

大数据时代高教研究的特征与新文化^{*}

郑宏

摘要: 大数据时代高等教育研究已经或即将呈现一些新的特点,主要表现为大数据的发现与收集将成为高等教育研究的基础;大数据分析能力将提升高等教育研究的质量和科学性;大数据利用能力是高等教育研究能够服务实践的关键。这些特征意味着我们需要倡导并培育一种基于大数据的高等教育研究新文化:重数据而不唯数据,见“数”又见“人”;重视相关关系和因果关系;安全使用数据,注重数据伦理;形成一种开放、共享、合作的研究模式,从而增强研究的科学性,促进高等教育质量的提升。

关键词: 大数据时代; 高等教育研究; 特征; 新文化

当很多人尚在迷惑信息时代究竟是什么的时候,大数据时代已经翩然而至。2011年6月麦肯锡全球研究院发布题为《大数据:下一个创新、竞争和生产力的前沿》的研究报告,在全世界宣告“大数据时代已经到来”。大数据在教育领域更是吸引了全球的关注。美国新媒体联盟(New Media Consortium)和美国高校教育信息化协会(EDUCAUSE Learning Initiative)联合发布的2013年地平线报告(高等教育版)^[1],以国际教育信息化为主题,指出未来5年将影响全球教育的6种新兴技术,即:学习分析技术、大规模网络课程、平板电脑、3D打印技术、游戏及游戏化和可穿戴技术,其中的学习分析技术和大规模网络课程就与大数据紧密相关。2013年英国将大数据作为国家的战略之一,指出大数据对经济发展和科技创新意义重大,要求所有高校将数据能力(Data Capability)作为大学生的核心能力之一,并研究如何结合不同学科培养学生的这种能力。2015年8月31日,中国国务院发布《促进大数据发展行动纲要》,指出“数据已成为国家基础性战略资源”,应该“完善教育管理公共服务平台,推动教育基础数据的伴随式收集和全国互通共享”^[2]。可见,大数据已经渗透到世界教育系统的方方面面,

与此同时,各学科领域的学者也纷纷开始探讨大数据所带来的影响,以及本学科研究范式与话语体系可能发生的深刻变革,如语言学、新闻传播学、政治学、经济学、心理学、社会学、管理学、营销学和历史学等学科都将大数据的理念引入进来。

但是在这股热潮中,高等教育研究似乎相对滞后。2016年12月15日,在中国引文数据库(期刊)以“big data”展开中英文论文扩展搜索,对10236篇文献进行学科分析,结果“高等教育”仅有219篇,勉强名列前20位。而这些高等教育学科的论文,大多是分析大数据给高等教育带来的影响而不是给高等教育研究带来的变化,“总体来看,教育学中运用大数据方法进行研究的成熟案例还不多”^[3]。因此,本文的着力点在于分析大数据时代,到底高等教育研究将呈现哪些新的特征,这些特征又要求我们培育一种怎样的高等教育研究新文化。

一、大数据时代高等教育研究的新特征

大数据时代,大数据与大数据分析被提升到了前所未有的高度。人们通常把大数据的特征归纳为4个V,即数据数量的规模性(volume)、数据种类的多样性(variety)、数据处理的高速性(velocity)

收稿日期: 2017-06-14

作者简介: 郑宏,厦门大学教师发展中心、麦可思-厦门大学高等教育大数据中心副教授,教育学博士。(福建厦门/361005)

^{*} 本文系国家社会科学基金“十三五”规划2016年度教育学重点委托课题“高等教育人才培养模式的国际比较研究”(WGA160006)的阶段性成果之一。

以及数据隐藏的价值 (value), 这些特征为高等教育研究带来了新的机遇, 也提出了新的挑战。中央电化教育馆王晓茛副馆长说“教育正在走向大数据时代, 谁能够发现数据, 谁就能够赢得未来的生存; 谁能够挖掘数据, 谁就能够赢得未来的发展; 谁能够利用数据, 并利用数据提供个性化的服务, 谁就能够赢得未来的竞争。”^[4]可以说, 正在走向大数据时代的高等教育研究, 将呈现新的特点。

(一) 大数据发现与收集将成为高等教育研究的基础

大数据时代, “让数据说话 (Data Tell the Truth)” 是最明显的特征, 数据为王。除了定性研究, 传统的抽样定量研究将逐渐被大数据研究所取代, 大数据的发现与收集将成为高等教育研究的基础。“基于大数据的研究突破了以样本推断总体的传统测量统计方法, 直接对总体进行分析, 并更注重结论的相关性和实时性。”^[5]相对于传统的定量研究, 大数据研究不存在样本, 总体就是样本。为什么高等教育领域目前缺乏大数据的研究成果, 一个原因是大数据不足, 没有数据想做大数据研究只能是“无源之水、无本之木”, 当然数据不仅包括数字, 视频、文本、图像、音频都是可以记录的数据。

仅从数字数据来说, 目前有几个主要的来源: 其一是来自政府层面的, 比如教育部每年对高等教育的统计, 即《高等教育学校 (机构) 统计报表》, 统计的是每所高校培养的各类学生数、师资数、固定资产量等; 学科建设、科研成果等没有纳入这一套报表。其二, 各省市及地方教育主管部门会根据不同的职能, 定期或不定期要求高校报送各种统计数据。其三, 有些高校自己因为战略决策需要或评估等需要, 统计自己高校的某些数据。比如厦门大学自 2005 年开始, 学校教务处参照教育部本科教学评估的做法, 坚持开展年度教学评估, 形成了常态化的教学质量“年检”制度。结合师资情况分析、学生情况统计、教学质量报告、课程教学评价、学生学习经历调查、教师发展中心建设等多种手段, 学校不仅建立了系统的可持续的动态的教学质量监测与评估机制, 还积累了厦门大学有关教学状态的一系列数据。第四, 目前有一些数据中心、科研机构、评估中心和咨询公司也在收集教育数据, 如麦可思数据有限公司比较早就开始进行专门的教育数据收集, 教育部高等教育质量评估中心也正在进行

高等教育质量监测国家数据平台的建设。

除了传统的数据收集方式, 现在大量的网络数据源源不断且数量庞大, 是今后重要的高等教育研究资源。而且这些网络大数据, 正在使一些用传统方法无法进行的研究选题成为可能。比如以前研究大学办学水平, 主要是考察教授数量、生源质量、校舍面积、科研经费、科研项目多寡、论文数量等硬指标, 对大学社会声誉、大学满意度等软性指标的衡量相对缺乏, 原因一方面是由于大学软实力的体现往往是隐性的、还没有直接生成的数据, 只能通过问卷、访谈等形式进行采集, 但是要对中国几千所高校进行这样的数据采集几乎不可能; 同时中国高校类型多样、数量众多且分布地域广泛, 不同社会群体对高校关注的情况千差万别, 调查样本难以选择, 但是如今大数据为进行大学的社会声誉研究提供了可能性。截至 2016 年 12 月, 我国网民规模达 7.31 亿^[6], 网民在互联网上对大学的检索、评论、关注使得大学社会声誉的隐性数据变为显性数据, 研究者不再需要通过调查问卷等传统方式去临时生成关于大学社会声誉的数据, 而只需要采用大数据手段直接采集互联网上已经生成的现成数据即可。^[7]

(二) 大数据分析能力将提升高等教育研究的质量和科学性

显然, 大数据时代对研究者的数据处理能力提出了更高要求。高等教育研究者获取数据之后, 需要快速高效地将原始数据转换成可行并且有用的数据, 以便更好地进行科学研究。所以, 高等教育研究者首先必须具有大数据思维, 能够去伪存真, 清理数据、管理数据、合并数据、分析数据并使数据可视化, 最后得出结论, 这是大数据研究者应该具备的能力, 否则, 只能“望据兴叹”, 无能为力。当然, 这并不是说每个研究者都必须在一夜之间成为大数据专家, 这既不可能也不现实。有了大数据思维, 研究者不能完成的一些纯技术工作可以交给专业的技术人员去操作, 包括运用计算机进行运算分析等等, 但是大数据思维是每个研究者都必须具备的。所谓大数据思维, 就是“一切皆用数据来观察, 一切都用数据来刻画, 人们以数据的眼光来观察、思考、解释这个纷繁复杂的世界”^[8]。

没有大数据分析能力和大数据思维, 数据背后隐藏的价值将无从体现, 所以, 我们应该认识到, 大数据时代对研究者的理论素养要求并没有降低,

从某种角度来说,反而是提高了,理论素养本身就是大数据思维的一部分。首先,研究问题是什么,需要收集哪些数据,如确定变量或参数,是研究者根据自己的理论假设或观察到的现实事先确定的,当然可能在研究过程中还会发生改变;其次,即使是基于大数据的研究,也需要研究者具有一种对高等教育现象的深刻洞察力,以及对高等教育基于政治、经济、历史与文化体验的深度诠释能力,同时还需要具备理论的表达与建构能力。大数据研究与传统的定性研究、案例研究并不矛盾,大数据不是告诉我们高等教育如何运作,而仅仅是呈现给我们信息,如何对数据进行解读与挖掘,还需要具有高等教育知识结构和理论背景的研究者的分析与推理,特别是大数据当中大量的非结构性材料,难以被计算机处理,需要研究者质性的解读,并把它们转化成结构性数据。如果不具备一定的理论素养,研究者不仅找不到真正的具有学术价值的高等教育问题,还很可能迷失在海量的数据中不知所云。比如,现在高等教育普及化逐渐进入人们的视野,那么普及化量的规定性跟以前提出普及化理论时有没有改变?是否还是沿用老的标准?大规模的网络课程是否会对普及化量的规定性产生影响?这些问题在分析数据时都必须考虑周全。数据终究是死的,只有人的解释和分析才能揭示数据之间的关系,建立模型,并进一步分析变量之间的因果关系,使得这些数据具有说服力和解释力。

传统的高等教育定性研究和“抽样”的定量研究,一直被人诟病逻辑性和科学性不足,大数据的出现,高等教育研究逐渐从抽样模式走向全样本模式,这为高等教育研究走向科学性奠定了坚实的基础。

(三) 大数据利用能力是高等教育研究服务实践的关键

通常高等教育研究的目的有两个:一是理论目的,通过研究促进高等教育理论的发展和认识的深化;二是实践目的,通过研究揭示高等教育发展的趋势,提出预测或预警,为政策制定或改革实践提供指导。大数据时代的研究当然也要促进高等教育理论的发展,但是通过研究助力宏观和微观的高等教育决策更加科学,促进高等教育改革的顺利进行,始终是研究者不可推卸的社会责任。因此,利用大数据研究,对未来做出预测,就成为高等教育研究能否服务实践并发挥智库功能的关键。

大数据的突出价值之一即是对于事物发展的预测功能。如前所述,大数据分析是一种全样本的分析,而不是通过小样本的统计推论来对整体进行预测,所以大数据研究注重的是事物发展的先后顺序,而不是事物的因果联系,并由此进行较为准确的预测。我们不能保证大数据一定能够揭示教育规律,但是不可否认,大规模的数据的确使我们更加逼近事实,发现更多的规律性或者高等教育的发展趋势,从而为我们预测未来和科学决策奠定基础。这种预测可以是宏观的也可以是微观的,宏观的比如中国高等教育大众化和普及化,有没有什么特殊性?哪些高等教育要素会发生重大变化,我们应该如何应对?微观的比如对大学生的学习分析。“大数据可以支持对学习者的个性发展的研究,数据的分析可以提供给我们关于每一个学习者的学习需求、学习风格、学习态度乃至学习模式等信息,因此我们可以相应的提供适合不同学习者发展的学习内容和学习指导,促进其个性发展从而实现真正意义上的个性化教育。”^[9]

各个国家在启动大数据教育研究时都不约而同地指向实践的需求,突出表现在大学教学和院校管理两方面。如美国学校管理者协会 AASA 携手学校网络联合会 COSN 以及全球性的信息技术研究和咨询公司 Gartner,共同实施了一个名为 Closing the Gap: Turning Data into Action 的项目,旨在促进学校对学生信息系统和学习管理系统中大数据的使用;美国教育部(U. S. Department of Education)在2012年10月发布了《通过教育数据挖掘和学习分析促进教与学》报告,其直接目的就是面向教学实践;两位美国学者曾经对约1000门在线课程及约6000名学生进行大数据分析,发现在某个时间段学生表现不佳或者处于放弃的边缘,由此分析怎么样的学生群体较容易处于这种危险之中,从而对学生进行适时的帮助,改变教学策略,降低辍学的比例。关于院校管理,别敦荣教授的看法具有代表性。他认为院校研究中的数据分析,不仅可以帮助院校研究人员和管理者获取关于研究对象的一般认知,还可以获取对研究对象的本质认知和未来认知,从而帮助大学的决策者从传统的主观决策、经验决策和指令决策走向理性决策和科学管理。^[10]我国《促进大数据发展行动纲要》明确要求“探索发挥大数据对变革教育方式、促进教育公平、提升教育质量的支撑作用”^[11]。高等教育研究应该切实运用大数据,改

善高校的管理,促进教学改革和教学质量的提高,而不是将大数据研究成果束之高阁。

二、大数据时代高等教育研究的新文化

大数据时代高等教育研究呈现或即将呈现的这些新特点,意味着掌握数据、分析数据和利用数据的能力将成为每个高等教育研究者应该具备的数据素养,我们急需倡导并培育一种基于大数据的高等教育研究新文化。

数据文化是指一个教育组织或系统内部崇尚数据对于各个层面决策过程的重要性的学习环境,它包括相应的价值观、态度和行为准则。^[12]布鲁金斯学会的一篇论文指出:大数据在教育决策中的使用日益增多,而数据驱动决策(data-driven decisions)确实能使学校做出的决策更合理,更让人信服。^[13]我们已经进入一个“数据驱动决策”的时代,高等教育研究者要用自己的大数据研究,使学校管理者认识到数据会让他们的管理决策更加科学高效;给教师们的教学实践带来帮助,促使整个高等教育系统产生对数据的信任感,形成一种基于数据的科学决策文化。对高等教育研究来说,我们需要培育一种在人的理性的指引下合理运用技术的价值观、态度和方法,也就是一种高等教育研究的新文化,主要包括四个方面。

第一,重数据而不唯数据,见“数”又见“人”。大数据时代我们强调用数据说话,实质就是强调实事求是,重视研究的客观性和科学性,但是与此同时,我们也必须意识到大数据是我们认识高等教育的一种工具,它不能代替研究者的理性思考。我们更不能见“数”不见“人”,这是教育学与自然科学的不同,因为高等教育的目的是培养人,所以当我们解读教育数据时,首先不能只把它理解为“冷冰冰”的数据,教育数据反映的其实是“活生生”的人的信息。大数据对于人的感觉、情感、经验、体验和创造性也具有收集和分析能力(如多媒体数据和行为数据),这让我们触摸到大数据的温度和质感,依靠这些数据进行研究,最终目的是帮助我们理解人的培养过程,掌握人才的成长规律。教育与人永远是不可分的。其次,我们还要看到个体的人,不能只见“森林”(大数据)不见“树木”(个体)。大数据让我们更好地掌握研究对象(教育者、受教育者、管理者等等)的共性和发展趋势,但是同时这些研究对象是一个个有自身特点的人,

我们要利用大数据,发现个体的差异性,促进个性化的教育改革,真正实现管理和教育的科学性和人性化。最后,我们还要看到这些个体是动态发展变化的,而不是一成不变的,因为收集的数据往往反映的是个体的过去,而且由于大数据保存的便利性,很可能某个个体几年前甚至更长时间的信息都得以保存,所以当我们审视这些数据时,要意识到该个体的情况已经发生了变化,必须以发展的眼光来分析这些数据,同时考虑人的巨大潜能,否则就可能得出僵化或片面的结论。

第二,重视相关关系也探究因果关系。不可否认,大数据揭示的相关关系帮助我们相对准确地了解现象并预测高等教育的发展动态,这是以前的定性研究或样本研究难以做到的,但是相关关系只是因果关系的基础,只有弄清相关关系背后的因果关系才能真正把握高等教育规律。高等教育与社会政治、经济、文化存在错综复杂的关系,高等教育内在系统本身也充满复杂性和动态性,所以大数据呈现出来的高等教育领域的相关关系也是多种多样的,有些可能互相矛盾甚至完全颠覆常识,这个时候因果关系的探讨就显得特别重要,否则就可能导致结论的简单化和片面化。我们“需要深度理解大数据分析的局限性,以避免陷入数据产生歧义之陷阱”^[14]。况且高等教育领域几乎不存在单一原因导致单一结果的简单关系,所以去粗取精、去伪存真的分析至关重要。数据并不直接导向决策,只有同时揭示相关关系和因果关系,才能保证高等教育研究的质量和科学性,有效地将大数据的预测应用到高等教育实践,服务国家高等教育决策,这将是我們长期关注的课题。

第三,安全使用大数据,注重数据伦理。这是所有学术研究都要注意的问题,但大数据时代的数据伦理问题愈加凸显。高等教育数据是宝贵的教育财富,但也不可避免地涉及管理者、教育者和受教育者的隐私,所以合理安全地使用教育数据,保护个人隐私是不可回避的社会伦理问题。目前很多国家都出台了保护数据隐私的相关法律,欧盟也推出了《个人数据保护》(Personal Data Protection),这些法规一方面希望开放数据服务于公共事业,另一方面保护个人对自身数据的控制。中国目前还缺乏相关法规或条例。另外“在教育领域运用大数据,至关重要的是必须在互信的环境中,尤其要得到学生和家长的信任”^[15]。这里强调的不仅是让学生和

家长允许研究者收集数据,更重要的是根据大数据研究出来的结果要能够让学生和家长信服。比如根据大数据预测某学生适合某专业或某职业,很可能与学生或家长的预期不一样,那么家长会不会认为这剥夺了孩子的未来呢?孩子会不会认为这违背了他/她的兴趣和理想呢?尤其当我们要用大数据的结果去实施一项重大高等教育改革的时候,更要谨慎行事。

第四,形成一种开放、共享、合作的研究模式。长期以来,我们的高等教育研究注重描述性和思辨性研究,如比较教育研究领域,“在基于证据的研究的时代显得方法上偏‘软’,证据不够充分,结论不太客观,同时,与国际社会在研究范式上的这种差异,也使我国的比较教育研究成果很难得到国际同行的承认,难以与国际同行在同一个话语体系中对话”^[16]。可以说,大数据为我们提供了一个契机,可以与国际同行分享数据,共同研究。“比较教育学之父”朱利安很早就提出了获取、整理、对比和分析各国教育统计数据的一系列设想,由于当时条件所限无法得以实现,所以以前的比较高等教育研究主要集中于某国或某几国的资料,或者仅仅是某个高校的数据,现在依托分布存储和云计算,高等教育研究者将不再受地域和国家的限制,通过互联网就能够在世界上任何一个地方迅速地获取整个高等教育领域和相关领域的海量数据,使得真正的跨国研究和高等教育比较研究成为可能。共享大数据将使我们具有更加广阔的国际视野,探索世界高等教育的规律,而不是仅仅局限于中国特色的研究。

世界很多发达国家已经意识到共享大数据、促进研究深入的重要性,开始一系列数据开放的行动。如英国分别于2013/2014和2014/2015年度为100余所大学额外投资三千七百万英镑建立开放的科学研究数据平台^[17];欧洲委员会、美国政府、澳大利亚政府成立了研究数据联盟;欧洲委员会出台开放数据战略以及2020地平线开放数据计划。中国应该顺应这一趋势,积极主动地开放资源,寻找更多的国际合作研究伙伴。

我们知道,大数据的魅力除了数据之海量,更在于它将多学科、跨领域的数据结合在一起,因此突破了传统方法的限制,实现了研究的创新。大数据科学本身就横跨多门学科,如信息科学、网络科学、统计学、管理学、心理学、伦理学等等。大数据时代到来,社会、经济和科技发展一体化的程度

越来越高,一方面,社会科学各学科之间,社会科学与自然科学之间不断交叉、融合;另一方面,许多学科越分越细、新学科不断涌现,这就要求不同学科、不同领域之间在方法和视角上相互借鉴,加强协作。近年来,各个国家在开始培养大数据人才时都几乎采取了跨学科的方式。据不完全统计,2012年全世界有近170所大学开设了大数据相关专业,美国各大学的大数据硕士课程从2011、2012年开始大量招生,这些课程大多数是与商业管理、社会学、新闻学等学科结合起来培养大数据人才,有专家称之为数据科学家,而不仅仅是单纯的数据技术人才。如哈佛大学在2008年推出数字人文先导计划,2010年再进一步成立“人文2.0”实验室,也是面向校内的人文院系,发展数字化的研究工具、建立讨论平台或是提出跨领域的合作计划。^[18]

因此,大数据时代的到来,高等教育研究必须主动寻求跨学科的合作,联合信息科学、网络科学、统计学、伦理学和管理学等相关领域的学者,夯实大数据基础理论,切实提高研究者数据收集、数据分析、数据挖掘、数据应用的能力。我国已出台政策“鼓励高校设立数据科学和数据工程相关专业,重点培养专业化数据工程师等大数据专业人才,鼓励采取跨校联合培养等方式开展跨学科大数据综合型人才培养,大力培养具有统计分析、计算机技术、经济管理等多学科知识的跨界复合型人才”^[19]。跨学科的合作,不仅可以培养高等教育大数据人才,还有可能创生新的交叉学科,就像“计算社会学”日趋流行一样,有学者预言,假以时日,也许会诞生一门“计量高等教育学”,开辟一个全新的高等教育研究方向。

三、结语

大数据正在深刻地影响和改变着高等教育的发展,也对研究者的理论与实践带来新的挑战与机遇。大数据研究作为一种发展趋势,对中外高等教育领域来说都是新生事物。如果说我们的定量研究曾经落后于大多数西方发达国家的话,我们的大数据研究几乎是和国外站在同一个起跑线上,能否抓住机遇,实现高等教育研究的一次飞跃,关键取决于高等教育研究者数据素养的提升和高等教育大数据研究文化的培育和形成。这绝非一朝一夕之功,只能积跬步而至千里,以开放的心胸开展跨学科、跨国界的合作,逐渐增强高等教育研究的科学性,为高

等教育质量的提升和高等教育改革的实践提供更为坚实的理论与事实的支撑。

参考文献:

[1] Project N H. NMC Horizon Report: 2013 Higher Education Edition (II) [R]. Journal of Guangzhou Open University 2013.

[2][19] 国务院关于印发促进大数据发展行动纲要的通知 [EB/OL]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

[3] 南豪峰. 大数据在社会研究中的应用现状 [J]. 江汉大学学报(社会科学版) 2015(10) .

[4] 王晓波. 大数据促进教育变革与创新——专访中央电化教育馆王晓芜副馆长 [J]. 中小学信息技术教育 2013(10) : 10 - 11.

[5] 郭文革. 中国网络教育政策变迁 [M]. 北京: 北京大学出版社 2014: 265.

[6] 我国网民达 7.3 亿 [EB/OL]. <http://news.163.com/17/0123/04/CBEI5S4600018AOP.html>.

[7] 谭维智. 大数据时代的中国大学社会声誉研究 [EB/OL]. <http://paper.dzwww.com/dzrb/content/20160128/Article120011MT.htm>.

[8] 黄欣荣. 大数据时代的思维变革 [J]. 重庆理工大学学报(社会科学版) 2014(5) : 13 - 18.

[9] 刘凤娟. 大数据的教育应用研究综述 [J]. 现

代教育技术 2014(8) .

[10][11] 张俊超. 院校研究如何通过数据分析为大学管理决策服务——“院校研究数据分析的对象、内容和方法研讨会”暨 2013 年中国院校研究会年会综述 [J]. 高等教育研究 2013(8) : 106.

[12] 沈学珺. 大数据对教育意味着什么 [J]. 上海教育科研 2013(9) : 9 - 13.

[13] Grover J. Opportunity through Education: Two Proosals [J]. Brookings Institution 2011: 181 ~ 186.

[14][15] 维克托·迈尔-舍恩伯格; 肯尼恩·库克耶. 与大数据同行: 学习和教育的未来 [M]. 赵中建, 张燕南, 译. 上海: 华东师范大学出版社 2015: 120, 132.

[16] 刘宝存, 杨尊伟. 大数据时代比较教育研究范式的转型 [J]. 比较教育研究 2015(10) .

[17] Great Britain. Department for Business, Innovation and Skills (BIS). Seizing the data opportunity: a strategy for UK data capability [EB/OL]. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf.

[18] 李启虎, 尹力, 张全. 信息时代的人文计算 [J]. 科学 2015(1) .

(责任编辑 刘第红)