相关部门的信息对等,从而建立彼此的信任,民众通过可信的信息和数据支持能做到有“据”可依,不盲从评价者的决策和干预。

二、数据驱动的教育评价体系构架

教育评价活动的实施大体都包含前期的准备工作、评价实施和评价决策三个阶段,每个阶段都由相互关联的若干程序组成。互联网+、大数据技术、人工智能技术的发展改善了传统的教育评价体系,数据驱动的教育评价体系构架(如下页图2所示)使得评价流程更具科学性、客观性和精准性。

(一) 评价的准备阶段

在评价的准备阶段,首先要明确评价目的,即评价活动想要实现的结果,是鉴定、分等,还是监控与激励;然后确定评价对象,也就是评谁?评什么?新技术的使用扩大了评价的范围,无论是微观层面的教学效果评价、学生综合素质评价、课程质量评价,还是宏观层面的区域教育质量水平分析、国家教育竞争力对比等,都能根据评价目的确定相应的评价对象。评价目标、评价指标体系、评价标准的确立是教育评价的关键环节,评价指标体系要尽可能全面且可操作,评价标准需紧扣目标。另外,教学平台、评价系统的成熟运用为教育评价提供了全新的评价工具。

(二) 评价的实施阶段

评价的实施阶段主要包括数据采集、数据处理以及数据分析三个阶段。首先需要采集与评价对象、评价内容、评价目标相关的信息数据,全过程数据的采集是数据驱动教育评价的重要环节,不完整的数据采集会影响教育评价的实施和评价结果的准确性。教育数据来源十分多元,根据教育数据来源和范围的不同,评价数据大体分为国家教育大数据、区域教育大数据、学校教育大数据、班级教育大数据、课程教育大数据和个体教育大数据。目前,物联感知技术、平台采集技术、视频录制技术和图像识别技术可对多元化数据进行

全自动、全过程、全样本、全类型的采集。全自动体现在评价数据通常是学生、教师、管理者和家长等评价主体在运用各种教育装备中自动产生,而非根据特定的评价目标选择性采集;全过程体现在充分借助先进技术平台,持续记录评价对象的学习轨迹,动态监测评价对象学习的全过程,在保

证数据有效的基础上,细化数据粒度;全样本体现在数据分析所基于的样本量不是纯微观抽样,而是全样本的总体数据,这也是大数据时代所强调的整体性思维;全类型体现在评价数据的类型除了结构化数据,还有非结构化数据和半结构化数据。

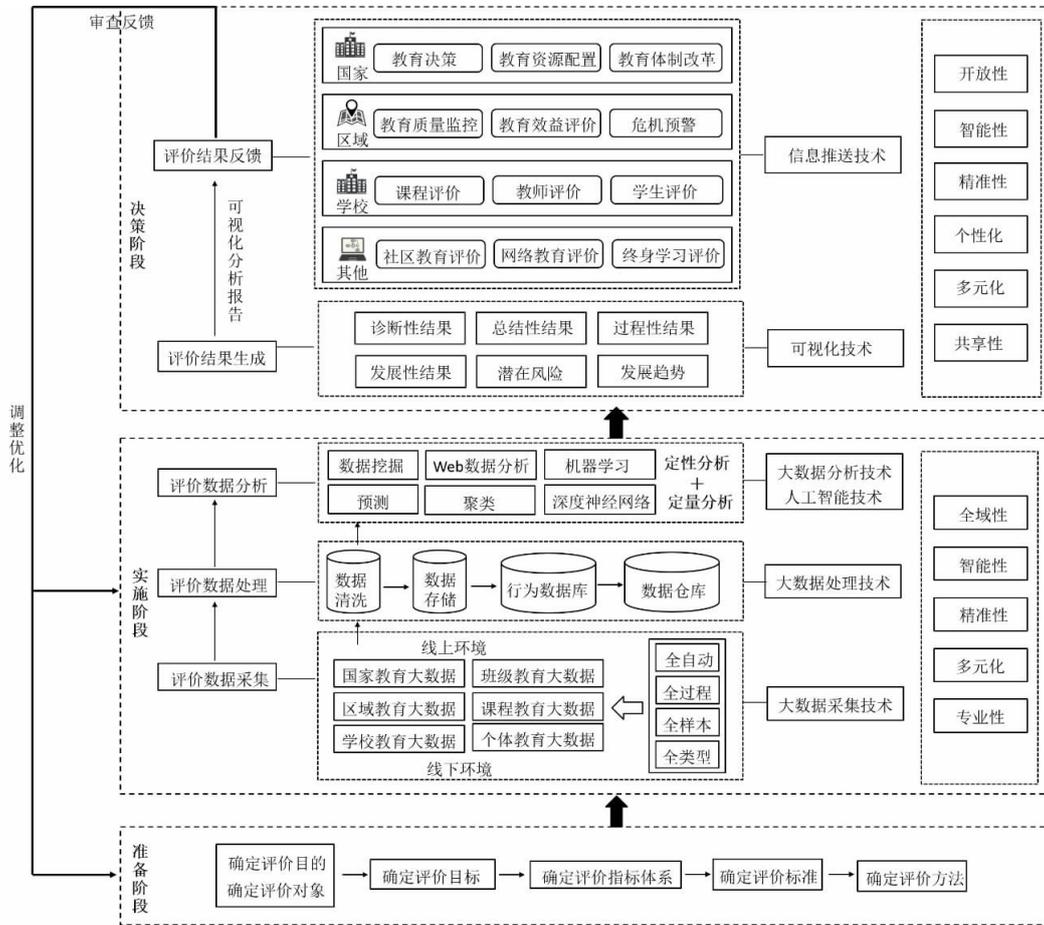


图2 数据驱动的教育评价体系构架

高质量的数据经过处理分析才能得到高质量的结果,因此需将采集的数据进行预处理,去掉低质量、不合要求的数据,然后将其余规范、合理的数据进行整合,集中存放在数据库中。数据存储需要根据数据类型和特点,对大规模数据进行结构化存储,存储的方式和特点将直接决定数据利用的效率。很多学习系统或社区都有各自成型的数据管理策略,学习者在不同的学习系统或社区中注册并留下学习轨迹,如果系统之间缺乏有效

数据链接,那么获得学习者完整的学习数据将十分困难。这就需要将不同来源的多样性数据加以整合,并将这些多来源的数据导入到同一个分析框架中,实现大规模数据的语义存储。

对于评价数据的分析既要综合运用传统的数据分析方法与工具,又要合理采用专门针对大数据处理的新方法与新工具。传统方法主要包括聚类、预测、关系挖掘、社会网络分析、文本挖掘、Web 数据分析等,新方法主要包括深度神经网络

分析、智能代理、机器学习、计算智能方法、图像分析、语音识别等,综合运用数据分析方法,对评价数据进行智能化处理,^[12]科学指导教育决策。大数据的使用一方面保留了教育评价的数量特征,保证了评价的客观性,同时使人文社会科学方法在评价中的应用有所创新,比如应答评价模式中评价者与评价对象之间的应答过程,教学平台提供了网络空间让评价对象及评价的相关人员针对某一问题各抒己见,网络空间可以容纳大范围、大规模的相关人员参与讨论,然后通过聚类、内容分析等技术快速统整意见,提炼核心观点。当发表意见的人数足够多时,就会削弱该方法带来的主观性影响,在响应人文主义思想的同时,也保证了评价结果的准确性。

(三) 评价的决策阶段

在评价的决策阶段,将教育评价结果以可视化的形式反馈给不同评价主体,以服务国家、区域、学校等进行教育决策。分析报告中呈现的内容主要包括诊断性评价结果、过程性评价结果、总结性评价结果、发展性评价结果、风险预测和发展趋势。诊断性评价结果是为了帮助决策者确认评价对象的初始发展水平;总结性评价结果是为了感知评价对象的整体发展情况;过程性评价结果是为了从细节处把握评价结果的精准性,通过了解评价对象发展的动态过程,以评价对象教育过程中发生的变化为事实依据,分析判断并给予评价对象实时反馈;^[14]发展性评价则是为了展示评价对象自身发展变化的轨迹,以初始发展水平为参考,呈现当前表现与过去表现之间的差距,便于决策者寻找影响评价对象发展的因素和促进其发展的策略。总结性评价结果、过程性评价结果、发展性评价结果等从整体到局部全方位挖掘评价对象在教育过程中的发展情况,精准了解当前的教育问题及潜在风险,及时进行预警并采取措施以有效规避风险。

数据驱动的教育评价应用具有多角度、全方位、多层次的特点。在内容上,既包括学业评价,也包括对学习态度和运用知识能力的评价,以及适当的价值观评价;在过程上,既包括课程评价,也包括教学及其结果评价;在层次上,涵盖微观、中观和宏观评价,从学生、教师评价到学校管理、

地方教育的行政评价以及在国家层面上教育政策、投入和教育改革等方面的评价等。

国家层面上,评价结果主要用于教育决策、教育资源配置和教育体制改革。在教育决策方面,通过对关键年龄段学生大规模抽样测试结果进行分析与数据挖掘,及时发现潜在的教育质量问题,并以高度可视化的方式呈现给决策者,以便决策部门调整教育政策,控制教育行为,从而稳定提高国家教育质量;在教育资源配置方面,教育管理者实时收集区域教育经费、学校分布、人口流动等数据,经过科学分析全面掌握不同区域的教育资源配置情况,管理者针对不同区域特点合理配置资源,促进教育资源均衡发展以及教育公平;在教育体制改革方面,教育管理者根据评价结果反映的问题重新审视当前的教育体制,及时修正教育体制中不合理的成分,通过了解教育供给中学习者的需要,为学生个性化学习提供精准支持,深化教育供给侧改革。

与国家层面评价相比,区域层面上由于教育发展差距较小,影响因素较少,获得的评价结果相对更为准确,可以为教育行政与业务部门采取相应的教育改革措施提供更准确的施政指标与证据,逐渐实现教育均衡、现代化以及提高教育质量等目标。评价应涵盖区域范围内的相关教育问题,包括学生参与性和进步程度、学生成就和素养、教育机构的组织和管理、教育机构的课程和发展趋势、机构的财政情况和人力资源等。

学校层面上,数据驱动的教育评价应用包括:(1) 对学生的评价。针对学习者相关数据的分析结果,评估其学习效果和学习状态,让教师更好地了解学生的知识掌握情况,预测学习者可能出现的学习风险,帮助学生查漏补缺,寻找相应的解决对策。(2) 对教师的评价。针对教师与学习者相关数据的分析结果,可以对教师的教学效果、教学策略的有效性、教学指导的满意度以及教学风格的接受度等进行评价,教师可以据此调整自己的教学方法与教学策略,促进教师的专业发展。(3) 对课程和管理绩效的评价。针对相关数据分析结果,为教育管理者提供更为深入的教学分析,并作为决策依据以改善现有的学校管理考核方式、课程设置等。

由于各种评价方式在具体实施中存有一定偏差和失误,为了保证教育评价的正确导向,需要对教育评价进行多元化处理,对既定的评价活动做出必要的鉴定和监控,使其更加合理、完善。大数据技术的介入使评价过程更加公开、透明、开放与灵活,通过审查、跟踪教育评价活动的每一个流程,及时调整优化评价程序,使评价结果有源可查。

三、数据驱动的教育评价实践进展

如何更好地采集和分析教育数据,并将分析评价结果应用于教学实践,教育工作者在国家、区域与学校各个层面进行了积极探索,并已取得了一些成果。

(一) 国家层面

在国家层面,基于大数据的教育评价主要用于质量认证、教育问责和学校改进,从宏观上评估教育体系的结构、效益等各要素之间的一致性,为制定国家教育政策提供一定的数据支撑和方向指引。

经济合作与发展组织(Organization for Economic Co-operation and Development)认为荷兰是教育体系最为发达的国家之一,主要得益于荷兰坚持以数据为基础开展科学化测评。荷兰的基础教育评价主体由学校、政府(教育部及其委派机构、省市两级)和非政府(第三方评价机构和国际组织)组成,包含学生评估、教师鉴定、学校评价和体制评价四个维度。^[15]

学生评估是教育评价的第一步,受学校的委托,第三方评价机构采集初等教育离校考试成绩数据、中等教育国家和校本考试成绩数据以及中小学生学习监测系统中的数据,对学生的学业成就进行客观、专业的评估,并为后续教师评价和学校评价提供数据支持。值得注意的是,荷兰的第三方评价机构——国家教育测量研究院在评估领域占据主导地位,它覆盖了85%的小学考评工作。荷兰的教师鉴定完全依赖于学校董事会的内部评价,外部评价的缺乏使得教师鉴定结果缺少一定的真实性和可靠性。学校评价主要途径有学校自评和学校督导,学校自评没有统一的标准,学校自行选择评价形式和评价方法,评价结果为学校督

导提供依据,从而助推学校的改进。体制评价是对国家教育体制的整体情况进行价值判断,其中国内评价依赖于学生评价和学校评价中积累的各项数据,通过数据分析精准反映国内教育质量,同时参与一些权威的国际评估项目(如PISA、TIMSS)进行教育质量认证,根据国际间的比较寻找差距,明确发展方向,教育督导局会将各评价报告、核心数据、发展趋势等内容公开在国家信息平台上,保证公众的知情权。其中学生评估、学校评价和体制评价都有第三方教育评价机构的介入,第三方评价的结果用于支持政府的外部问责。与荷兰不同,芬兰的国家教育评价机构——芬兰教育评价中心得出的评价结果主要用于帮助评价对象自我改进,^[16]它并不扮演督察员的角色。芬兰、荷兰开展评价的目的都是为了促进评价对象持续不断的改进,且愈加重视内部评价与外部评价相结合的方式,注重发挥各方面的作用。

(二) 区域层面

为落实党在十九大报告中提出的“发展公平而有质量的教育”,包头市于2017年正式启动了普通高中的增值评价工作,通过学生的进步程度和学校的努力程度评价学校的教育质量,每轮评价周期为三年。2018年2月,包头市根据基线测试数据已经完成了教育发展质量起点报告。目前包头市仍在稳步推进增值评价工作。

增值评价最关键的部分在于采集学生一段时间内的基线数据和最终成绩数据,基线测试数据包括学业成绩数据和学生学习态度数据两部分。就学业成绩来说,至少要包含数学或阅读成绩等核心课程科目,而包头市选取的科目涵盖语文、数学、英语等九门课程的中考成绩,因此学业数据的获取是较为理想的。在态度数据方面,全市对起点对象进行了在线问卷的调查,问卷涉及学生16个方面,如学习适应、师生关系、同伴关系等,这些数据能帮助教师了解学生在非智力因素上的差异,但是无法掌握学生的态度问题,因此最好用经典的态度问卷对起点对象进行调查,三年后可进一步计算态度增值分数。通过对基线数据的深入分析,形成起点报告、增值分析报告以及各市间的差异分析报告,教育管理者、教师等通过解读分析报告精准诊断当前的教育现状及相关问题,并有针

对性地给出可行性建议和解决对策,为了保证评价质量,包头市针对不同人员组织了各类培训会,帮助各校的教师、管理者充分挖掘起点报告价值。计算增值主要依靠两个数据,一个是基点的中考成绩,一个是终点的高考成绩,通过多水平模型、多元回归等统计方法精准估算学生的残差分数,以此了解学生的相对进步。由于未到2020年高考,因而尚未采集终点数据,当前的增值评价工作的重点主要是尝试剖析起点报告中存在的问题,并为其他地区开展增值评价提供借鉴依据。

不少人在阅读包头市教育发展质量起点报告时会产生一些“误解”,如增值是终点成绩与基点成绩的差,教师、学生、校长的背景调查数据对分析学生的进步作用不大等。由此可见,在大数据时代,要想利用评价结果数据驱动教育质量的提升,教育管理者 and 广大教师必须不断提高数据应用意识和数据解读能力,善于挖掘数据背后的价值,以更好地服务于学校改革、区域改进以及科学决策。

(三) 学校层面

在学校层面,互联网技术的迅猛发展变革了教育评价的手段,使其更加电子化、智能化,同时在评价内容方面也愈加注重对学生的综合评价。

电子答题器是衡水中学用来进行课堂测验的神器,教师只需在系统中完成试卷的编制,学生即可利用答题器进行答题,系统会及时反馈试题答案,帮助教师第一时间掌握学生的答题情况,这有效提高了教师批改作业的效率,同时也节省了学校用纸的费用。对教师而言,随着学生答题的结束,电子白板上会出现每位学生答题的时间和答题的准确率,根据学生的答题情况有针对性地去调整教学内容,将重点、难点内容着重讲解。对学生而言,提交作业或测验后,能快速查看参考答案,通过与自己作业的对比,便于自我改进和自我提高。电子答题器为学生的诊断性评价提供了极大的便利,但也存在一些问题,如系统只能批改客观题,对于主观题,衡水中学采用学生互评的形式完成,这对考试、作业不断的学生来说无疑增加了负担;对于随堂答题的功能,光靠答题总时间和答题率不一定能准确反映学生对知识的掌握情况,考虑到一些偶然性或学生对某一知识点理解模棱

两可的情况,该系统最好能提供学生完成每道测试题所用的时间,帮助教师进行精准判断与科学指导。

不同于传统基于单一学科成绩的评价,广东省佛山市盐步中心小学尝试采用“学生综合发展在线评价系统”,该评价系统的应用大大提高了评价结果的服务能力。这一系统包括品德发展水平、学业发展水平、身心发展水平、兴趣特长养成和学习负担状况5个一级指标,与《教育部关于推进中小学教育质量综合评价改革的意见》中指标完全一致,参考20项关键指标,盐步中心小学根据培养要求、地方特色等因素设置了19个二级指标,67个三级指标,并将三级指标细化为229个具体可观测的点,教师和家长根据229个观测点及时记录学生的表现,形成全方位的学生发展性评价,学生综合素质发展成长报告书将进一步服务教师、学生、家长决策。

大数据有能力让教育教学变得更科学、精准,利用智能化的评价手段、测评系统能够帮助教师厘清每位学生的问题和需要,从而提供个性化的教学策略。但由于技术的限制,当前智能工具的功能与智能程度有待进一步提高,随着对此类工具探索的越来越多,基于大数据的智能测评系统会越来越完善。

四、数据驱动的教育评价实施建议

数据驱动的教育评价为学生自我发展、教师教学反思、学校质量提升等提供基于数据分析的实证支持。大数据支持的教育评价在国内外教育实践中取得了一定成效,但依然面临着诸多挑战。为此,结合教育改革需求,本研究提出四条实施建议。

(一) 开展数据素养专项培训,提升评价者基于数据的评价能力

大数据时代,数据素养是一名合格的评价者所必须具备的技能。首先,建议政府部门组织各院校进行数据素养的网络课程学习,学习结束后需通过一定考核才能授予结业证书,考核主要涉及数据的基础知识、常用的数据分析工具以及数据处理的一般流程等理论知识;其次,第三方教育评价机构要积极提供个性化的培训服务,运用大

数据分析技术了解不同评价人员的培训需求以及在运用数据驱动评价过程中存在的问题,有针对性地提供数据收集、数据处理、数据分析、数据驱动决策等专题培训;最后,各教育相关部门可以设计一些真实的“数据驱动决策”项目,类似美国一个非营利机构 TERC 的“使用数据”项目和荷兰的“数据团队”项目,^[17]由不同的评价人员组成学习共同体,在具体的实践情境中提升数据素养。

(二) 注重教育评价数据的开放与共享,确保教育评价的公正与精准

大数据带来的价值愈加凸显,开放和共享是实现数据增值的有效途径。首先,建议政府部门建立数据开放平台,平台提供一定的数据检索和数据导航服务,^[18]帮助公众快速获取科学数据。为了便于公众进一步处理数据,也可对部分数据(不涉及隐私、机密)提供下载权限。政府在开放数据建设过程中要确保评价主体及评价相关人员参与其中,大众与政府统一步调有助于政府更便捷地进行教育治理,同时监督政府决策的透明公正。因此,政府应积极利用网络平台收集公众对开放数据的需求以及教育决策的建议。其次,国家可以发起“数据共享计划”,各院校、区域根据已有的教育评价数据制定整体的数据使用规划,包括具体的评价数据指标、详细的数据目录,针对不同评价目标提供选用数据、数据维护的方法等,公众依据数据规划能更有计划地重用数据并进行精准评价。

(三) 建立大数据智能测评分析系统,促进教育评价效率的提升

海量数据是机器智能的核心要素,^[19]人工智能技术的发展使得教育服务的智能化程度不断提高。一方面,大数据智能测评分析系统能与第三方录播系统深度对接,全面采集结构化数据和非结构化数据,对评价主体的发展过程进行全方位分析;另一方面,该系统集多类型评测管理于一体,如学生综合素质评价、教师增值评价、学校效能评价等,将不同的评价指标进行分类管理,系统根据不同的评价标准灵活组合评价指标,不同主体使用该系统时只要完成评价信息的填写,系统就会自动生成可视化评测报告,评测报告不仅包含基本的描述性统计,还有对评价主体的质性分

析、预测和关联分析,系统根据评价结果会智能提供个性化的服务,如根据学习者的学习情况提供知识图谱,根据教师的增值数据提供专业发展路径规划等。

(四) 倡导增值性评价,完善数据驱动教育评价体系

增值性评价是一种相对更客观、公平的评价模式,也是新时期教育评价中重要的方式。首先,教育局应联合外包公司开发增值性评价分析系统,该系统能准确测量每个学生的基础认知能力,同时监测学生在教育过程中的“增值”情况,系统基于基线数据分析和增值数据分析形成测评分析报告并将其反馈给学生、教师、学校、政府等不同主体,通过建立科学的增值数据应用体系,确保增值数据得到切实有效的应用,如学生根据进步情况反思自身的学习过程,学校根据全体学生的整体进步情况评估教育质量,政府则按照增值指标进行学校的问责等;其次,增值性评价分析的数据来源于多次有效的测验,^[20]各地区在进行增值性评价时建议与高校、研究所、教育机构等开展合作,通过专业机构完成测验的设计、后期的数据分析处理,以保障评价的规范性和合理性,当各地区具备独立实施增值性评价的能力时,则可逐步脱离专业机构的帮助;最后,建议学校和有关教育部门在实施教育评价时参考美国俄亥俄州的评价方式,将增值性评价指标纳入已有的学生评价或学校评价体系中,注重考察学生的进步程度、学校的努力程度以及家长的教育投入度等。^[21]

参考文献:

- [1] 习近平全国教育大会重要讲话金句速览[EB/OL]. [2019-04-23]. <http://politics.people.com.cn/n1/2018/0910/c1001-30284629.html>.
- [2] 张曦,王磊,叶赛君.数据互联互通助力教育管理向教育治理转变——以宁波教育GIS系统开发、功能及应用为例[J].浙江教育技术,2018(3):12-16+32.
- [3] 张岩.“互联网+教育”理念及模式探析[J].中国高教研究,2016(2):70-73.
- [4] 肖远军.目标导向课程评价模式探析[J].教育科学,2002(3):22-25.
- [5] 肖远军.教育评价原理及应用[M].杭州:浙江大学出版社,2004:34-35.
- [6] 周谦.教育评价与统计[M].北京:科学出版社,2002:190-191.
- [7] 邱均平,王碧云,汤建民.教育评价学[M].北京:科学出版社,

- 2016: 6-7.
- [8] 邱均平, 柴雯, 马力. 大数据环境对科学评价的影响研究[J]. 情报学报, 2017, 36(9): 871-877.
- [9] 邢蓓蓓, 杨现民, 李勤生. 教育大数据的来源与采集技术[J]. 现代教育技术, 2016, 26(8): 14-21.
- [10] 杜瑛. 价值多元化视野下的我国高等教育评价改革思考[J]. 教师教育研究, 2008(5): 19-24.
- [11] 刘尧. 中国教育评价发展现状与趋势评论[J]. 中国地质大学学报(社会科学版), 2003(5): 59-62.
- [12] 上超望, 韩梦, 杨梅. 基于大数据的在线学习过程性评价设计研究[J]. 现代教育技术, 2018, 28(10): 94-99.
- [13] 郭光亮, 朱德全. 教育评价发展的多元路径探析[J]. 中国高校科技, 2018(11): 66-69.
- [14] 赵蔚, 张赛男, 裴晓杰. “动机、过程、效果”三位一体式网络学习评价体系构建研究[J]. 现代远程教育, 2012(4): 53-58.
- [15] 尤铮. 质量保障视野下荷兰基础教育评价体系研究[J]. 比较教育研究, 2018, 40(9): 103-110.
- [16] 丁瑞常. 芬兰教育评价中心: 社会第三方参与教育评价的新模式[J]. 比较教育研究, 2017, 39(7): 56-62.
- [17] 李青, 任一姝. 教师数据素养能力模型及发展策略研究[J]. 开放教育研究, 2016, 22(6): 65-73.
- [18] 武彤. 基于数据生命周期的美国研究图书馆科学数据开放共享服务研究[J]. 图书与情报, 2019(1): 135-144.
- [19] 李扬. 人工智能发展趋势: 融合平台、智能大脑、情感计算[J]. 智能机器人, 2017(2): 27-28+46.
- [20] 边玉芳, 孙丽萍. 教师增值性评价的进展及在我国应用的建议[J]. 教师教育研究, 2015, 27(1): 88-95+112.
- [21] 教育部关于推进中小学教育质量综合评价改革的意见[EB/OL]. [2019-05-23]. http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/s7054/201306/xxgk_153185.html.

The Architecture and Practice of Data-driven Education Evaluation System in the Internet + Era

YANG Xianmin¹, GU Jiani¹, XING Beibei²

(1. School of Wisdom Education, Jiangsu Normal University, Xuzhou 221116, China;

2. School of Chemistry & Materials Science, Jiangsu Normal University, Xuzhou 221116, China)

Abstract: In the era of “Internet +”, educational evaluation has shifted from the enlightenment period, the creation period, the critical period and the specialization period to the period of precision. Compared with the traditional education evaluation, data-driven education evaluation has a great change in evaluation function, evaluation object, evaluation content, evaluation method, evaluation data source, data analysis, evaluation process, and evaluation result, etc. Six features could be summarized as follows: value-added, diverse, comprehensive, real-time, accurate and fair. Combined with the preparation stage of the education evaluation process (evaluation purpose, evaluation object, evaluation target, evaluation index system, evaluation standard, evaluation method), implementation stage (evaluation data collection, evaluation data processing, evaluation data analysis) and decision-making stage (evaluation result generation, evaluation result feedback), a data-driven educational valuation framework has been constructed and the progress of educational evaluation has been analyzed from the national, regional and school levels respectively. In order to promote the development of data-driven education evaluation, four suggestions were put forward: carrying out special training on data literacy to improve the data-based evaluation ability of evaluators; paying attention to the opening and sharing of educational evaluation data to ensure the fairness and precision of educational evaluation; establishing big data intelligence evaluation and analysis system to promote the efficiency of educational evaluation; advocating the concept of value-added evaluation to perfect the data-driven education evaluation system.

Key words: Internet +; education evaluation; data-driven education evaluation; framework; progress in practice

(责任编辑 张丽珍)