

发展教育大数据：内涵、价值和挑战*

□杨现民 唐斯斯 李冀红

摘要：我国教育领域的发展与改革正面临前所未有的挑战，大数据与教育的结合成为时代发展的必然要求。教育大数据是指整个教育活动过程中所产生的以及根据教育需要采集到的，一切用于教育发展并可创造巨大潜在价值的集合。在当前国际形势下，教育大数据从战略高度应定位为推动教育变革的新型战略资产、推进教育领域综合改革的科学力量以及发展智慧教育的基石。教育大数据的最终价值应体现在与教育主流业务的深度融合以及持续推动教育系统的智慧化变革上，具体表现在驱动教育管理科学化、驱动教学模式改革、驱动个性化学习真正实现、驱动教育评价体系重构、驱动科学研究范式转型、驱动教育服务更具人性化。然而由于教育系统自身更强的独特性和复杂性，我国教育大数据的发展也面临应用落地、数据安全、数据治理与运营等诸多挑战。只有正视并深入研究这些问题，才能推动我国教育大数据持续健康发展。

关键词：教育大数据；教育改革；教育信息化；内涵界定；战略定位；价值潜能；现实挑战

中图分类号：G434 文献标识码：A 文章编号：1009-5195(2016)01-0050-12 doi:10.3969/j.issn.1009-5195.2016.01.007

*基金项目：国家社会科学基金教育学青年课题“开放环境下学习资源进化机制设计与应用研究”（CCA130134）。

作者简介：杨现民，博士，副教授，江苏师范大学教育研究院，江苏省教育信息化工程技术研究中心（江苏徐州 221116）；唐斯斯（通讯作者），副研究员，国家信息中心（北京 100045）；李冀红，硕士研究生，江苏师范大学教育研究院（江苏徐州 221116）。

2012年，联合国在发布的《大数据促发展：挑战与机遇》白皮书中指出：“大数据时代已经到来，大数据的出现将会对社会各个领域产生深刻影响。”大数据正在实现人类工作、生活与思维的大变革，其“威力”也强烈地冲击着整个教育系统，正在成为推动教育系统创新与变革的颠覆性力量。教育领域已有学者探讨了大数据时代背景下教育模式的转变（杨森淇等，2014；郑立海，2015）、教育变革的趋势与转向（喻长志，2013；胡弼成等，2015）、教育研究的新范式（祝智庭等，2013）、学习方式的变革（徐鹏等，2013）等，还有学者探讨了教育领域具有广泛应用价值的大数据技术，比如学习分析（顾小清等，2012；Shum et al., 2012）、教育数据挖掘（葛道凯等，2012；Winne et al., 2013）等，也有学者探讨了大数据技术在教育中的应用模式与实践案例（杨现民等，2015）。

毫无疑问，大数据理念正在逐步深入人心，大

数据技术正在快速融入各行各业，行业大数据呈现出不同的特征、不同的数据采集方式、不同的发展趋势以及应用模式。我国教育领域的发展与改革正面临前所未有的挑战，大数据与教育的结合已是时代发展的必然要求。然而，目前教育界对教育大数据还没有形成清晰的认识，有一些关键问题亟需澄清，比如教育大数据究竟是什么，有何独特性？需要采集哪些教育数据？如何进行教育大数据的战略定位？教育大数据的价值如何体现？我国的教育大数据发展面临哪些现实挑战？基于此，本文将着重对我国教育大数据发展面临的几个关键且基础性问题进行探讨，期望能促使同行研究者形成共识，更好地应用大数据深化我国教育领域的综合改革。

一、教育大数据的内涵界定

1. 教育大数据的概念

大数据是一个新生事物，目前尚处在逐渐被认

识、被应用的初始阶段，还未形成公认的定义。其产生之初是一个IT行业的技术术语，被定义为所涉及的数据量规模巨大到无法通过人工在合理时间内达到截取、管理、处理并整理成为人类所能解读的信息（Wikipedia: Big Data, 2014）。大数据的核心特征常被概括为“4V”，即数据量大（Volume，一般认为在T级或P级以上）、输入和处理速度快（Velocity）、数据多样（Variety）和精确性（Veracity）（刘雍潜等，2014）。大数据技术几乎在所有领域都拥有非常广阔的应用前景，通过对海量数据进行模型构建，有利于挖掘事物的变化规律，准确预测事物发展趋势，并进行及时有效地干预。

随着大数据理念的传播及其应用的逐步深入，大数据的内涵也在不断变化和拓展。大数据不仅仅是一种技术，还是一种能力，即从海量复杂的数据中寻找有意义关联、挖掘事物变化规律、准确预测事物发展趋势的能力。大数据更是一种思维方式，即让数据开口说话，让数据成为人类思考问题、做出行为决策的基本出发点。

教育大数据是大数据的一个子集，特指教育领域的大数据。文献调研发现，目前还未有学者对教育大数据进行明确的概念界定。所谓教育大数据，是指整个教育活动过程中所产生的以及根据教育需要采集到的，一切用于教育发展并可创造巨大潜在价值的数据集。

首先，这里的教育是“大教育”的概念，具有全员（从全日制学生到全民，面向所有人）、全程（从学前教育到终身教育，服务各个教育阶段）、全方位（家庭、学校、社会“三位一体”教育，无处不在的教育，虚实融合的教育）的特点。

其次，教育大数据有四大来源：一是在教学活动中直接产生的数据，比如课堂教学、考试测评、网络互动等；二是在教育管理活动中采集到的数据，比如学生的家庭信息、学生的健康体检信息、教职工基础信息、学校基本信息、财物信息、设备资产信息等；三是在科学研究活动中采集到的数据，比如论文发表、科研设备运行、科研材料采购与消耗等记录信息；四是在校园生活中产生的数据，比如餐饮消费、上机上网、复印资料、健身洗浴等记录信息。

再次，教育大数据要能服务教育发展，具有教育目的性，而非盲目地囊括一切数据。教育活动过程中也会产生大量无意义的噪声数据（Noisy Data），因此需要根据教育应用目的进行数据过滤与整理，为后期深度挖掘与分析做好准备。

最后，教育大数据之“大”并非指数量之大，而是强调“价值”之大，即能从繁杂的教育数据中发现相关关系、诊断现存问题、预测发展趋势，发挥教育大数据在提升教育质量、促进教育公平、实现个性化学习、优化教育资源配置、辅助教育科学决策等方面的重要作用。

2. 教育大数据的独特性

与传统教育数据相比，教育大数据的采集具有更强的实时性、连贯性、全面性和自然性，分析处理更加复杂多样，应用更加多元深入。

传统教育数据的采集往往是阶段性的，多在用户知情的情况下（非自然状态）进行，分析的手段多采用简单的汇总统计和比较分析，关注的重点是受教育者的群体特征以及国家、区域、学校不同层面教育发展的整体状况。在大数据时代，移动通信、云计算、传感器、普适计算等新技术将逐步融入教育的全过程，可以在不影响师生教学活动的情况下实时、持续地采集更多微观的教与学的过程性数据，比如学生的学习轨迹、在每道作业题上逗留的时间、教师课堂提问与微笑的次数等。教育大数据的数据结构更加混杂，常规的结构化数据（如成绩、学籍、就业率、出勤记录等）依旧重要，但非结构化数据（如图片、视频、教案、教学软件、学习游戏等）将越来越占据主导地位。

与电子商务、交通、医疗、金融保险等领域的大数据相比，教育大数据的独特性表现为三个方面：

第一，教育大数据的采集呈现高度的复杂性。教育活动是人类社会中一种特殊的实践活动，主客体关系复杂、不稳定，教育过程呈现复合结构（教的活动与学的活动并存）（顾建军，2000）。教育业务复杂，无标准化的操作流程和模式，创新人才的培养又需要更多元化、创新性的教学模式与方法。由于缺少商业领域标准化的业务流程以及学习方式的多样性和学习地点的不确定性，导致教育大数据的采集将变得异常复杂。

第二，教育大数据的应用需要高度的创造性。大数据在重塑教育方面具有无限的潜能，而潜能的发挥需要打破数据分析与应用的常规思维，发挥更多的创造性。当前我国教育发展面临公平、质量、减负、择校等一系列重大现实难题，直接影响人民群众对教育的满意度。教育关乎国计民生，而教育问题又异常复杂，需要一大批教育大数据研究者与实践者，充分发挥其创造性，将数据挖掘、学习分析、人工智能、可视化等先进技术与教育现实问题相结合，方可破解当前教育发展之难题。

第三，不仅注重相关关系，更要强调因果关系。国际大数据专家维克托·迈尔·舍恩伯格博士认为，大数据时代的一个最重要转变便是从因果关系转向相关关系，不再需要从事实中寻求原因，而要从看似无关的数据中发现某种相关关系。对于商业领域而言，注重相关关系的挖掘或许可以更迅速、更直接地达成预期目标。然而，教育以培养人为根本目的，不仅要“知其然”，更要“知其所以然”。唯有洞察到教育问题产生的本质原因，才可能从根本上寻求解决之道。

二、教育大数据的分类与结构

随着“三通两平台”建设、教学点数字教育资源全覆盖、中小学教师信息技术应用能力提升、精品开放课程建设等一系列国家信息化工程的大力开展，以及全国各地数字校园、区域教育云的建设与完善，为教育数据的采集提供了便利条件。此外，国内在线教育市场规模不断壮大，传统教育培训企业和互联网企业纷纷瞄准在线教育，推出多样化的在线教育产品和服务，为学校外非正规教育数据的采集提供了基础。教育数据每时每刻都在产生，然而教育领域究竟包含哪些数据？如何采集？不同的数据又指向何种教育应用？这些问题的解决是教育大数据建设与应用的关键。

1. 教育大数据的分类

教育数据有多种分类方式。从数据产生的业务来源来看，有教学类数据、管理类数据、科研类数据以及服务类数据。从数据结构化的程度来看，包括结构化数据、半结构化数据和非结构化数据。结构化数据适合用二维表存储，图片、视频、文档等

非结构数据则不适合用二维表存储。从数据产生的环节来看，还可以分为过程性数据和结果性数据。过程性数据是活动过程中采集到的、难以直接量化的数据（如课堂互动、在线作业、网络搜索等），结果性数据则常表现为某种可量化的结果（如成绩、等级、数量等）。

多年来，国家采集的教育数据主要以管理类、结构化、结果性的数据为主，重点关注宏观层面教育发展整体状况，在一定历史时期对我国制定教育政策、推动教育发展起到了积极的作用。然而，随着大数据时代到来，国际社会对教育大数据作为战略资产的地位越来越认可和重视，教育数据的全面化采集与深度挖掘分析就变得越来越重要。教育数据采集的重心将向非结构化、过程性的数据转变，这些数据无论从数量和增长速度上，还是潜在的价值上，都将远远超越传统的教育数据。

2. 教育大数据的结构模型

为了更清晰地认识教育大数据的概貌，笔者构建了如下图所示的教育大数据结构模型。整体来说，教育数据可以分为四层，由内到外分别是基础层、状态层、资源层和行为层。其中，基础层存储国家教育基础性数据，包括教育部2012年发布的7个教育管理信息系列标准中提到的所有数据，比如学校管理信息、行政管理信息、教育统计信息等；状态层存储各种教育装备、教育环境以及教育业务的运行状态信息，比如设备的能耗、故障、运行时间、校园空气质量、教室光照、教学进程等；资源层存储教育过程建设或生成的各种形态的教学资源，比如PPT课件、微课、教学视频、图片、游戏、教学软件、帖子、问题、试题试卷等；行为层

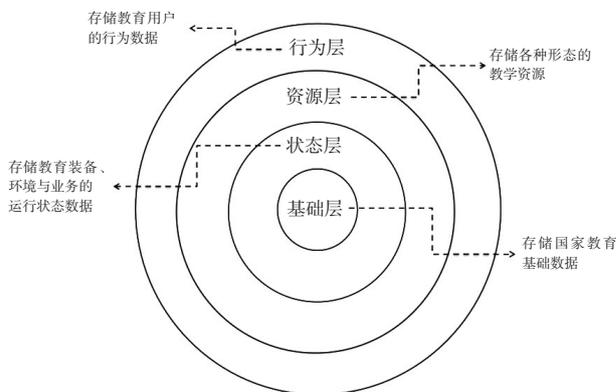


图 教育大数据的结构模型

存储广大教育相关用户（教师、学生、教研员、教育管理者等）的行为数据，比如学生的学习行为数据、教师的教学行为数据、教研员的教学指导行为数据、管理员的系统维护行为数据等。

不同层教育数据的主要采集与生成方式、应用场景也有所不同（见下表）。数据采集的难度从内向外逐步增加，尤其是行为层数据的采集最为复杂多变，对于传统的面授教学或者不使用技术工具的学习行为，很多数据仍无法采集。

表 教育大数据采集方式与应用场景比较

数据层次	数据采集方式	数据应用场景
基础层	人工采集、数据交换	宏观掌控教育发展现状、科学制定教育政策、合理配置教育资源、完善教育体系等
状态层	人工记录、传感器感知	教育装备的智能管理、教育环境的智能优化、教育业务的实时监控等
资源层	专门建设、动态生成	各种形式的教学与培训，如课堂教学、教师培训、网络探究学习、移动学习、协作学习等
行为层	日志记录、情境感知	个性化学习、发展性评价、学习路径推送、教学行为预测等

基础层数据。一方面通过定期地人工采集实现国家规定的教育基础数据的逐级上报，如每年的招生、教师招聘等新产生的教育数据；另一方面通过系统之间的数据交换实现教育基础数据采集与更新，比如学籍系统、人事系统、资产系统等定期进行自下而上的系统数据更新。基础层数据属于高度结构化的教育数据，主要用于宏观掌控教育发展现状、科学制定教育政策、合理配置教育资源、持续完善教育体系等。其中，有些基础数据（如学籍、人事、资产等）具有高度的隐私性和保密性，属于国家重点保护的教育数据。

状态层数据。采集方式有人工记录和传感器感知，当前主要以人工记录为主，随着传感技术的普及应用，未来的教育装备、教育环境以及教育业务的运行状态将实现全天候、全自动化的记录监控。状态层数据可用于高效管理与维护教育装备，打造更具人性化的教育环境，全面掌控各项教育业务运行状况等。

资源层数据。总量巨大，形态多样，大都属于非结构化数据。资源的产生主要有两种途径：一是专门建设，包括国家组织的精品开放课程资源、企业自主开发的各种学习培训资源与工具、个体自发

建设的教学课件等；二是动态生成，在教学过程中产生的各种生成性资源（如讨论、试题、笔记等）。海量优质的资源是实现教学模式创新和教学方法变革的基础，比如当前基于微课的翻转课堂、基于MOOC的开放创新教学、基于电子书的移动学习等。随着移动与开放教育浪潮的兴起，微课、电子书、APP应用、慕课等将成为未来重要的学习资源。

行为层数据。教育行为有很多种，比如收发公文、录入成绩、设备报修、财务报销、教师备课、学生写作业等。其中，教学行为数据（包括教师的教和学生的学）在所有行为层数据中占据主导地位。大数据时代可以采集更多、更细微的教学行为数据，比如学生在何时何地应用何种终端浏览了哪些视频课件、观看了多长时间、先后浏览顺序、是否跳跃观看等细颗粒度的行为都将以日志记录的形式被保存下来。GPS定位、情境感知、移动通信等技术使得各种教与学行为的日志信息更加丰富，不仅仅可以记录什么人在什么时间什么地点做了什么，还可以采集到行为发生时周边的环境信息、个人体征信息、情绪状态等。这些看似无用的数据都将成为后期数据挖掘与学习行为分析的宝贵数据源，为个性化学习、发展性评价、学习路径推送、教学行为预测等提供数据支持。

三、教育大数据的价值潜能

2015年8月31日国务院发布《促进大数据发展行动纲要》，文件指出“数据已成为国家基础性战略资源”，并在启动的十大工程之一“公共服务大数据工程”中明确提出要建设教育文化大数据。由此可见，教育大数据的重要性已经上升到国家战略层面，引起社会各界的广泛关注和高度重视。

1. 教育大数据的战略定位

大数据时代，教育数据的价值正在被广大教育者重新认识和评估。教育数据不再仅仅是一堆用作统计的简单“数字”，其正在成为一种变革教育的战略资产和科学力量。

(1) 教育大数据是一种教育战略资产

随着大数据理念在全球的发酵，以美国、英国、法国等为代表的发达国家率先将大数据作为新型战略资源，视其为“未来的石油”。数据作为战

略资产的观念被越来越多的国家所认可，国际社会纷纷通过“公共数据开放”运动激发数据活力，以期创造更大价值。

理论上讲，任何领域有了人的活动，都可以持续不断地产生大数据，教育领域也不例外。随着全球教育信息化的快速发展，教育数据正在以几何级的规模递增。以一节40分钟的普通中学课堂为例，其中一个学生所产生的全息数据约有5~6GB，而其中可归类、标签并进行分析的量化数据约有50~60MB（张韞，2013）。除了传统学校教育产生的数据外，互联网教育市场每天也在产生海量的教育数据。目前我国教育规模位居世界首位，仅在校生就有2.6亿（人民网，2013）。如此大规模的教育，必将产生世界量级的教育大数据，而如何发挥这笔“资产”的价值则成为我国教育赶超欧美的关键。

教育大数据是一种无形的资产，是一座可无限开采的“金矿”，充分的挖掘与应用是实现数据“资产”增值的唯一途径。西方发达国家已经先行一步，我国也应该加速部署教育大数据战略，强化教育大数据战略资产意识，让每个人都成为教育数据的缔造者和受惠者，并顺应数据开放趋势，通过教育大数据的适度公开，汇聚广大民众、企业、政府等多方智慧，使教育数据资产实现源源不断的增值。

（2）教育大数据是教育领域综合改革的科学力量

当前我国教育还不完全适应国家经济社会发展和人民群众接受良好教育期盼的要求，存在一系列发展难题，比如：中小学生课业负担过重，素质教育推进困难；学生创造力不足；城乡之间、区域之间教育发展不均衡；教育公平问题长期存在；高等教育规模飞跃式扩张导致本科教学质量下滑；各地校园安全事件频发等等。改革是解决教育发展难题的唯一途径，党的十八届三中全会提出要“深化教育领域综合改革”，将促进教育公平和提升教育质量、考试招生制度改革和教育管理体制改革作为改革的重点任务（顾明远，2014）。

虽然我国教育改革的攻坚方向和重点举措已经明确，但是如何科学、有序、有效地全面推进改革则成为亟需解决的关键问题。教育改革是复杂的系统工程，需要综合考虑经济、文化、社会

等因素，而大数据最擅长的就是关联分析，即从繁杂的交叉领域数据中寻求有意义的关联。大数据是一股创新的力量，是一股时代变革的力量，也是一股推动教育领域全面深化改革的科学力量。因此，确立教育大数据的战略地位已是教育领域综合改革的必然要求。

改革既要有胆魄，更需要科学的依据。教育大数据将汇聚无数以前看不到、采集不到、不重视的数据，对这些混杂数据进行深度挖掘以及与其他领域（如公安、交通、社保、医疗等）的大数据进行关联分析。教育决策将不再过度依靠经验、拍脑袋和简单的统计结果，而转向基于数据的科学决策。招考制度、管理体制以及教育公平与质量提升，无论是宏观的制度与体制改革，还是微观的教学方法和管理方式的改革，都可以通过科学的数据分析寻找问题的症结所在，识别不同地区教育发展的独有规律，然后对症下药，实施改革。

（3）教育大数据是发展智慧教育的基石

世界范围内的教育信息化建设正在走向融合创新的深层次发展阶段。在物联网、云计算、大数据、移动通信等新一代信息技术的推动下，世界上多个国家和地区已将智慧教育作为其未来教育发展的重大战略。新加坡在《iN2015计划》中提出实施智慧教育战略计划，韩国于2011年颁布了“智慧教育推进战略”的国家教育政策，美国在2010年发布的《国家教育技术计划》虽未提到智慧教育，但其倡导信息技术支持下教育系统的全方位、整体性变革的理念与智慧教育不谋而合。技术变革教育的时代已经到来，从数字化教育转向智慧教育正在成为全球教育发展的重要趋势。

智慧教育是依托物联网、云计算、无线通信、大数据等新一代信息技术所打造的物联化、智能化、感知化、泛在化的教育生态系统，是数字化教育的高级发展阶段（杨现民，2014）。各种智能型技术是构建智慧教育“大厦”的技术支柱，其中大数据是实现教育智慧化的灵魂所在。教育大数据汇聚存储了教育领域的信息资产，是发展智慧教育最重要的基础（柯清超，2013）。教育大数据将促进教育发生几个方面的重要转变：其一，教育过程从“非量化”到“可量化”，教与学的行为信息将越来越

精确地被记录下来；其二，教育决策从“经验化”到“科学化”，数据驱动的决策将变得越来越可靠；其三，教育模式从“大众化”到“个性化”，学习分析技术将赋予教师认识每个“真实”学生的能力，实现因材施教；其四，教育管理从“不可见”到“可视化”，通过可视化技术将实现更直观、更准确、更高效的教育资源与业务管理。

2. 教育大数据的价值体现

在政府、企业、学校、研究机构、行业协会等社会力量的推动下，大数据在社会各行各业战略价值正在逐步凸显。教育领域的研究者和实践者也在积极探寻大数据技术与教育最适合的结合点和实施方式。教育大数据的最终价值应体现在与教育主流业务的深度融合以及持续推动教育系统的智慧化变革上。

(1) 教育大数据驱动教育管理的科学化

当前我国的教育管理信息化仍属于“人管、电控”的管理模式，智能化程度不高，管理水平和效率有待提升（杨现民等，2014）。随着国家教育管理公共服务平台的建设与运营，我国教育数据的采集工作将越来越规范化、有序化和全面化。大数据时代，教育数据的分析将走向深层次挖掘，既注重相关关系的识别，又强调因果关系的确定。大数据技术能够从海量的教育数据中发现隐藏的、有用的信息，反映教育系统中实际存在的问题，从而为做好教育管理和决策工作提供科学的数据支持。大数据在教育管理业务中的应用价值主要体现在三个方面：一是教育的科学决策，二是教育设备与环境的智能管控，三是教育危机预防与安全管理。

大数据除了可以对各级各类教育单位的人员信息、教育经费、学校办学条件、运维服务管理等数据进行图表式的统计与分析外，还可以基于各级各类教育机构长期的数据积累，整合社会人口分布、经济发展、地理环境等从各类跨行业操作级的应用系统中提取有用的数据，通过数据统计、指标展现、横向对比、趋势分析、钻取转换等技术方法将数据转化为知识，为各级管理人员的科学决策提供数据支持。美国政府早在2002年就通过立法的形式确定了教育数据在支持教育科学决策方面的重要地位。纵观我国十几年的新课

程改革历程，虽然在课程内容、教学方法、教学环境等方面取得了进步，但实际的改革效果远未达到预期状态。其中的要因之一便是忽视了教育数据在课程改革诸多决策上的重要性，使改革更趋向于理性思辨和经验决策。

可以采取如下措施促使教育大数据驱动教育管理的科学化：通过设置全方位的传感器，对教育管理过程中的教学活动、人员信息、学校资产及办学条件等数据进行采集、汇总、挖掘与分析，并对数据分析结果进行可视化处理。根据各级各类教育管理机构的需要，建立自上而下的教育管理和调度指令发布功能。对各级各类教育管理机构所需的各方面信息与数据、资产设备、教学活动、企业运维服务管理数据等进行远程可视化质量监控与管理。通过对教育设备的智能化管控，实现设备的科学使用，降低能耗和管理负担，节约开支。比如江南大学建设了“校园级”智能能源监管平台，该平台通过物联网、通信、信息、控制、检测等前沿技术智能化监管能源，将原来能源管理过程中的“模糊”概念变成清晰数据，为管理者提供更好、更科学的决策支持，打造低碳绿色校园（新华网，2010）。

近年来校园安全问题已成为社会关注的热点。通过对传感设备所采集的数据以及信息系统所汇聚的数据进行实时监控与对比分析，可以对校园的安全运行状况进行预警，以提前预防、妥当处理教育危机，提高教育安全管理水平。此外，大数据在提升学校网络安全（北京民教信息科学研究院，2013）、改善教学和科研管理（宓咏等，2013）、完善学生救助体系（张越，2013）、促进区域教育均衡发展（刘雍潜等，2014）等方面也有极大的应用价值。

(2) 教育大数据驱动教学模式的改革

通过应用大数据技术对海量教学数据进行分析与预测，将改变传统千篇一律的教学模式，有利于真正实现个性化教育。以翻转课堂、MOOCs等为代表的新型教学模式的成功开展，离不开大数据的支持。通过对学生学习历程记录的分析，教师能够快速、准确掌握每位学生的兴趣点、知识缺陷等，从而为设计更加灵活多样、更具针对性的学习活动提供数据支持。传统预设的固化课堂教学将转变为动态生成的个性化教学。在大数据的支持下，教师

能够更好地认识自己和学生，以不断改进其教学模式与策略，并且在学生进行自我导向学习时，真正变成学生学习的促进者与协作者。利用大数据技术可以对教师进行全面考核，跟踪教师成长过程，还可以运用回归分析、关联规则挖掘等方法帮助教师分析教学方法和手段的有效性，使教师及时调整教学方案，优化教学方法，提高教学质量。

近年来应用大数据技术改善课程教学日渐流行。美国奥兰治县的马鞍山社区学院通过“高等教育个性化服务助理”系统，运用学生数据成功实施了个性化教育。该软件为每个学生建立详细档案，记录其完整的在校期间的日程信息、跟随导师学习的经历以及其他个人信息；接着对这些信息进行分析，提出对时间管理、课程选择的建议，以及分析其他有助于学生在学业上获得成功的要素（陈律，2013）。美国普渡大学（Purdue University）早在2007年就启动了“课程信号项目”（Pistilli et al., 2010；马红亮等，2014），通过将数据从学生信息系统、课程管理系统和课程成绩单中提取出来，按照学习表现进行分类，并利用数据分析技术对那些极有可能不及格或辍学的学生提供针对性的辅导服务。

大数据技术的突出优势在于其拥有强大的“预判”能力，从商业计划到选举结果，大数据发挥着越来越精准的预言能力。随着教育大数据应用的逐步深入，预测性分析在消除教育不确定性、提供提前干预方面潜力巨大，越来越多的公司推出了基于数据驱动的个性化学习和教学工具。如著名的个性化教育服务公司 Desire2Learn 利用其分析平台“学生成功系统”（Student Success System，以下简称“S3”），可以提前几个月预测学生的期末考试成绩，甚至能精确到小数点后两位数字。学生可以利用 Desire2Learn 的学习管理平台阅读课程材料、提交家庭作业、做练习、提问题、交流互动等，所有这些活动数据将自动存储下来。基于上述原始学习过程数据，“S3”可以预测学生是否适合一些新的课程，从而辅助学生做出选课决定。“S3”具有多样化的模型管理、预测和数据可视化能力，教师不仅可以查看整个班级表现的报表，并能像传统联机分析工具 OLAP 那样，对教学报表进行钻取，筛选学生或课程的详细信息，以检视学生的能力、参与

度和评估分数（移动 LABS，2013）。此外，“S3”综合应用多种预测模型，能够基于每门课程来准确预测每位学习者的课程学习情况，辅助教学人员提供恰当的教学介入，提高学生课程学习的成功率。

（3）教育大数据驱动个性化学习的真正实现

大数据将使得教师和机器能够真正了解每个学生的真实情况，从而为其提供真正个性化的学习资源、学习活动、学习路径、学习工具与服务等。网络学习虽然具有天然的“个性化”优势，然而缺少大数据的支持，机器将无法真正了解每位学习者，也就无法实现个性化资源与服务的推送。如果说互联网促进了教育的民主化，那么大数据将实现教育的个性化（翟博，2006），而教育个性化的首要体现便是学习的个性化。

当前学习管理系统正在向智慧型学习平台发展。通过集成教育数据挖掘与学习分析技术，能够持续采集学习者的学习行为数据并进行智能分析，如依据学习者模型推送适合的学习资源，进行个性化的学习评价，提供准确的诊断结果，给出适合学习者的个性化学习建议。大数据技术使得学习行为的记录更加精细化，可以准确记录到每位用户使用学习资源的过程细节，如点击资源的时间点、停留了多长时间、答对了多少道题、资源的回访率等信息。这些过程数据一方面可用于精准分析学习资源的质量，进而优化学习资源的设计与开发；另一方面，学生可以对自己某一段时期内的学习情况（包括学习爱好、业余活动等非结构化的学习行为）进行分析和预测，以便尽早通过这些预测做出最适合自身发展的决策，更好地开展适应性学习和自我导向学习（Johnson et al., 2012）。

事实上，无论是研发中的还是市场上推广应用的各种适应性学习系统，要实现为不同学习者提供个性化学习服务的目的，都依赖于大量学习过程数据的采集与深度分析。例如，美国纽约一家在线教育公司 Knewton，其主打产品是适应性在线学习系统，核心技术是适配学习技术，通过数据收集、推断及建议“三部曲”来为学习者提供个性化学习建议（Knewton，2014）。其中，数据收集阶段会建立学习内容中不同概念的关联，然后将类别、学习目标与学生互动集成起来，再由模型计算引擎对数据进

行处理。推断阶段会通过心理测试引擎、策略引擎及反馈引擎对收集到的数据进行分析，分析的结果将提供给建议阶段进行个性化学习推荐使用。建议阶段则通过建议引擎、预测性分析引擎为教师与学生提供学习建议并提供统一汇总的学习历史。

(4) 教育大数据驱动教育评价体系重构

随着大数据时代的到来，教育评价正在从“经验主义”走向“数据主义”，从“宏观群体”评价走向“微观个体”评价，从“单一评价”走向“综合评价”。

智慧学习环境中通过新一代信息技术可以采集到教与学的全过程数据，不仅仅包括网络教学平台上记录的档案数据，还包括更多学习情境数据（如地点、时间、个体特征、所用设备、周围环境等），为开展中小学学业成就评价提供了更全面的数据支持。每个学生拥有相伴一生的学习档案袋，持续存储每个学期、每门课程、每节课、每次学习的表现数据。每个教师拥有一个教学档案袋，全面记录每个学期、每门课程的教学表现。基于云计算技术，将档案袋数据永久存储在云端，同时通过科学的评估模型，对教师和学生的发展进行定期评估，提出更具针对性的发展建议。学校不仅要对学生在校期间的学业成就进行评价，还要持续跟踪学生毕业后的发展情况，以期为学校教学质量评估提供更全面、更准确的科学数据。

目前，我国教育部已经建立了较为完整的国家基础教育质量数据库和多级数据采集网络。全面客观地记录学生成长轨迹，沉淀和积累多维度的学生成长数据，让反映学生发展状态的数据完整显示出来，将能引导学生培养模式和教育质量管理方式的科学健康发展。上海作为试点区域于2011年开始建立了“上海市中小学生学业质量绿色指标”体系，“绿色指标”在收集学生学业水平数据的基础上，还收集了有关学生家庭背景、学习动机、学业负担和师生关系以及教学方式和校长领导力等信息，并及时向区县和部分学校反馈评价结果，以合理引导教育管理、教学指导以及各种教与学行为。

现阶段已有不少地区尝试将基于大数据的学习评价方式应用于教学中。如田纳西州增值评价系统（Tennessee Value Added Assessment System，

TVAAS）是一个通过对学生的成绩进行多年连续追踪分析来评估学区、学校和教师效能的系统。在TVAAS中，3~12年级的每个学生都要参加一系列的测试，如语言、数学、科学等，该系统采用增值评价方法分析每个学生的学业进步，并依此评估各区、各校、各教师对学生学业进步的贡献大小（Ballou et al., 2004）。TVAAS可以为教育决策者提供大量的诊断信息，有利于促进形成性评价的开展与实施；可以计算每所学校或学校子群体在各科上的成绩进步率，然后参照以往的进步率，发现那些没有取得足够进步的学生群体，并对其进行干预；可以预测每个学生在各科上将来可能获得的成绩，有助于学校管理者与教师尽早发现那些将来有可能会达不到毕业要求的学生，使他们有足够的时间为这些学生制定不同的课程和教学策略；还可以帮助教师提前了解那些即将在他们班就读学生以前的成绩，有助于教师根据学生的学习情况制定更加适合本班学生的教学方案，从而提高教师的教学质量。

(5) 教育大数据驱动科学研究范式转型

随着大数据技术的不断成熟，各种科研数据的获取将更加便利，传统科研的复杂性将大大降低，也在一定程度上破解了科研经费投入、数据分析以及科研管理等难题，为科研工作者提供了更为便捷的技术支撑与人性化服务，大大提高了研究的效率和结果的可信度。

在自然科学领域，大数据为发现更多“意外”的结果提供了可能，使传统研究中很多的不可能逐步成为可能。美国的大数据研究计划中专门列出寻找“希格斯粒子”（被称为“上帝粒子”）的大型强子对撞机实验（李国杰等，2012），实验目标是至少要在1万亿个事例中才可能找出1个希格斯粒子。2012年7月4日，终于发现新的玻色子，标准差为4.9，被认为可能是希格斯玻色子（承认是希格斯玻色子粒子需要5个标准差，即99.99943%的可能性是对的）。大数据处理是整个实验的一个预定步骤，为实验的成功提供了数据保障。

在社会科学领域，由于实验设施及人员精力的限制，研究者总是通过人工搜集、资料查找等方式获得科研数据，并根据对“抽样”数据的分析，推算一般规律。这种研究范式导致社会科学研究一直

备受逻辑性和科学性不足的诟病。随着大数据的出现，越来越多的社会科学变得可以被量化，社会科学的研究范式正在从抽样模式走向全样本模式，成为一门实实在在的实证科学（北大新媒体，2013）。

此外，大数据技术还可以为科研人员便捷地获取个性化的学术文献资源、自动寻找研究同行和合作伙伴、组建跨学科跨地域的国际研究团队提供支持和便利。

（6）教育大数据驱动教育服务更具人性化

教育大数据是一笔难以估量的潜在的巨额财富，而数据挖掘技术与学习分析技术的发展将实现这笔财富的无限增值。教育数据的合理、合法、有效、创新应用正在催生越来越多样化、智慧化的教育服务。

上海闵行区教育局依托大数据进行数字化校园建设，全面客观地记录学生成长轨迹，沉淀和积累多维度的学生成长数据，让反映学生发展状态的数据完整显示出来，以推进教育质量观的转变，引导学生培养模式和教育质量管理方式的科学发展（吉安教育网，2014）。为了全面搜集数据，学校为每一位学生发放了电子学生证，学生的日常行为数据将被动态抓取和实时记录，并能在系统内自动生成各种数据统计图表，从而让学生隐性的状态和需求显性化，让家长和老师能够更直观地看到学生的点滴进步或潜在的问题。比如：学生早晨上学是从哪个校门进入，中午吃了什么，选了哪些兴趣课，有没有去图书馆，喜欢看哪些课外书？这些情况学生家长从家里的IPTV数字电视中都能看到。

百度教育上线了包含1336所院校信息的高考院校库，首期院校信息已基本覆盖全国“985”和“211”等重点高校，通过搜索数据分析与资源整合，利用大数据与自然语言算法将搜索数据与个性化需求相匹配，可以为考生在估分、报考、专业选择上提供帮助和指导。除了填报高考志愿，大数据在促进就业方面也大有用武之地。在美国，只要点击进入“Where are the jobs?”网站（<http://www.where-are-the-jobs.com/app.php>），就可以一目了然地看到美国各领域的就业人员数和平均工资。该网站是根据美国劳工部开放就业数据开发的一个应用系统，利用美国政府手中掌握的海量教育数据进行

分析和挖掘，并最终可视化的图表进行形象的展现，让公众与政府官员都能看得懂，从而为其提供自身决策需要的有关依据。这个过程的关键是海量教育数据以机器可读的方式完全对公众开放，数据资源被“激活”，教育数据服务社会的价值被很好地体现出来。

四、教育大数据面临的挑战

与商业、交通、环境、医疗等领域相比，教育领域具有更强的独特性和复杂性，大数据技术在教育领域的应用推广仍存在诸多难题。当前，我国教育大数据的发展主要面临应用落地、数据安全、数据治理与运营等诸多挑战。

1. 教育大数据的应用问题

当前大数据的理念正在被各行各业所接受，广大教育工作者已充分认识到大数据在推动教育决策科学化、实现个性化教育、加快区域教育均衡发展、提升教育质量等方面具有重大的应用价值。国内在经济困难学生预警、大学报考难度预测、学校资产智能管理等方面已经出现了一些教育大数据应用案例。然而，当前的教育大数据应用还比较零散，处于起步探索阶段。大数据在教育领域究竟该如何全面“落地”，有无可推广的成熟应用模式，仍是困扰教育界的一大难题。

我国应尽快组织教育专家和大数据专家等制定《教育大数据发展建议与应用指南》，从国家层面加大对教育大数据应用和推广的支持力度，将教育大数据应用提升到更高的战略层面；提炼总结若干典型、有推广价值的教育大数据应用模式与案例，明确教育大数据应用思路，引导全国各地教育行政机构、学校、教育培训企业等合理应用教育大数据；借鉴美国教育部推动教育数据应用方面的经验（美国教育部教育规划、评价与政策办公室从2006年开始每年都发布一个关于学校数据应用的报告），定期发布教育领域数据应用的研究报告，以系统有序推进教育数据应用的深入发展。此外，还应组织教育学、管理学、计算机科学、统计学等多学科的研究人员成立专门的教育大数据研究机构，集中优势力量破解教育大数据应用推广过程中存在的热点、难点问题，同时结合教育发展的战略需求，开

展前瞻性研究，通过研究指导引领教育大数据应用快速健康发展。

2. 教育大数据的安全问题

教育数据既是一笔宝贵的教育资产，同时也涉及到教育者和受教育者的隐私，保护不当则会带来严重的安全事故。美国的教育大数据存储机构 inBloom 仅仅运行了 15 个月便关闭，主要原因便是教育数据开放过程中导致了安全问题。公共数据开放运动已是国际趋势，教育作为一项公益性事业，教育大数据也应适度向社会开放。国家应在推进教育数据开放运动的同时，高度重视教育数据的隐私保护与安全管理，不断努力采取更先进、安全系数更高的措施来保障教育数据安全，保护教育隐私数据不外泄、不被恶意使用。

我国需加快制定《教育大数据安全管理办法》，从体制、机制、技术、方法等多个层面制定管理细则，切实保障个体、机构、国家的教育数据安全。建立健全教育大数据安全管理架构，包括数据生产部门、数据使用部门、数据管理部门等，并明确各部门的安全管理职责。建立教育数据的保密等级，按保密等级规定采取相关处理措施。加快研制教育专用的大数据存储系统，提供高性能、高可靠、高灵活性、易管理的教育数据存储服务。实时监控教育大数据的应用状况，对非法应用、侵犯用户隐私与国家安全的单位和个人进行处罚。

3. 教育大数据的治理问题

信息时代每时每刻都在产生海量的、各种来源、多种类型的教育数据，如何协同多方力量进行高效的教育数据治理是摆在每位教育管理者面前的现实问题。教育数据治理的目的是提升教育数据质量，保护教育数据隐私安全，保障教育数据合理应用，促进教育数据合法共享。

各级教育管理者应加强“数据治理”理念，发挥政府、学校、企业、研究机构、公众等多主体在发展教育大数据上各自的优势。我国应尽快出台教育大数据治理的相关办法，建立完整的教育大数据治理模式，指导教育数据的获取、归档、保存、互换以及重复利用。制定教育基础数据采集和质量管理标准，形成清晰的数据治理机制、流程与质量管理办法。确立教育数据集的归档和长期保存的机制

与方法，将教育数据库、教育服务平台、教育资源平台等产生的数据进行规范化采集与汇聚共享，形成全国一体化的教育大数据中心。建立教育大数据开放平台，发挥企业、个体、教育机构等主体的创造性，开发大量特色化的教育应用，让社会力量共同参与到教育数据的治理与创新应用中。

4. 教育大数据的运营问题

教育数据是一种无形的国家资产，其所有权、使用权属于广大民众。因此，教育大数据应当受到国家保护，同时也应适度向民众开放。面向社会合理开放教育数据，能够鼓励、促使更多的研究机构、企业以及个人加入到挖掘教育数据价值的队伍中，提升教育数据的价值，增加公众对教育的认知，促进教育的改革发展。

从国家层面来看，教育数据的安全性不亚于金融数据。教育数据的开放程度、开放范围、开放对象都应进行深入论证，以确保教育数据的使用是合法的，是有利于教育创新发展的，而非阻碍、破坏、危害到国家安全。为此，政府部门应对教育数据加强监管，可参考通信领域颁发运营商牌照的方式，制定教育数据运营商准入标准，颁发运营牌照。非官方机构在使用教育数据进行研究时，需要满足相关规定并取得相关许可，才能获得教育数据的使用权利。

参考文献：

- [1]北大新媒体(2013). 大数据助力社会科学研究[EB/OL]. [2015-12-10]. <http://www.loooker.com/archives/6765>.
- [2]北京民教信息科学研究院(2013). 大数据技术如何处理大学校园 IT 安全问题[EB/OL]. [2015-12-10]. http://www.cnein.ac.cn/html/special/2013/0423/109.html?pc_hash=0xvnuG.
- [3]陈律(2013). 大数据背景下学习分析技术对教学模式的变革[J]. 中国教育信息化, (24): 15-17.
- [4]葛道凯, 张少刚, 魏顺平(2012). 教育数据挖掘: 方法与应用[M]. 北京: 教育科学出版社.
- [5]顾建军(2000). 浅析教育的双主体性特征[J]. 教育科学, (1): 1-5.
- [6]顾明远(2014). 教育领域综合改革的宏观视野[J]. 教育研究, (6): 4-9.
- [7]顾小清, 张进良, 蔡慧英(2012). 学习分析: 正在浮现中的数据技术[J]. 远程教育杂志, (1): 18-25.

- [8]胡弼成,王祖霖(2015).“大数据”对教育的作用、挑战及教育变革趋势——大数据时代教育变革的最新研究进展综述[J].现代大学教育,(4):98-104.
- [9]吉安教育网(2014).大数据时代教育管理谋求信息化变革[EB/OL].[2015-12-10].<http://www.jaedu.com/Article/ShowInfo.asp?InfoID=10122>.
- [10]柯清超(2013).大数据与智慧教育[J].中国教育信息化,(24):8-11.
- [11]李国杰,程学旗(2012).大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J].中国科学院院刊,(6):647-657.
- [12]刘雍潜,杨现民(2014).大数据时代区域教育均衡发展新思路[J].电化教育研究,(5):11-14.
- [13]马红亮,[英]袁莉,郭唯一等(2014).反省分析技术在教育领域中的应用[J].现代远程教育研究,(4):39-46.
- [14]宓詠,赵泽宇(2013).大数据创新智慧校园服务[J].中国教育信息化,(24):3-7.
- [15]人民网(2013).教育部:目前我国教育规模位居世界首位2.6亿学生在校[EB/OL].[2015-12-10].<http://politics.people.com.cn/n/2013/1015/c1001-23206346.html>.
- [16]新华网(2010).江南大学全国首创“校园级”智能能源监管[EB/OL].[2015-12-10].http://wx.xinhuanet.com/2010-12/25/content_21716917.htm.
- [17]徐鹏,王以宁,刘艳华等(2013).大数据视角分析学习变革——美国《通过教育数据挖掘和学习分析促进教与学》报告解读及启示[J].远程教育杂志,(6):11-17.
- [18]杨淼淇,孙纳新,柴华(2014).大数据时代教育模式的研究[J].计算机工程与科学,(S1):272-273.
- [19]杨现民(2014).信息时代智慧教育的内涵与特征[J].中国电化教育,(1):29-34.
- [20]杨现民,王榴卉,唐斯斯(2015).教育大数据的应用模式与政策建议[J].电化教育研究,(9):54-61,69.
- [21]杨现民,余胜泉(2014).论我国数字化教育的转型升级[J].教育研究,(5):113-120.
- [22]移动 LABS(2013).学习管理平台 Desire2Learn 野心不小,想改变大学生学习生涯规划方式[EB/OL].[2015-12-10].<http://labs.chinamobile.com/news/94091>.
- [23]喻长志(2013).大数据时代教育的可能转向[J].江淮论坛,(4):188-192.
- [24]翟博(2006).教育均衡发展:理论、指标及测算方法[J].教育研究,(3):16-28.
- [25]张越(2013).华东师大:让教学生活更智慧[J].中国信息化,(16):44-45.
- [26]张韞(2013).大数据改变教育[J].上海教育,(4):8-11.
- [27]郑立海(2015).大数据时代的教育管理模式变革刍议[J].中国电化教育,(7):32-36.
- [28]祝智庭,沈德梅(2013).基于大数据的教育技术研究新范式[J].电化教育研究,(10):5-13.
- [29]Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for Student Background in Value Added Assessment of Teachers [J]. Journal of Educational and Behavioral Statistics, 29(1): 37-65.
- [30]Johnson, L., Adams, S., & Cummins, M. (2012).The NMC Horizon Report: 2012 Higher Education Edition[R]. Austin, Texas: The New Media Consortium.
- [31]Knewton[EB/OL].[2014-07-01].<http://www.knewton.com/>.
- [32]Pistilli, M. D., & Arnold, K. E. (2010). In Practice: Purdue Signals: Mining Real-Time Academic Data to Enhance Student Success[J]. About Campus, 15(3): 22-24.
- [33]Shum, S. B., & Ferguson, R. (2012). Social Learning Analytics[J]. Educational Technology & Society, 15(3): 3-26.
- [34]Wikipedia: Big Data[EB/OL].[2014-07-13].http://en.wikipedia.org/wiki/Big_data.
- [35]Winne, P. H., & Baker, R. S. J. (2013). The Potentials of Educational Data Mining for Researching Metacognition, Motivation and Self-Regulated Learning[J]. Journal of Educational Data Mining, 5(1): 1-8.

收稿日期 2015-12-04 责任编辑 刘选

The Definition, Potential Value and Challenges of Big Data in Education

Yang Xianmin, Tang Sisi, Li Jihong

Abstract: The education development and reform in China faces unprecedented actual challenges. The integration of big data and education has become an irresistible trend in this new era. Big Data in Education (BDE) is a large data set used for educational development with huge potential value in which all data are produced in educational activities, or collected according to educational needs. In a global context, BDE should be viewed as a new kind of strategic asset to promote education innovation, a scientific power to promote comprehensive educational reform, and the foundation of smart education. The ultimate value of BDE is reflected in its deep integration with educational businesses and in the promotion of intelligent education reform, including making educational management scientific, driving instructional reform, achieving truly individualized learning, promoting the reconstruction of educational evaluation system, driving the transformation of scientific research paradigm, and making educational services more human-oriented. However, since our educational system is peculiar and complex, the development of our national BDE still faces numerous challenges, such as the application, the data security, the data governance and operation. Therefore, only by facing these problems clearly, as well as conducting deep research, can we promote the sustainable and healthy development of BDE.

Keywords: Big Data in Education; Educational Reform; Education Informatization; Definition; Strategic Position; Potential Value; Actual Challenges

(上接第 32 页)

right relies on the development of lifelong education system, and the construction of a learning society. Citizens' learning right is a logical starting point for lifelong educational theory, and the attention paid to the realization of the learning right reflects the rethink of the traditional education theory and the conformity to the international trend of building a learning society. And its property contains endogenous freedom and external trait of social rights. Therefore, to lay stress on its free property in an isolated and vague way while ignoring its dependence on the society will inevitably fall into a passive embarrassment. In the implementation, there is a discrete dynamic balance of confuse and separation between the learning right and the traditional sense of "the right to receive education". The realization of learning right should start with the consciousness introspection and gradually moves towards rational return in the conflicts of outside rights claims and to be shaped, in order to agree with the system. With the strengthening of civil consciousness, to give the majority the right to appeal maximum learning in the most equitable way is the ought-to-be choice for the formulation and implementation of education policy.

Keywords: Lifelong Education; Citizens' Learning Right; Theoretical Context; Path to Achieve; Learning Society