

MOOCs 平台大数据的教育实证

李曼丽^{①*}, 黄振中^②

① 清华大学教育研究院, 北京 100084;

② 清华大学在线教育办公室, 北京 100084

* 联系人, E-mail: marylee@tsinghua.edu.cn

2014-10-15 收稿, 2014-12-24 接受, 2015-01-20 网络版发表

摘要 随着大数据技术和大规模在线开放课程(MOOCs)的不断发展, 基于MOOCs平台数据的教育研究也日益涌现. 然而, 相关研究在研究问题、理论基础和研究方法等方面还有待进一步探索. MOOCs平台大数据将为教育问题的研究提供不同类型的数据, 有助于打破以往的一些限制. 目前来看, 基于MOOCs平台大数据的教育实证研究可能会出现聚焦MOOCs本身的研究、大规模在线开放式研究和应用于经典教育研究框架内的研究3种类型.

关键词大数据
MOOCs
教育研究

近年来, 信息技术的不断进步对于社会发展的影响日益显著, 其中, 尤以“大数据”技术受到各界关注. 有专家认为, “大数据是人们获得新的认知, 创造新的价值的源泉; 大数据还是改变市场、组织机构以及政府和公民关系的方法……大数据开启了一次重大的时代转型”^[1]. 毫无疑问, 大数据对教育发展也将产生重要影响, 国内许多研究者指出, “大数据”对教育的影响涉及教育决策、教育评估、教育研究、课堂与教学、个性化教育等各个方面^[2,3].

大规模开放在线课程(MOOCs)的迅速发展, 为大数据与教育研究的结合提供了新的条件. 专家和媒体普遍认为, 依托于互联网技术的MOOCs将为教育研究提供大规模的数据基础. 在大数据相关技术日益发展的背景下, 特别是教育数据挖掘和学习分析技术受到相关领域研究者的日益关注, MOOCs展现了通过开拓教育研究以助力教育水平提高的路径和美好蓝图, 教育界和学术界对此寄予厚望. 但是, 由于MOOCs目前还是新鲜事物, 使用MOOCs平台大数据来进行教育的研究尚处于起步和摸索阶段. 那么, 现有的基于MOOCs平台数据的教育研究呈现怎样的特点? 教育数据挖掘和学习分析如何利用MOOCs平台大数据来进行教育研究? 本文将对此进

行初步探讨与分析.

1 MOOCs相关研究概况

1.1 早期MOOCs研究

MOOCs是“开放教育资源(open educational resource, OER)”运动发展的一种新模式^[4]. 一直以来, OER在中国的主要表现形式包括远程开放教育和网络高等教育. 随着其近年来的蓬勃发展, 相关领域研究也显著增加, 内容涉及基本理论、政策、教学和学习模式等. 然而, 部分研究在路径和方法上都还有很大的提升空间. 以学习模式的研究为例, 一方面缺乏系统性实证研究; 另一方面无论是思辨性的还是归纳性的分析都缺乏较为坚实的理论论证^[5]. 当然, 必须注意到在中国开放教育的语境下, 基于学生的背景和学习数据已有大量的研究. 例如, 基于数据挖掘技术, 在E-Learning数据挖掘方面, 典型应用就有学习者特征识别、学习者在线学习行为分析和师生交互分析等^[6]. 这些研究的样本能达到一定规模, 也形成了相对稳定的分析技术和具体的分析问题. 不过, 由于这一时期主要实践形式是远程开放教育和网络高等教育, 相关研究样本也集中于该体系内.

引用格式: 李曼丽, 黄振中. MOOCs 平台大数据的教育实证. 科学通报, 2015, 60: 570-580

Li M L, Huang Z Z. Exploration of educational empirical research based on big data from MOOCs platforms (in Chinese). Chin Sci Bull, 2015, 60: 570-580, doi: 10.1360/N972014-00983

伴随开放教育运动的推进,自2008年第一门MOOC推出以来,短短几年间MOOCs迅猛发展,到2011年,斯坦福大学“人工智能导论”课程吸引了16万学习者,随后的2012年主要的MOOCs平台开始在美国出现,时至今日,全球MOOCs平台涌现,覆盖了数以千万的学习者¹⁾。Rodriguez^[7]把MOOCs分成2类:遵循联通主义的cMOOCs和符合行为主义的xMOOCs。一般认为,cMOOCs多见于2012年以前,强调学习者共享学习资源和深入讨论的社会化学习;自2012年edX和Coursera兴起后流行的是xMOOCs,以讲授性的视频和丰富练习为主要形式。这一划分类型得到普遍使用^[8-10]。

与此同时,MOOCs相关研究也不断涌现并初步呈现阶段特点。有国外研究者系统检索和回顾了经同行评审、并于2008~2012年期间发表的45篇MOOCs相关论文,将它们归纳为8个研究兴趣领域,包括介绍性的、概念、个案研究、教育理论、技术、聚焦参与者、聚焦提供者以及其他问题。就研究方法和数据来源而言,该文归纳了其中21篇包含个案研究成分的文章(共研究13门不同的MOOCs),发现大部分个案研究运用多种方法收集数据,主要是在线调查,也包括电子邮件访谈、焦点小组、Moodle数据记录、论坛数据、博客及观察记录、虚拟民族志和叙事研究法等方法^[11]。而通过对知网数据库的检索²⁾,发现在2012年及以前,只有8篇相关文献,涉及领域主要是介绍性的、概念和技术。总的来说,由于2012年以前主要是cMOOCs,相关平台也没有发展起来,因此利用平台的数据为主进行的研究还比较少。

1.2 2012年以后的MOOCs研究

2012年以后,随着MOOCs的更快速兴起和发展壮大,相关研究也日益增多,在形式、方法和关注的内容方面也呈现多样化的特点。研究成果主要有MOOCs课程报告、大规模用户调查和MOOCs发展报告、会议类研究报告、发表于学术期刊的论文和专著等类型。

(i) 课程报告。典型的课程报告如2013年上半年公布的杜克大学《对生物电学课程量化视角的分

析》^[12]、爱丁堡大学的《MOOCs@爱丁堡》^[13]以及麻省理工学院(MIT)的《“世界课堂”中的学习:对edX首门MOOC的研究》^[14]。这3份报告根据课程的在线数据、并结合对学习者的问卷调查的结果,对课程的设计、运营、学习者背景、学习过程等进行了分析,集中探讨了学习者面貌(地域分布、年龄、性别、学历)、学习愿望和动机、学习行为和状态等问题^[4]。2014年初,以课程数据为基础,MIT和哈佛研究团队推出了围绕其在edX上首批17门课程的《HarvardX和MITx:在线开放课程的第一年》系列课程工作报告^[15],系统分析了每门课程平均200亿字节的数据,访谈了授课教师、开发团队和学生,结合问卷调查结果,给出了课程注册者的人口学统计数据及其课程活动情况,较为深入地描述了每门的人口学统计数据及其课程活动情况,包括注册者的课程活动(类型)、人口学特征、证书完成情况等。该报告提出了一个基于学习行为的分类框架用作进一步的学习者分析。

(ii) 大规模用户调查和MOOCs发展报告。上述报告使用了平台的课程数据并结合问卷调查的结果,而单纯使用问卷的方法对于研究学习者总体状况也有重要意义。例如2013年国外机构发起的“全球学习者”调查^[16]和国内机构发起的“MOOC中文用户调查”(http://mooc.guokr.com/opinion/437530/, http://mooc.guokr.com/opinion/437642/)等,关注学习者基本习惯、学习动机、MOOCs使用情况以及对MOOCs的评价等内容。就MOOCs发展状况而言,大型的报告如哥伦比亚大学教师学院发布《MOOCs:期望和现实》(http://cbcse.org/wordpress/wp-content/uploads/2014/05/MOOCs_Expectations_and_Reality.pdf)报告,主要运用访谈法,确认各机构使用MOOCs的目的包括扩大受教育范围、建立并维护品牌形象、降低教学成本且提高效率和提升学习效果等,发现这些目标的达成程度各有不同。

(iii) 会议类研究报告。学术会议等活动是推动研究的有效途径。盖茨基金会资助的“MOOC研究行动(MRI)”于2013年启动,支持了22项不同主题研究(http://www.moocresearch.com/reports)。2014年在欧洲举办了“MOOCs利益相关方峰会”、并发布了论文报

1) 仅Coursera的注册用户已超一千万(http://video.cnbc.com/gallery/?video=3000328114)

2) 根据笔者于2014年10月20日以“主题”为检索条件,分别以MOOC、MOOCs、大规模在线课程、大规模开放在线课程、大型开放式网络课程、大规模在线开放课程、慕课等为检索词进行检索的结果

告集^[17]。它们关注的话题最主要的是学习者参与情况,包括坚持(persistence)、完成率(completion rate)、退出(dropout),以及这些行为与动机等因素之间的关系;除此之外,还涉及到讨论区参与、自主学习和自我导向学习、同伴互评、课程视频及使用分析、个人化学习、MOOCs数据、MOOCs与职业教育、混合学习/SPOC和MOOCs分析框架等各类主题。其中,很多研究使用了MOOCs平台的数据,这将在后文做更详细的介绍。

(iv) 发表于学术期刊的论文。仅就国内研究而言,根据前文同样的检索方法,2013年开始大量涌现、共有记录353条,2014年的记录共有1050条³⁾——文献数量的迅猛增长反映了学界对MOOCs关注程度的迅速提升。从研究主题来看,对文献题目初步分析的结果主要有:(1)以介绍和思辨性分析为主,主要从宏观层面讨论MOOCs发生的原因、条件、发展状况和对未来的影响等问题;(2)围绕平台的研究,如MOOCs平台与典型网络教学平台的比较、MOOCs平台的特点和创新发展的建议等;(3)实践总结类研究,如课程建设与实施、视频制作、教师参与MOOCs实践等;(4)不同实践形式如SPOC、混合式教学有关的介绍及案例分析。总的来说,利用平台数据的研究还相对较少,实证类的研究以案例分析、实践经验总结为主,开始出现少量的基于平台数据的研究。

(v) 专著。有关MOOCs的专著尚不多见。中文专著如《解码MOOC》^[4]、《大学的革命——MOOC时代的高等教育》^[18]等,主要采用文献法、案例研究和调查法等:前者从教育学视角考察了MOOCs的兴起、影响、平台及运作模式、课程教学设计和制作、以及相关研究探索等问题;后者全面介绍了国内外MOOCs发展总体状况,透彻分析了它对于高等教育和经济社会发展的深刻影响。

综上,在MOOCs快速发展的背景下,相关研究在类型、主题和内容方面都呈现多样化特点,可以看出有越来越多的研究采用在线平台数据为基础,问卷调查、案例研究也是重要的方法和手段。本文关注的重点是基于MOOCs平台数据的教育研究,这方面国内目前相对较少,而国外有较多的成果已发表,大致包括学习模式和学习者分类、学习的坚持/完成/退

出、聚焦MOOCs具体的功能等主题。

(i) 学习模式(pattern)及学习者分类研究。目前,分析学习者的学习状况(过程和结果等),进而对学习者的学习模式进行分类是利用MOOCs平台数据研究的热点。一般根据学习者是否取得证书这一结果可以将学习者简单区分为两类。如果考虑学习行为,那么对学习者的类别则更为复杂。基于研究者对“Change11 course”这门cMOOC的学生进行有关其学习行为的问卷和访谈结果,Phil Hill将学习者分为爽约者(no-shows)、观察者(observers)、插班生(drop-ins)、被动参与者(passive participants)和主动参与者(active participants)5类(<http://mfeldstein.com/emerging-student-patterns-in-moocs-a-revised-graphical-view>)。不过,该框架的基础还不是对平台数据的计算分析。综合考虑学习行为和学习结果,并使用MOOCs平台的大数据,研究者们提出了不同的分类方法。斯坦福大学的Kizilcec等人^[19]运用聚类法,对3门MOOCs的注册学习者及其与课程视频和测验的交互程度进行分析,将学习者的行为模式分为完成(completing)、旁听(auditing)、辍学(disengaging)和取样(sampling)4类轨道(trajectories)。《HarvardX和MITx:在线开放课程的第一年》分析了edX平台上两校课程的学习者数据^[15],依据其点击平台“课件”内容的比例和最终分数是否达到合格分数线这两个指标,将他们分为注册者(registered)、观看者(viewed)、挖掘者(explored)和持证者(certified registrants)4类,并运用这一框架分析了各门课程。这一分类框架也被应用于中文MOOCs平台“学堂在线”的学习者分析^[20]。最近,斯坦福和康纳尔大学的研究者^[21]根据Coursera平台6门课程的学习者数据,以观看视频和完成测验数量的比例关系为“投入(engagement)”的操作化定义,将学习者分为了观看者(viewers)、答题者(solvers)、全参与者(all-rounders)、搜集者(collectors)和旁观者(bystanders)5类,并探讨了学习者投入与其课程成绩的关系。

上述研究对基于学习者行为模式的分类框架提出了不同的解决方案。各研究的数据来源和类型大致相同,主要包括学习者学习的过程和结果两类。区别主要在于数据分析的不同角度和方法,例如聚类

3) 以直接检索结果(包括论文、报刊文章等各类文章)的记录数为准,2014年的记录截至9月25日

分析和依据定义的数据比例结算等,各研究也有相似的结果.基于这些发现,研究者们对有关MOOCs通过率较低的质疑进行了回应,他们认为学习者的不同模式的行为体现了学习者需求的多样化,这无法由通过率这样的单一指标涵盖,MOOCs的价值体现和评价不应局限于此.

(ii) 学习者完成或退出课程的相关问题研究.上述聚焦学习模式的研究对MOOCs完成率的质疑提出了不同意见,与此同时,研究者也关注影响学习者完成或退出课程的因素,以及学习者退出课程行为的预测因子等问题.

一个由美国和印度的研究者组成的小组发现MOOCs中的社交因素对学习者的退出存在影响^[22],他们运用探索性因子分析方法和社交网络分析的思路,基于Coursera上一门课程学习者的发帖行为和社交网络行为构建了一个生存模型(Survival Model).而MIT研究人员则发现发帖行为中只有帖子长度这一指标对学习者的成功完成课程具有较高预测度,而更显著的预测指标是个体学习者的作业提交时间和每周作业的提交比例等.他们使用大数据和机器学习的方法分析了edX首门课程“6.002X”5万名学习者产生的共计13亿的各类在线点击数据,在此基础上研究影响学习者退出课程的因素^[23].同样是关注学习者退出的预测,斯坦福大学的一个研究小组在将“退出”定义为缺席课程超过1个月或者观看视频少于一半的基础上,分析了不同领域12门课程学习者的各种行为特征变量,设计了一个包含跳过视频、跳过测试、拖延和课程表现等4个变量的学习者退出预测模型:该模型预测活跃的和缺席课程超过14天的这两类学习者退出行为的准确率分别达到40%~50%和90%以上,教师可以利用该模型的预测结果对有退课风险的学习者进行及时干预^[24].

由此可见,预测学习者持续参与课程的指标是多样的,包括时间、讨论区表现等更为具体和细致的行为,目前相关研究还没有取得一致结论.仅就MOOCs平台数据的范围而言,研究者需要从多种在线行为记录中提取最主要的变量以表征学习者的退出可能性.从上述研究可以发现,研究的理论框架、研究设计和分析方法对于最终结果的差异有极为重要的影响.不过,虽然上述研究最终提取的指标完全不同,但相似的结论是学习者已经投入的程度(参与各类在线学习活动的范围和深度)对于其持续参与课

程的程度有显著的作用,这也证明了促进在线学习者的参与度有助于课程完成率的提高.由此可见,有关预测指标的探讨能够加深对学习者持续学习影响因素的研究.同时,这对教师的快速干预、师生交互、课程设计等实践有指导意义,也为MOOCs平台需完成哪些即时的数据分析与反馈有所启示.基于即时反馈的个性化教学,正是各界对MOOCs的期望所在.

(iii) 关于具体的学习行为和功能的研究.除了上述两类主要议题之外,还有很多基于MOOCs平台数据的教育研究涉及了更为具体的在线学习的分析.MIT和哈佛课程工作报告提出以“课程点击量”、“学习者活动日”和“讨论区发帖数”3个指标来描述课程的活动参与情况^[15].对于xMOOCs来说,视频是主要的学习资源和途径,Seaton等人^[25]利用平台提供的视频交互的不同类型数据,分析了edX首门课程6.002x的证书获得者使用视频的特点.讨论区参与也是MOOCs学习的重要组成部分,研究者们一般利用平台提供的学习者发帖、投票等讨论区行为的量化数据,并结合内容分析等质性研究方法展开有关讨论区的研究,例如:对活跃发帖人的行为特点、相关因素及影响^[26]的分析,基于计算语言学模型的思路对帖子内容进行分析、进而测量学习者的动机和认知投入^[27],使用情感分析(sentiment analysis)的方法利用讨论区的文本数据研究学习者对课程的评价^[28],等.在测验与评估方面,有利用平台数据对短文自动评分系统(automated essay scoring)的研究^[29]和对同伴互评(peer assessment)功能的研究^[30]等.

综上所述,目前研究者已开始利用MOOCs平台数据进行了很多实证研究.从主题来看,主要集中于对学习者的描述和行为模式的探索,对应于美国教育部报告^[31]中所提出的学习者经历建模、用户档案、趋势分析等教育数据挖掘和学习分析的应用领域,也可能受到互联网影响而体现出用户研究的导向.然而,很多研究的问题不具备显著的“教育意涵”,也缺少教育、心理学等相关理论支撑.从数据来源和分析方法来看,使用的数据类型和层次还比较有限,在数据挖掘技术方面上述研究涉及了统计分析、聚类、预测、文本挖掘等常见的教育数据挖掘和学习分析技术.MOOCs平台则能提供更大规模的数据支持相关研究.而面对越来越大的超过传统研究规模的数据时,很多研究者会受到既有研究套路的束缚、甚至可

能会感到无所适从。在此过程中,教育和心理的基本理论和基本问题对研究的定位、范围选择等还是具有很强的指导意义,利用MOOCs平台大数据进行教育研究不应与经典的教育研究框架相脱离、割裂。为此,下文将围绕经典的教育实证研究框架和以MOOCs为代表的、具有大数据特点的新教育情境之间的关系,探讨在MOOCs背景下如何开展教育研究。

2 MOOCs与经典教育实证研究框架的关系

利用MOOCs平台数据进行有关教育的研究,不应脱离教育研究的基本问题。经典的教育实证研究框架,对此类问题的澄清、方法的选用等都具有启发性。一般教育研究可以从3个层面的视角出发:在微观层面关注个体学习与课堂教学,在中观层面关注学校的组织管理和运行,在宏观层面关注整体的教育政策和教育系统的问题。结合MOOCs平台的实践状况,下文将重点探讨学习理论、教学理论和教育政策等领域中经典教育实证研究框架和MOOCs平台数据研究的关系。

2.1 学习理论

作为社会发展和个体成长的必然要求,学习既是人类的一个永恒主题,也是当代很多学科的主要关注领域之一。自古至今,学习研究逐渐成为了一门典型的跨学科研究领域,来自哲学、心理学、教育学、神经科学、信息技术等各个学科的专家学者,从不同角度对有关学习的各种问题进行了思考、分析和研究,见仁见智地提出了不同的学习理论,不断地推进我们对于学习的本质和规律的理解与认识。在教育研究中,学习(理论)既是主要的研究主题之一,也是重要的理论基础来源。其中,教育心理学是以学习研究为主要领域的代表学科。一般认为,学习理论可大致分为行为主义、认知主义、建构主义和人本主义等流派。不管采取怎样的视角,不同的理论在以教学促进学习的问题上还是存在一些共同点的^[32]。

对于MOOCs来说,也是以一定的理论为支撑的,并非无源之水、无本之木。根据前文所述,目前主流的xMOOCs和早期的cMOOCs,其理论基础可分别追溯至经典理论流派行为主义和近年来兴起的所谓“联通主义”。MOOCs实践本身,则为相关理论提供了新的情境和更多可能的研究问题。

以短视频和及时反馈等为主要特点的xMOOCs,符合斯金纳的学习理论。斯金纳基于其操作性条件反射理论和强化等核心概念,撰写了大量将这些理论应用于教育领域的专著,并发展了程序教学。程序教学一般要遵循积极反应、小步子、即时反馈、自定步调等原则,而这正是多数xMOOCs所呈现的特点^[4]。实际上,程序教学在实践中并不罕见,除了早期专门的教学机器外,在面授的课程中也是可以采用的^[32]。在MOOCs模式下,就可以利用大规模学习者的数据做进一步的验证和研究。此外,MOOCs学习是非强制的,这和正规学校教育中对学习者的强制要求有很大不同,这又对程序教学有着怎样的影响呢?这一问题也可以通过在线平台的数据进行研究。

最初出现的cMOOCs则是基于联通主义的学习理论而产生的。考察经典的学习理论,可以发现它们的一个共同特点是将学习的焦点放在学习者个体内部。而在互联网对生活影响日益显著的当代社会,就有研究者开始讨论学习的网络特性。Siemens^[33]系统地提出了学习的联通主义思想,突破将学习视为个体内部的视角,指出学习是节点和信息源间的连接过程。联通主义表达了“关系中学习”和“基于网状连接的分布式认知”的观点,关注个体之间思想联系之上的知识迁移与知识创造。然而,联通主义目前还不能严格地看作一种学习理论,更确切的讲还是一种“学习观”。正如Siemens在与国内学者通讯中指出:“我不知道联通主义是否能够很好地解释人类的学习。做这样一个描述和论断还非常早。然而,在我看来,联通主义能够更精确地解释学习而不是建立一种理论。”^[34]可以说,联通主义是在学习(者)外部连接上提出了一种分析视角,而没有系统地关注学习的产生及其机制。一方面,联通主义的观点还缺乏坚实的实证研究基础,另一方面“其倡导的教学模式目前还只是研究者的一种探索,尚未形成稳定、易于操作、可供一般教师使用的实践模式”^[35]。

MOOCs对其学习者也提出了更严格的有关“自主学习(self-directed learning)”和“自我导向学习(self-regulated learning)”的要求。受人本主义心理学的影响,发达国家在20世纪70年代提出了自主学习。目前,研究者对自主学习的内涵、影响因素和促进自主学习的方法等已达成一定共识,对自主学习的关注已主要进入实践应用阶段。研究者认为,如果学生的

学习动机是自我驱动的,学习内容是自己选择的,学习策略是自主调节的,学习时间是自我计划和管理,学生能够主动营造有利于学习的物质和社会性条件,并能够对学习结果做出自我判断和评价,那么他的学习就是充分自主的^[36]。基于MOOCs特性,对于学习者(特别是进行深入学习和取得证书者)学习行为数据的分析,能够以更大规模的数据来验证甚至推进已有的相关理论。在成人学习领域,更为常见的相关概念是“自我导向学习(self-regulated learning)”。在网络学习环境下,需要更多良好设计的研究来分析自我导向性学习及其对学习者的影响^[37]。尽管目前有关网络环境下自我导向性学习的研究还有很多是描述性的、方法有限的,仍已有研究探讨了在线环境下学生的动机、策略和学习表现之间的关系,发现高自我导向的学习者更可能取得学习成功;发现了其对在线学习支持的有效性。在MOOCs环境下,有条件进行更加深入的研究,以探讨不同因素对自我导向性学习及学习效果的影响。

2.2 教学与课程论

作为教育学的传统主干学科,教学与课程论既是教育领域的基础性研究,也受到学习理论发展的推动。一直以来,系统的教学与课程论主要关注的是基础教育领域;在高等教育研究中,大学教学论则一直是被忽视的领域;而正规学校体系外的辅导、培训、成人和职业教育,可能更不需要“教学论”,更多的是经验性的总结,或者是演绎甚至照搬基础教育教学与课程基本理论。

然而,根据本文第一部分所列有关课程报告,可以发现MOOCs的学习者多是成年人。由此,MOOCs为面向成年人的课程设计和教学规律与方法等的研究提供了很好的契机。目前的成人教学论研究,在研究方法方面存在对调查法、实验法等基本研究方法的运用还相对较少等问题^[38],该领域还比较缺少实证性的研究。在MOOCs环境下,大学和成人教学已有理论,可以通过实际的课程安排和教学设计进行检验;对课程实施、学习者实际学习的情况的相关数据进行分析,还可以丰富相关理论。

2.3 教育政策研究

上文从学习理论、教学与课程的角度探讨了MOOCs下教育研究的可能进展,层次相对微观。作

为在线教育当今发展的新形态,MOOCs对宏观教育格局也有着重要影响。清华大学校长陈吉宁认为:优质教育资源的大规模扁平化扩散,在带来教育低价优质的同时,也将深刻改变高等教育全球化的竞争模式。在教育多样性、学习多样性、学习自由性全面发展的同时,文化多样性和学习的选择性将有可能受到全面的挑战(<http://news.tsinghua.edu.cn/publish/news/4204/2013/20130605095021423650319/20130605095021423650319.html>)。因此,MOOCs在教育政策研究上也有相当价值。

在教育政策研究领域,公平和效率是教育政策追求的重要价值^[39,40]。教育公平与教育质量息息相关。宏观地看,教育质量即整个教育体系的质量,实质上是指其与系统规模、结构和效益等之间协调问题。也就是说,它以系统内部各要素之间是否协调一致为标准。什么时候系统各要素之间协调一致,什么时候就表现出较高的体系质量^[41]。教育公平是深受关注的研究领域,也有大量实证研究。另一方面,效率就是指投入与产出之比,因此教育的效率就可以理解为教育投入和教育产出之比,当然人们都希望投入越少、产出越高。然而,由于教育是培养人的实践活动,其投入与产出比纯粹的物质生产要复杂不少。首先面对的问题就是如何定义产出和投入;其次,用什么样的方法来衡量。在教育经济学领域,多采用成本收益分析的方法来直接计算教育的投入产出,包括宏观社会的和微观个体的;此外,也有借用这一方法来进行学习性投入产出的研究(而不是直接的经济成本)。

MOOCs的快速兴起,一个重要的原因在于它提供了一种低成本的优质教育资源传播的解决方案,有助于同时解决教育公平和教育效率的问题,从而提高教育质量。这也将是未来MOOCs有关研究中的重要问题。从目前已有的研究成果来看,很多对学习者的调查就涉及到了此问题。多数报告都发现了学习者的多样性,然而年初edX和滨州大学的报告结果,却被广泛的MOOCs研究证明学习者的集中性。国内的调查和平台数据结果也发现,用户集中于经济发达地区和高等教育水平较高的地区,这些区域往往有较好的网络和计算机条件。这些发现就和人们的期望并不一致。对于MOOCs是否能如人们所期望那样,真正益于有效地解决教育公平问题,还是相反地扩大技术鸿沟,还有待于MOOCs进一步发展和

更多相关研究加以证明。

3 MOOCs平台对教育实证研究中的数据支持

MOOCs为教育理论的实证分析提供了更为丰富和便利的数据支持。一方面,在线记录使得数据的收集更为便利,且具有全面、处理相对便捷、易于保存和客观等优点,大数据技术使得多元化的分析更为方便。另一方面,虽然从数据来源来看,MOOCs数据仅仅是由在线网络平台提供的,还不包括其他技术手段搜集到的数据;但是,开放、在线、数量庞大及内容丰富的名校课程,能吸引大规模的学习者,使系统可以全面记录和积累他们的在线行为数据,相对于传统的非在线数据支撑的研究来说,这些都是全新的资源。这里,数据的丰富性既指其规模容量,也包括数据的类型。目前,MOOCs平台能够直接提供的数据包括用户日志、用户IP地址、学习者成绩和大量的文本数据;可根据研究需要将用户日志转化为学习者行为、学习路径、或不同资源与功能的使用状况等信息。

从MOOCs平台所提供的数据来看,首先,基于用户日志的数据是目前研究中使用的主要基础。网站会记录用户注册、登录、点击链接、页面停留及跳转等所有行为,形成用户日志,在MOOCs这样的学习平台上,它们组成了用户学习行为的基础数据。早期的网络教育研究多以此为主要数据来源,通过学习者登录的次数、频率、时间和周期等各类数据来对研究中的概念进行操作化定义。研究者可按需整理MOOCs平台用户的行为日志,以获得各类学习数据。例如,xMOOCs中课程视频是一个重要的学习资源,通过对数据的整理,可以得到每一个学习者观看所有视频的情况,包括是否有停顿、反复、有选择性地观看、在不同学习阶段的观看特点等,数据全面、精确。目前Coursera和edX等主要平台已经或正在开发有关技术和可视化呈现的在线功能,国内也有研究者在做相关探索^[42]。又如,学习者使用讨论区的情况,包括是否看帖、是否参与讨论、讨论的深度(发帖、回帖等互动行为等)。再如,学习者点击课程内不同模块的情况,点击不同页面的频率、时间点、页面间的跳转情况等。对学习日志的记录、分析,可以弥补传统课堂教学研究的不足。传统课程教学研究一般采用观察法收集数据,需要制定观察评价的标

准,常常是观察者的主观评价,难免存在数据的遗漏、缺失和误读,此外研究者的在场会影响学习者的实际表现。

其次,根据用户的学习行为日志,可以进一步分析学习者个体或者群体的学习路径、学习模式和学习经历。学习日志是按时间记录的,能够用来描述每一位学习者的学习习惯,在此基础上可以分析群体的学习行为模式类型。对于在线学习产品来说,这些数据的重要价值就体现在为提高用户的学习体验提供了基础;同时,对于学习研究来说,又提供了新的数据来源,并拓宽了学习研究的领域。例如,可以探讨在线学习的规律,是否存在学习模式或者类型?影响因素又有哪些?等。如果把在线学习方式与传统学习方式做类比,可以把观看视频视为正式的课堂学习,而有关学习路径的分析就类似于传统中对课堂外学习行为的研究。这类研究传统上主要的数据搜集方法是研究对象的自陈,存在主观性偏差;而在网络学习的背景下,如前所述,所有行为都是有系统记录的,相对更加客观。在具体实现的技术规范层面,已有相关成果值得参考,例如美国“高级分布式学习”组织发布的xAPI规范等^[43]。

再次,在线系统中的习题和测验等模块体现了学习者学习的直接效果。一般来说,学习研究特别关注影响学习效果的因素。因此,对于MOOCs平台下的学习来说,可以将行为日志、学习路径等数据与学校效果数据结合起来,探讨在线学习情境中不同因素与学习效果之间的关系。与传统方法相比,MOOCs平台下学习效果类数据能够拓宽自身内涵,除了成绩本身之外,还可以考虑作业完成时间、提交答案次数等更多的因素。此外,在讨论区、wiki等体现MOOCs学习特征的功能中,还能够提供大量文本数据供研究进行分析。

对于教育研究来说,MOOCs数据的不足之处在于开放平台下的数据大都是行为和成绩类的,缺少背景信息类的数据,有关动机、态度和情感等方面的研究受到了很大限制,而这些因素又是教育研究不可避免的话题(例如学习者坚持或退出课程的行为就与动机密切相关)。从已有研究可以看出,研究者们多使用调查问卷或者内容分析等方法来弥补问卷的不足。随着技术的发展和研究的不断深入,一方面平台搜集数据的方法和工具将会越来越多样;另一方面,各类研究使用数据的层次和类型也会更加丰富,

同时会探讨利用客观的行为数据来挖掘学习者动机和意图等主观要素的可能。

4 MOOCs平台数据的教育实证研究类型

毫无疑问,在大数据背景下,MOOCs给教育研究带来了许多可能。有研究者认为,由于教育问题本身的复杂性以及对人类行为和思想的测量困难,定量研究一直受到较大的限制。因此,很多教育理论研究都停留在主观解释的阶段,缺少客观验证。而MOOCs的出现为突破教育测量这一难题带来了可能^[4]。MOOCs在带来更多样的在线学习数据同时,也拓宽了教育研究在更多领域和议题的实证基础。

具体而言,MOOCs对于教育研究的拓展可初步划分为3类。第一类是聚焦于MOOCs和在线教育自身的研究,例如本文第一部分所述各类研究,其主题多围绕MOOCs本身所面临的问题:学习者类型、MOOCs中的不同教学环节、MOOCs的课程设计和教学设计,等。第二类是所谓的“MOOR”,即大规模在线开放研究。有研究者认为(<http://mooc.pku.edu.cn/resource/?p=198>),MOOR是MOOCs的延伸,MOOCs模式下教学间即时的网状交叉互动,为大规模人群间思想碰撞、思维启发和灵感激发提供了很好的途径,进而有可能达到集体增智的目的;还有研究者认为,大规模在线开放研究本质上即是一种以研究为重点的大规模在线开放课程(<http://edu.163.com/14/0127/08/9JJ4J5JE00294MBE.html>)——不论做如何理解,MOOR的视角在于将研究本身与MOOCs融合,而不是将后者视为前者的对象。第三类则是以MOOCs平台带来的在线大数据为基础的、对经典教育研究问题和框架的探讨。

教育是对人的成长发展的引导与塑造,根本问题是人的发展问题。在这个意义上,教育研究的主要问题就包括作为个体的人的发展和与社会发展矛盾统一的人的发展。由此考察对于有关MOOCs的研究,传统的教育研究框架依然是有效的,MOOCs为传统的研究框架和经典问题提供更多的数据基础和创新机会,这正是MOOCs在教育发展领域能够带来重大变革的可能之所在。前面,本文已经讨论了经典的教育实证研究框架以及传统数据的局限,以下将列举讨论MOOCs平台数据与经典教育研究相结合的可能趋势。

(1) 我们经常看到以往有关学习动机的研究中,很多数据是来自于学习者的主观自陈报告。而MOOCs

平台提供的数据,在规模巨大的前提下,很有可能为研究动机与行为表现关系的这类问题提供更多的支持。

(2) MOOCs平台下的数据对于教育评价和测量研究将带来巨大的影响。有学习者认为,传统教育测量中,有关学生的学习动机基本相同和学生的心智水平大致相当这2个隐含假设,在MOOCs情境下不再成立(<http://www.yocsef.org.cn/sites/yocweb/yocltzw.jsp?contentId=2790000064253>);或者说,MOOCs开放条件下大规模学习者的现实放大了学生个体之间的差异,使其成为评价中无法再回避的问题。有研究者详细讨论了基于MOOCs的学习分析和教育测量思路^[44],指出“基于MOOCs的教育测量则可以精确跟踪和记录学习者在网络平台学习实时发生的行为”,这将拓宽我们对学习效果的理解,并使过程性评价建立在更坚实的基础上。此外,通过对学习者学习行为和即时反馈的分析,也有助于有关课程和教学设计的研究,并改进教学质量的评估。

(3) 一直以来,个性化教育都是教育研究和实践的一个焦点,也是信息技术支持下的教育创新的诉求。然而,真正个性化教育的实现还面临诸多困难。在MOOCs平台下,教育大数据挖掘和学习分析等技术的运用,为个性化教育的研究和实践探索提供了更多的途径;同时,个性化教育亦是以MOOCs为代表的在线教育发展不可回避的选择。

(4) 教育公平问题是社会发展长期无法彻底解决的难题。例如在高等教育,苏芑和罗燕^[45]认为,优质高等教育资源的短缺和分配不公已经成为全球性的问题……MOOCs则带来了解决这些问题、改变现有状况的希望和方法。助力于教育公平问题的解决,是社会各界热捧MOOCs的一个重要原因;而MOOCs平台数据,特别是立足于学习者背景角度的有关分析,也将拓宽教育公平和均衡问题的研究。

当然,上述只是列举了部分基于MOOCs平台数据已经或可能会出现的教育研究,并分别就MOOCs平台数据的价值和可能的研究方向做了简要的分析。总的来说,在研究领域和研究问题层面,传统、经典的教育问题对于MOOCs的研究来说依然是有效的,如果忽视了经典的教育问题,围绕MOOCs的教育研究在研究深度和价值等方面值得商榷。在方法层面,MOOCs平台的大规模、多样化和即时性的数据,将突破以往教育实证研究所面临的一些限制,使相关

研究得到更坚实的数据支撑,甚至有可能拓宽研究问题和视域,进而深化相关理论研究,助益于教育实践发展.值得注意的是,MOOCs平台对于加入实验性设计的研究非常方便,然而这需要课程教师和研究者的共同参与和配合;此外,平台方面还应该根据课程实践和研究的需要,开发新的功能,并改进数据搜集和整理方面的不足.相信随着MOOCs自身的发展,会涌现出越来越多的相关研究.

5 结论

本文回顾了自MOOCs出现以来有关的各类典型研究,发现其呈现了与MOOCs自身发展相呼应的2个阶段,即研究对象在2012年前后分别聚焦于cMOOCs和xMOOCs.基于MOOCs平台大数据的实

证研究日益增多,目前这些研究主要关注学习模式和学习者类型、学习者完成及退出课程的情况以及更具体的学习行为和MOOCs平台功能等3个领域;然而,多数研究还没有体现出明显的教育“意涵”.利用MOOCs平台大数据进行教育研究不应与该领域的基本理论和问题相脱离,经典的教育和心理研究框架对于这类研究具有启发意义,同时MOOCs为教育理论的实证分析提供了更为丰富和便利的数据支持,基于平台的大数据也有助于基本理论和相关研究的拓展与深化.目前来看,围绕MOOCs的教育研究将有很大的发展,并可根据主题大致划分为3类,包括聚焦MOOCs和在线教育自身的研究、大规模在线开放研究(MOOR)和以MOOCs平台大数据为基础并立足经典教育问题和框架的研究.

参考文献

- Mayer-Shcönberger V, Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think (in Chinese). Sheng Y Y, Zhou T, eds. Hangzhou: Zhejiang People's Publishing House, 2013 [Mayer-Shcönberger V, Cukier K. 大数据时代: 生活、工作与思维的大变革. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013]
- Lu J. Big Data and its Application in Education (in Chinese). Shanghai: Shanghai Research on Education, 2013 [陆璟. 大数据及其在教育中的应用. 上海: 上海教育科研, 2013]
- Zhu Z T, Shen D M. New paradigm of educational technology research based on big data (in Chinese). E-educat Res, 2013, 10: 5-13 [祝智庭, 沈德梅. 基于大数据的教育技术研究新范式. 电化教育研究, 2013, 10: 5-13]
- Li M L, Zhang Y, Ye F G. Decoding MOOC (in Chinese). Beijing: Tsinghua University Press, 2013 [李曼丽, 张羽, 叶赋桂. 解码MOOC——大规模在线开放课程的教育学考察. 北京: 清华大学出版社, 2013]
- Ding S H. Content analysis on learning model of distance education in the past ten years (in Chinese). Open Educat Res, 2009, 15: 93-97 [丁顺宏. 远程学习模式十年研究的内容分析. 开放教育研究, 2009, 15: 93-97]
- Ge D K, Zhang S G, Wei S P. Educational Data Mining: Methods and Application (in Chinese). Beijing: Educational Science Publishing House, 2012 [葛道凯, 张少刚, 魏顺平. 教育数据挖掘方法与应用. 北京: 教育科学出版社, 2012]
- Rodriguez C O. MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. Eur J Open Distance E-Learning, 2012, Retrieved: <http://eric.ed.gov/?id=EJ982976>
- Daniel J. Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. J Interact Med Educat, 2012, 3: 18
- Rosselle M, Caron P A, Heutte J. A typology and dimensions of a description framework for MOOCs. In: Cress U, Kloos C D, eds. Proceedings of the European MOOC Stakeholders Summit. Lausan: PAU Education, 2014. 130-138
- Jadin T, Gaisch M. Extending the MOOCvertistry: A multi-layered and diversified lens for MOOC research. In: Cress U, Kloos C D, eds. Proceedings of the European MOOC Stakeholders Summit. Lausan: PAU Education, 2014. 73-79
- Liyanagunawardena T R, Adams A, Williams S A. MOOCs: A systematic study of the published literature 2008-2012 (in Chinese). Zhang Y S, ed. Distance Educat China, 2014, 3: 6-16 [Liyanagunawardena T R, Adams A, Williams S A. 张永胜, 译. MOOCs已发表文献(2008-2012)的系统研究. 中国远程教育, 2014, 3: 6-16]
- Belanger Y, Thornton J. Bioelectricity: A quantitative approach (Duke University's first MOOC). Working Paper. Duke University, 2013. Retrieved: <https://www.era.lib.ed.ac.uk/handle/1842/6683>
- MOOCs@Edinburgh Group. MOOCs @ Edinburgh 2013: Report #1. Working Paper, University of Edinburgh. 2013. Retrieved: <https://www.era.lib.ed.ac.uk/handle/1842/6683>
- Breslow L, Pritchard D E, DeBoer J, et al. Studying learning in the worldwide classroom research into edX's first MOOC. Res Prac Assessm, 2013, 8: 13-25

- 15 Ho A D, Reich J, Nesterko S O, et al. HarvardX and MITx: The first year of open online courses. Working Paper, HarvardX & MITx, 2014. Retrieved: <http://ssrn.com/abstract=2381263>
- 16 Bartholet J. Worldwide MOOC investigation (in Chinese). *Sci Am*, 2013, 9: 51–52 [Bartholet J. MOOC 全球调查. 环球科学, 2013, 9: 51–52]
- 17 Cress U, Kloos C D. Proceedings of the European MOOC Stakeholders Summit. Lausan: PAU Education
- 18 Wu J P, Zhao K. Revolution of Universities: Higher Education in MOOC Era (in Chinese). Beijing: Tsinghua University Press, 2014 [吴剑平, 赵可. 大学的革命——MOOC 时代的高等教育. 北京: 清华大学出版社, 2014]
- 19 Kizilcec R F, Piech C, Schneider E. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. Proceedings of the third international conference on learning analytics and knowledge. New York: Association for Computing Machinery, 2014. 170–179
- 20 Zhang L. Research on data analysis of massive online education: An example from xuetaangX. (in Chinese). Comprehensive Dissertation Training for Undergraduates. Beijing: Tsinghua University, 2014 [张蕾. 基于学堂在线平台的大规模在线教育数据分析研究. 本科生综合论文训练. 北京: 清华大学, 2014]
- 21 Anderson A, Huttenlocher D, Kleinberg J, et al. Engaging with massive online courses. In: Proceedings of the 23rd International Conference on World Wide Web. New York: Association for Computing Machinery, 2014. 687–698
- 22 Yang D, Sinha T, Adamson D, et al. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In: Proceedings of the 2013 NIPS Data-Driven Education Workshop, 2013
- 23 Taylor C, Veeramachaneni K, O'Reilly U M. Likely to stop? Predicting stopout in massive open online courses. arXiv: 1408.3382v1[cs.CY], 2014
- 24 Halawa S, Greene D, Mitchell J. Dropout Prediction in MOOCs using learner activity features. In: Cress U, Kloos C D, eds. Proceedings of the European MOOC Stakeholders Summit. Lausan: PAU Education, 2014. 58–65
- 25 Seaton D T, Nesterko S, Mullaney T, et al. Characterizing video use in the catalogue of MITx MOOCs. In: Cress U, Kloos C D, eds. Proceedings of the European MOOC Stakeholders Summit. Lausan: PAU Education, 2014. 140–146
- 26 Huang J, Dasgupta A, Ghosh A, et al. Superposter behavior in MOOC forums. Proceedings of the First ACM Conference on Learning@Scale Conference. New York: Association for Computing Machinery, 2014. 117–126
- 27 Wen M, Yang D, Rosé C P. Linguistic reflections of student engagement in massive open online courses. In: Proceedings of the International Conference on Weblogs and Social Media. 2014
- 28 Wen M, Yang D, Rosé C P. Sentiment analysis in MOOC discussion forums: What does it tell us? In: Proceedings of Educational Data Mining, 2014
- 29 Corliss S, Reill E D. Promoting a higher-level learning experience: Investigating the capabilities, pedagogical role, and validity of automated essay scoring in MOOCs. 2014. Retrieved: http://www.moocresearch.com/wp-content/uploads/2014/06/MOOC-Research-Initiative_AES-system.pdf
- 30 Piech C, Huang J, Chen Z, et al. Tuned models of peer assessment in MOOCs. arXiv: 1307.2579, 2013
- 31 US Department of Education. Enhancing teaching and learning through educational data mining and learning analytics. 2012. Retrieved: <http://www.ed.gov/technology>
- 32 Schunk D H. Learning Theories: An Educational Perspective, 3E (in Chinese). Wei X M, ed. Nanjing: Jiangsu Education Press, 2003 [Schunk D H. 学习理论: 教育的视角. 韦小满, 译. 南京: 江苏教育出版社, 2003]
- 33 Siemens G. Connectivism: A learning theory for the digital age. *Instruct Technol Distance Learn*, 2005, 2: 3–10
- 34 Wang Y M, Zhu Z T. Connectivism: A new approach of learning Theory (in Chinese). *E-educat Res*, 2006, 3: 5–9 [王佑镁, 祝智庭. 从联结主义到联通主义: 学习理论的新取向. 中国电化教育, 2006, 3: 5–9]
- 35 Han X B, Zhai W F, Cheng J G. A dialectical analysis of cMOOC & xMOOC and their integration into the ecological chain of higher education (in Chinese). *Modern Distance Educat Res*, 2013, 6: 3–10 [韩锡斌, 翟文峰, 程建钢. cMOOC 与 xMOOC 的辩证分析及高等教育生态链整合. 现代远程教育研究, 2013, 6: 3–10]
- 36 Pang W G. Students' self-regulated learning (in Chinese). *J East China Normal Univ (Edu Sci)*, 2014, 20: 78–83 [庞维国. 论学生的自主学习. 华东师范大学学报(教育科学版), 2001, 20: 78–83]
- 37 Artino Jr A R. Self-regulated learning in online education: A review of the empirical literature. *Int J Instruct Technol Distance Learn*, 2007, 4
- 38 He A X. Review and reflection on the development of pedagogy of adults (in Chinese). *J Adult Educat College Hubei Univ*, 2006, 24: 42–45 [何爱霞. 成人教学论发展的回顾与反思. 湖北大学成人教育学院学报, 2006, 24: 42–45]
- 39 Pan M Y. Equity and efficiency: The bases of higher educational policy making (in Chinese). *Peking Univ Educat Rev*, 2003, 1: 54–57 [潘懋元. 公平与效率: 高等教育决策的依据. 北京大学教育评论, 2003, 1: 54–57]

- 40 Chu H Q. Educational equity and educational efficiency: The dual objectives of educational reform and development (in Chinese). *Educational Research*, 2008, 7-13 [褚宏启. 教育公平与教育效率: 教育改革与发展的双重目标. *教育研究*, 2008, 7-13]
- 41 Gu M Y. *Dictionary of Education* (in Chinese). Shanghai: Shanghai Education Press, 1998 [顾明远. 教育大辞典. 上海: 上海教育出版社, 1998]
- 42 Ma C H, Xu Y Y. Research and solution on data collection of learning analytics based on collecting the behavior data of learners when watching online video (in Chinese). In: 12th Conference Proceedings of China Higher Education Information Academy. Shanghai: East China Normal University, 2014 [马晨辉, 徐艳艳. 学习分析中数据采集的研究与技术实现——以采集学习者视频行为数据为例. 中国高等教育学会教育信息化分会第十二次学术年会论文集. 上海: 华东师范大学, 2014]
- 43 Gu X Q, Zheng L W, Jian J. Access to educational big data: Tracking and sharing the learning experience based on the xAPI specification (in Chinese). *Modern Distance Educational Research*, 2014, 5: 13-23 [顾小清, 郑隆威, 简菁. 获取教育大数据: 基于 xAPI 规范对学习经历数据的获取与共享. *现代远程教育研究*, 2014, 5: 13-23]
- 44 Zhang Y, Li Y. Learning analytics and education assessment based on big data generated by MOOCs (in Chinese). *Tsinghua Journal of Education*, 2013, 34: 22-26 [张羽, 李越. 基于 MOOCs 大数据的学习分析和教育测量介绍. *清华大学教育研究*, 2013, 34: 22-26]
- 45 Su P, Luo Y. Technology myth or educational revolution? Impact of MOOCs on higher education (in Chinese). *Tsinghua Journal of Education*, 2013, 34: 6-12 [苏芑, 罗燕. 技术神话还是教育革命——MOOCs 对高等教育的冲击. *清华大学教育研究*, 2013, 34: 6-12]

Exploration of educational empirical research based on big data from MOOCs platforms

LI ManLi¹ & HUANG ZhenZhong²

¹*Institute of Education, Tsinghua University, Beijing 100084, China;*

²*Online Education Office, Tsinghua University, Beijing 100084, China*

We reviewed numerous course reports and research papers generated through massive open online courses (MOOCs) and generally found two phases of MOOCs research, with a turning point in 2012, the so-called year of the MOOCs. Among these sources, research analyzing educational big data of MOOCs can be sorted into three groups: (1) Learning pattern and learner categories; (2) learners completing or dropping out of courses; (3) specific learning behaviors and different functions of the MOOCs platform. However, we find limitations in the existing data. This insufficiency is partly derived from the lack of conventional educational and psychological theoretical frameworks in MOOCs research based on big data from online platforms. We point out that researchers could use MOOCs platforms as valid data resources to further verify and to extend conventional educational research theories and frameworks, such as learning theories, pedagogy and educational policy. We also analyze different data types that MOOCs platforms offer and their roles in educational data mining (EDM). MOOCs data include user logs, learning paths and performance results. It is important in future research to combine conventional theories with options presented by new data. This paper presumes there will be three types of research based on big data from MOOCs platforms. Such research will mainly be on: the MOOCs platform itself; MOOCs design and delivery; EDM and learning behavior analytics.

big data, massive open online courses, educational research

doi: 10.1360/N972014-00983