

学习分析技术应用:寻求数据支持的学习改进方案

顾小清^{1,2} 刘妍² 胡艺龄^{1,2}

(1. 华东师范大学 上海数字化教育装备工程技术研究中心, 上海 200062;
2. 华东师范大学 教育信息技术学系, 上海 200062)

[摘要] 层出不穷的信息通信技术创造并支撑着丰富的学习模式,在产生大量教育数据的同时,教学模式的中心也逐渐发生变化。作为助学者、服务者,教师迫切需要了解支持服务与学生需求间的差异。借助数据支持的决策所提供的引导与帮助,成为目前学习分析技术应用于教学实践领域最有价值的目标。本文立足于此目标,通过阐明学习分析技术用于提高教学和学习效果的实践需求,以本团队开展的评估、预测和干预三类应用案例为载体,突破对数据的浅层理解和理论阐述,并总结可能的实践应用方向及关键技术和特点。研究成果一方面可为改进教学实践提供可操作的方法,同时也能教育研究者进行实验设计和课程设计提供参考。

[关键词] 学习分析;教育大数据;改进教学实践;案例研究

[中图分类号] G434 **[文献标识码]** A **[文章编号]** 1007-2179(2016)05-0034-12

一、改进教学的学习分析诉求

随着信息技术在教育领域的深入应用,在线学习平台、学习管理系统、自适应教学系统、教育云服务应用迅速崛起,相应的行为数据和学习数据以指数级增长。学习分析技术以为学习者提供个性化学习服务、为教学过程提供准确的改进策略为目标,成为技术增强学习(Technology - Enhanced Learning, 简称 TEL)领域又一重要技术(Ferguson, 2012)。在学的方面,学习分析技术能整合与学习者相关的背景信息、学习过程数据,运用多类分析方法和数据模型解释与预测学习者的学习表现(Lee et al., 2016),也能基于行为数据和学习轨迹分析,提供基于个性化需求的学习路径和适应性帮助(Jovanovic et al., 2008);在教的方面,学习分析技术可以用于评估和优化课程,为教师提供基于证据的针对性教学干预(Chen et al., 2005),以及更为深入的教学建议。可以看到,学习分析技术具有支持学生自我改进学习,支持教师优化教学的潜力(顾小清

等,2012)。因此,如何利用学习分析技术有效促进教学实践,以及如何借助成熟的理论与模型将其在教育应用中“落地”,成为研究者关注的热点。

自首届“学习分析与知识国际会议”(The International Conference on Learning Analytics & Knowledge, 简称 LAK)2011年在加拿大阿尔伯特省班芙市举办以来,迄今已六届,但无论是最初探讨学习分析技术的定义、内涵、相关理论(LAK, 2011; 2013),还是后来探寻理论发展与实践发展的交叉点(LAK, 2014),以及如何实现从大数据到大效应的影响扩大(LAK, 2015),该会议主题都围绕学习者的行为分析、结果评价等展开,不曾深入涉及教学实践分析。直到2016年,该会议才出现研讨教学分析与教师分析(Teaching and Teacher Analytics)方面的主题,专家开始关注基于数据的教学分析,以提供更好的教学干预与教师帮助。这一趋势显示,学习分析技术已逐步在教学实践及学习改进方面获得重视。然而,除了上述关于学习分析研究的会议及相关成果外,学习分析技术能够用于支持教学实践的

[收稿日期] 2016-07-12 **[修回日期]** 2016-08-20 **[DOI 编码]** 10.13966/j.cnki.kfjyyj.2016.05.004

[作者简介] 顾小清,博士,教授,博士生导师,华东师范大学教育信息技术学系,研究方向:学习科学与学习技术、计算机支持协作学习、学习分析;刘妍,华东师范大学教育信息技术系在读博士生;研究方向:学习分析技术、自适应学习和设计;胡艺龄,博士,华东师范大学教育信息技术系晨晖学者,研究方向:学习分析技术、计算机支持问题解决。

具体成果尚不多(金慧等, 2016)。来自实践的诉求——如何以学习分析的研究成果为改进教与学提供指南——越来越强烈。学习分析如何才能支持和实现教学实践的改进? 在教学中还有哪些可能的应用? 本文试图对此进行探究。

二、以学习分析改进教学的数据可能性

(一) 学习分析技术概览

学习分析技术作为从海量数据中提取隐含、潜在信息的工具, 能够通过对这些海量数据的处理获得对学习过程更全面且深刻的认识, 同时也能够为学习者提供更智能和个性化的教学辅助策略。2012年, 美国发布的《通过教育数据挖掘和学习分析促进教与学》报告指出, 大数据在教育中的应用包括两大领域: 教育数据挖掘和学习分析技术。教育数据挖掘作为一个新兴研究领域, 借助统计学、机器学习和数据挖掘等方法(如构建预测模型、学习者聚类分析和关系数据挖掘等), 分析教与学过程中产生的数据; 另一方面, 学习分析技术需要测量、收集、分析和报告有关学生及其学习环境、过程的数据, 涉及信息技术学、社会学、计算机科学、统计学、心理学和学习科学等多个学科, 其应用能直接影响教学实践, 促进个性化教与学的实现, 优化学习路径, 改进教育决策(Siemens & Long, 2011; Siemens et al., 2011)。两者区别主要在于: 教育数据挖掘的目的在于开发新的算法和模型, 学习分析技术则是应用已有的模式和方法理解学习过程和学习行为。特别是, 学习分析技术注重监测和预测学生的学习成绩, 及时发现潜在问题, 为教学过程提出有针对性的改进策略和教育决策, 比如通过仪表盘或构建图形化模型等方式对学习过程进行评估、追踪预测和分析以实现个性化学习(胡艺龄等, 2014)。可见, 学习分析技术是回归教育本质, 以学习者对象, 以教育数据驱动的方式改进教学实践、促进学习者个性化发展的方法和技术。

(二) 学习分析技术的数据需求

正如上文所述, 学习分析技术具有对数据进行总览性描述和分析、推动个性化学习的潜力。然而, 对于学习分析框架与技术流程的设计, 学者们的观点不同。西蒙斯(Siemens, 2010)认为, 学习分析的过程是以数据为核心的分析环路, 包括收集数据

(如学习者数据、智能数据、个人学习档案)、存储数据、清洗数据、整合数据、分析数据、数据可视化和执行(如预测、干预、教学改进)等; 布朗(Brown, 2011)将学习分析过程整合为数据收集、分析、学习者学习、报告反馈和干预指导五个步骤; 沙缇(Chatti et al., 2012)指出, 学习分析是将教育数据转化为有用的行为来促进学习的过程, 这个过程是循环、迭代的, 其模型包括数据收集和预处理、分析与采取措施以及后处理。

综合上述理论框架可以看出, 已有学习分析框架大多涉及技术方法和可用数据, 特别强调数据的重要性。西蒙斯(Siemens, 2010)就认为获取大量数据是学习分析产生的驱动力, 包括学习者发布的数据、智能语义数据和关联数据等; 弗格森(Ferguson, 2012)也提出, 大数据技术、教育、政治和经济因素是驱动学习分析发展的主要动力, 数据是学习分析技术的核心要素。

智能终端、社交软件和传感器等硬件技术在教育情境中的应用, 成为基于移动互联网的学习数据剧增的重要原因, 同时也在逐步改变着学和教的方式。首先, 学习管理系统、课程管理系统等学习技术系统存储了大量“线上数据”, 人们借助技术手段和应用软件可以跟踪和获取学生行为、学习轨迹等数据, 如 CourseSmart 公司与五个教科书出版商共同推出智能课程分析工具(Junco & Clem, 2015), 该分析工具能在学习者与在线文本互动时紧密跟踪学生活动, 并为教师提供相应数据的解释, 提供学生参与特定文本的评价; 其次, 社交网络中的学习行为也呈急剧增长态势, 为学习分析的精细化提供多元、多样本的数据; 此外, 课堂教学产生的可通过智能终端及传感器获取的“线下数据”, 记录线下学习行为、学习者自我报告考试测试等。

另一方面, 学习分析的受益者也对学习分析技术提出了更多需求。学习者需要数据分析为其提供反馈, 或为其推荐符合个人特征的学习内容、课程和活动(He, 2013); 或需要优化的学习路径、合适的学习帮助和建议及个性化、智能化的学习环境。教师希望能借助学习行为记录和数据分析的优势, 从不同维度对学生进行群体分类、学业表现评估和学业预测(Abdous et al., 2012; Huang & Fang, 2013), 或者获得及时的教学反馈和改进建议, 以提

升教学效果 (Leong et al., 2012; Worsley & Blikstein, 2013)。教育研究者也同样会关注数据来源、分析技术和应用, 从不同受益者角度提出对教育数据的需求及对其价值的期待。由此来看, 无论从学习分析理论层面, 还是面向服务和受益群体来看, 数据对于学习分析的重要意义毋庸置疑。

(三)“教育大数据”为学习分析带来的可能性

数据一般指通过科学实验、检验、统计等方式所获取的用于科学研究、决策为目的的数值 (胡德维, 2013)。“大数据”一词源于美国国家航空航天局 1997 年对数据的描述, 意为超级计算机所生成的海量、实时和全样本的数据。大数据, 一般被认为是由于信息获取技术、物理信息系统、互联网、物联网和社交网络等技术的突飞猛进, 所引发的数据规模爆炸式增长的产物。2011 年, 美国著名的麦肯锡咨询公司发布《大数据: 创新、竞争和生产力的下一个前沿领域》, 阐述了大数据概念, 详细列举了大数据的核心技术, 并分析大数据带来的可能性 (Bienkowski et al., 2012)。2012 年 5 月, 联合国发布大数据政务白皮书《大数据促发展: 挑战与机遇》, 全面分析了各国特别是发展中国家运用大数据促进社会发展所面临的机遇与挑战 (Pulse, 2012)。同年 9 月, 美国布鲁金斯学会 (Brookings Institution) 在《教育大数据: 数据挖掘、数据分析和网络仪表盘》一文中, 提出通过数据挖掘、数据分析和网络仪表盘等手段, 改进研究过程、评价机制, 进而使得以“教育大数据”挖掘学习者学习过程信息、深入认识和理解学习者学业表现、优化学习过程成为可能 (Manyika et al., 2011), 突破传统教学中依赖定期测试表现和教师主观性评价的局限。

由此可以看出, “大数据”已经得到各个领域的重视。特别是在教育领域, “教育大数据”已成为 2016 年度教育信息化十大热点词之一, 无论是学生的课堂言行数据还是线上平台终端保存的行为数据、测试数据, 都属于教育数据范畴, 成为“教育大数据”的来源。“大”数据不仅意味数据容量大, 还包括其获取数据快及结构多元化, 因此对教育领域的研究提出了新挑战, 如怎样获取来源广泛的教育数据, 如何挖掘以数字学习资源为载体的学习经历数据的教育价值, 这些都成为利用“教育大数据”进行学习分析研究的新方向 (顾小清等, 2014)。但是

海量数据为何受到教育信息化研究的“钟爱”, 或者说, “教育大数据”对学习分析技术的重要性如何都有待深入探究。

目前, 学习分析技术开始应用于教育教学过程, 尤其是随着可用数据量的递增、日益成熟的技术和精细的可视化分析表征设计, “教育大数据”在学习分析过程中扮演着越来越重要的角色。托巴拉等 (Tobarra et al., 2014) 认为学习分析技术能够有效地改善学习方法和路径; 教育数据为学习分析提供实现教育决策的可能性, 如学习过程记录的行为数据 (在线学习视频点击量、登录时长、页面跳转序列等) (Siemens, 2013; Wolff et al., 2013)、学习者参与社区讨论或者在线学习支持的过程性评价。此外, 数据更是成为改进课堂教学过程的重要依据, 如学生回答问题的次数、互动频率、作业时长和正确率等。维伯 (Verbert et al., 2012) 借助学生自我报告和 e-tutorial 系统中提取的学习行为数据, 构建学习者模型, 开展成绩预测和学习资源推荐等方面的研究。尼斯特等 (Nistor et al., 2014) 通过在线学习过程中学生的自我效能感数据预测学业表现。洛伦斯等 (Lourens & Bleazard, 2016) 利用学习者的历史数据 (如交互数据、家庭背景信息等) 预测辍学的概率, 并采取相应的干预措施; 根据学习者登录学习管理系统的时长和观看视频时长、参与活动和完成测试题目的数据, 建构学习者模型, 进而推测其通过率、辍学率。综上各类数据的应用与教学实践的改进, 无不表明教育大数据为学习分析的深入研究提供了多方面的可能性。

正如 2013 年新媒体联盟地平线报告 (高等教育版) 所述, 学习分析技术能够应用数据为教育系统的各级决策提供参考 (Johnson et al., 2013)。鉴于此, 在教育大数据背景之下, 学习分析技术如何落地到真实教学环境中为教学改进提供依据? 本文从教学中的具体问题出发, 从不同角度展示如何应用学习分析技术为改进学习提供依据或方法。

三、如何改进教学: 案例研究

以学习分析为教学实践的改进提供依据, 是本研究团队学习分析研究项目的核心研究目标。本文选取三个已完成的从不同角度开展学习分析的研究

案例,关注的研究目标分别侧重于评估、预测和干预,最终以改进教学为研究立足点。

案例一:以问题解决能力为核心的学习分析研究

1. 问题提出

复杂问题解决能力作为高阶思维能力的核心要素,如何对其进行真实有效的评价是教育工作者关注的话题(OECD, 2013)。对于问题解决能力,许多心理学家与行为理论学家将其作为一种思维与行为加工过程。不同学生使用不同学习策略,会得到不同的解决路径和问题解决速率。因此,真实有效的测评就需要对问题解决过程进行观察、评价。传统测评较多以总结性评价(如回答问题正确数)为主,缺乏客观、以学习分析技术视角进行的过程数据的实时记录与评价。本研究案例立足于问题解决过程的全景式观察,通过系统记录和收集过程数据,对问题解决能力的评价提供深层次的实践改进建议。

2. 学习分析需求

计算机模拟以其多重表征的优点,成为当前世界范围内辅助学习的重要工具。同样在复杂问题解决领域,计算机模拟的微观世界能很好地还原复杂问题情境,并以动态链接表征的形式加深学习者对概念的理解,最重要的是其基于技术支持的过程观察特征,使研究者不仅能够获得问题解决的结果及每次操作的时间、内容和编码含义,更能够获得关于问题解决过程的行为序列,如操作移动顺序、跳转次数、游历方式等。这些微观行为的观察有效记录了学生思维加工过程,还原学生问题解决的真实行为过程,但如何将行为数据转化为可执行的见解,还需要借助学习分析技术中的相关模型与技术进行分析。

3. 研究设计与分析方法

在基于学习分析技术的复杂问题解决能力测评中,研究设计主要考虑测试环境设计与评价模型设计两个维度。在本研究案例中,测试环境的设计在计算机模拟微观世界的基础上,融入问题解决相关的内在表征到外在表征的映射规则及真实问题解决行为→在线问题解决行为→知识建构的逻辑结构;评价模型设计采用基于认知结构的理论框架,实现对行动、技能、表现的多维度分析。

实验选取某小学 554 名三至五年级学生进行测

试,实验所用评测系统见图 1,此外,还配合发放了相关的人口学背景信息及问题解决行为的影响因素调查问卷。

在分析方法上,本研究整合了被试测试系统操作日志数据、测试结果和问卷统计进行分析,针对评价模型中的不同层次采用最适应的学习分析技术进行操作,包括统计分析、认知诊断、行为模式识别等。

4. 结果分析与应用

基于复杂问题解决评价模型的评测分析能够获得学生个体和整体水平等不同视角的分析结论。从个体视角来说,评测分析可以就问题解决表现、认知技能精熟程度及问题解决序列模式三个维度得出诊断结论。以复杂问题解决的表现水平为例(见图 2),评测分析可以获得某学生在认知水平、元认知水平、认知效率和总的问题解决表现。从整体水平诊断的视角来说,评测分析采用多种渠道获取数据以混合式分析方法对问题解决能力表现进行分析,如聚类对群体分类、认知诊断分析对技能掌握情况进行分析、序列模式分析法剖析知识建构过程(胡艺龄, 2016),能够获得对整体问题解决能力水平的评估,及行为模式的差异化对比。

总的来说,对被试的分析结果表明:1)被试整体的复杂问题解决能力处于中等水平,大多数学生的认知结构发展还在元认知增强期及认知结构整合期;2)学生的问题解决能力在性别上无显著差异,但在元认知水平上有差异;3)青少年时期,年龄是影响问题解决能力的显著因素,随着年龄增长认知效率与元认知水平都会提高。

案例二:基于在线学习平台的预测模型

1. 问题提出

传统课堂教学过程中,学习评价和干预多以总结性评价、主观的人为评价为主,缺乏客观的、以数据分析为依据的评价。LMS 平台的混合式教学能够捕获影响学生学习的关键行为和因素,建立影响学习成绩的学习者行为模型。LMS 系统拥有支持教学的资料 and 工具,包括课程管理、设计、学生互动、学习管理等功能,同时能够满足及时反馈和评价的需求(肖明等, 2010)。借助平台的行为数据和学生信息,学习分析技术能够预测学生能否顺利完成该门课程,以便为该学生提供及时的学习帮助以及必要的课程内容调整。本研究案例利用 LMS 混合式学

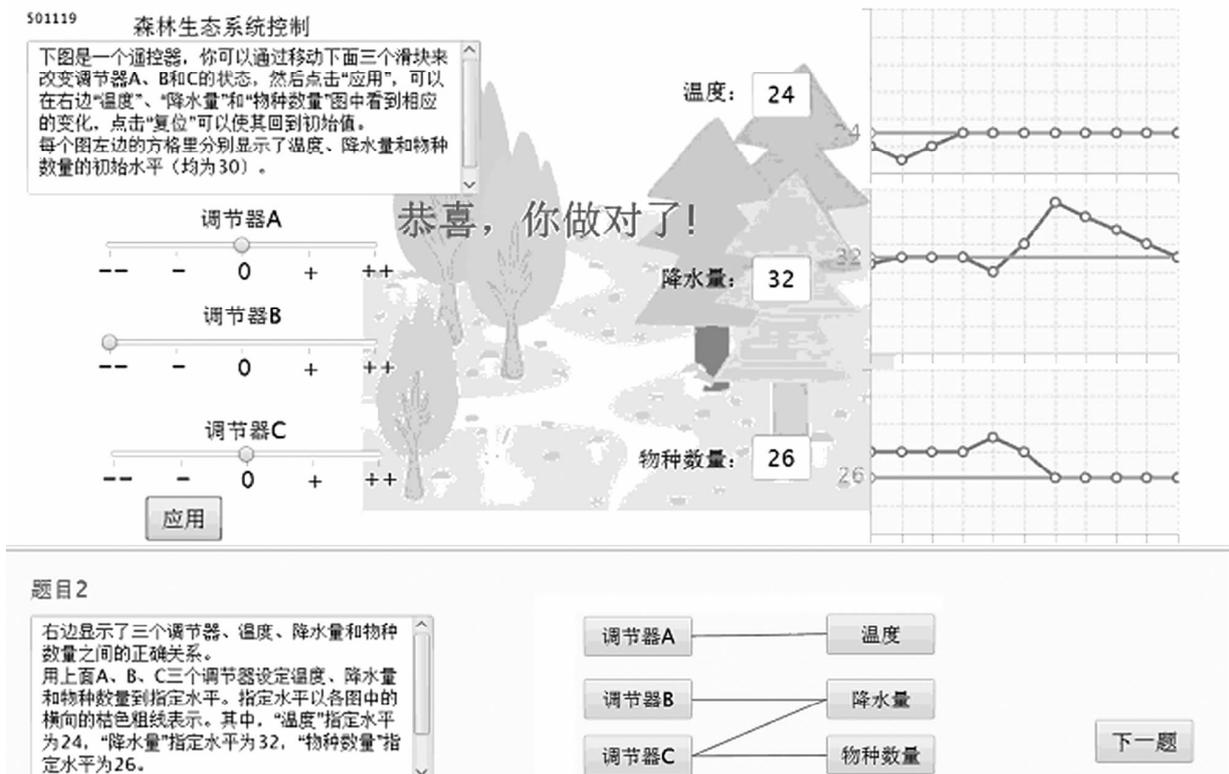


图1 问题解决系统界面

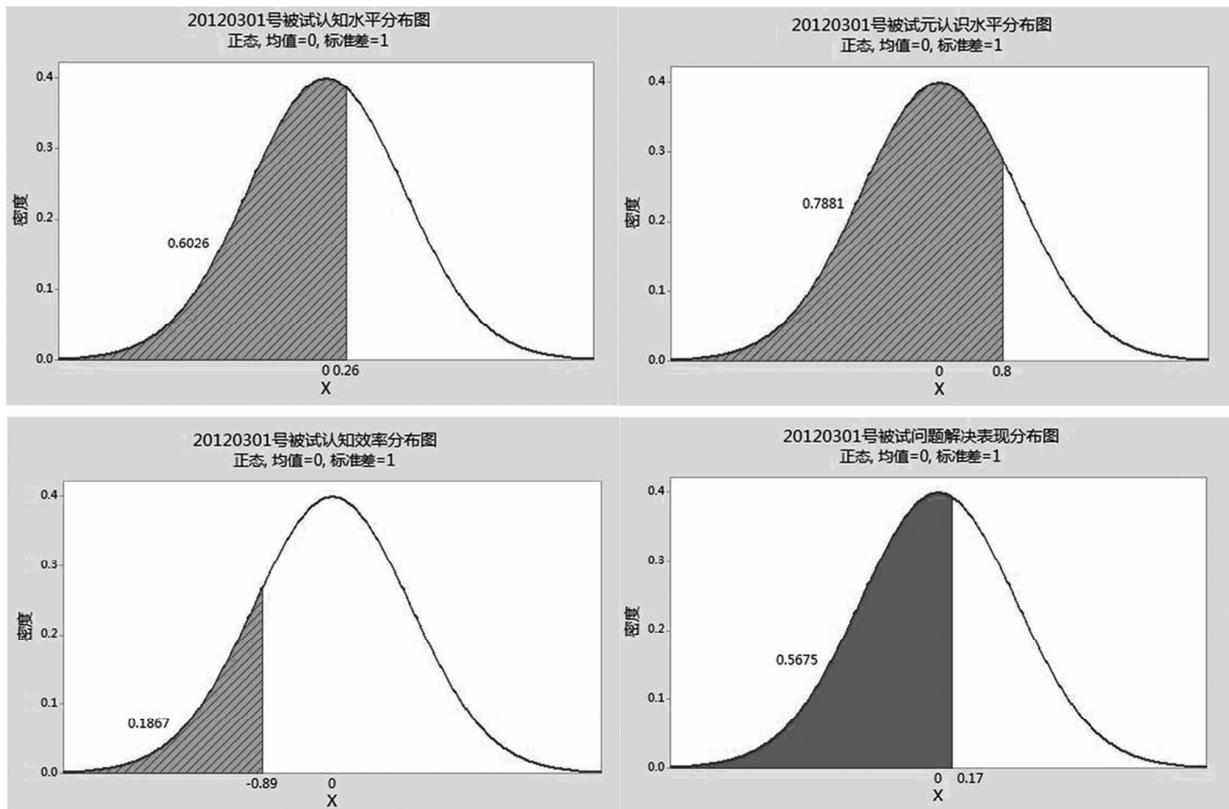


图2 问题解决分析结果

习的优势,就关键在线学习行为开展研究,预期达到提升学习效果和改进教学实践的目的。

2. 学习分析需求

混合式学习平台能够根据学生的学习活动状态,动态捕获其行为,记录停留时长、观看次数以及搜索何种资料等信息,并且能够持续记录学生的个人信息、社交信息、学习状态等静态数据,以提供丰富的数字化资源和系统功能支持。学习者能够根据反馈结果及时调整和控制学习的内容和步调。学生的学习效果很大程度上依赖于学生的自主性,但是也需要教师的关注和针对性评价;如果缺乏这些关注,学习者的参与度和主动性会有所下降。因而关注在线学习行为及其影响,还需要从学习分析视角收集数据、建立相关学习者模型进行分析。

3. 研究设计与分析方法

本研究案例的课程资料主要来源于教师制作的课件以及推荐的网络学习资源,同时提供大量习题与测试题。除每周两个课时的面授课外,学生还需要利用课余时间通过 Sakai 平台完成资源的浏览、作业和提交测试题目。与此同时,本研究跟踪用户在 Sakai 平台上的行为活动,包括学习态度(登录频率、资源访问频率、任务完成频率)、互动和协作(发布新帖数量、回复帖子数量、课程论坛讨论数量和阅读帖子数量)、资源应用(查看资源记录、应用查找功能频次、发送邮件数量)等,通过模糊理论结合层次分析法(Analytic Hierarchy Process, 简称 AHP)建立预测模型,确定学生的学业表层水平,实现对学生的学业表现水平的预判,达到提前发现“困难”学生

的目的。

某高校“C 语言”课程是该校计算机专业入门课程,修读学生的程序设计知识基础一般。实验选取大学本科一年级学生 49 名,运用学习分析技术将其行为数据、测试数据等进行统计分析后,从学习者的行为数据中抽取关键的在线行为,定义相关的指标体系结构,建构学业表现的预测模型,并与实际成绩进行比对分析,以检验模型的有效性。

4. 结果分析与应用

实验结果借助学生成绩与预测成绩比较曲线得出(见图 3)。预测模型的有效性是借助成对比较矩阵与模型的一致性比率计算得出,根据一致性指数(Consistency Index, 简称 CI)和随机指数(rRandom Index, 简称 RI)计算一致性比率(Consistency Ratio, 简称 CR),实验结果为 0.022,满足偏差分析。

此外,本研究选取教师的意愿和意图作为学生学业成就的影响因素之一。实验结果表明,与教师意图有关的跟踪变量和学习者在学习任务上花费时间(登录次数和在线时间)与其学业表现正相关,即教师意愿和在线学习时长,能够共同预测学业表现,但是预测所需变量的效度质量还取决于课程平台的设计和教学目标的优劣。

案例三:混合学习环境下学情的反馈干预

1. 问题提出

混合学习将面对面的课堂教学和在线学习有机整合,但是研究者通过一段时间的教学实践发现,学生线上学习的出勤率、参与度和通过率较低,学业成绩也不理想。相关研究表明,在线学习需要学习者



图 3 测试分数与预测分数比较

具有较强的动机和自主性。而学习动机和自我效能感很大程度上取决于个人兴趣、良好的学习习惯或对课程内容的认同感,进而可能影响学习投入以及最终影响学习成绩。因此,本研究案例借助学习分析技术从自我效能感视角进行探究和学习干预。

2. 学习分析需求

借助 Sakai 平台,研究者获取和收集学生在平台中的测试数据,利用学习分析技术进行处理,并以可视化方式为学生呈现分析报告。这一方面使学生增强对知识的掌握程度,提升自我效能感;另一方面满足教师掌握学生学习的需求,帮助其对学习资源和教学时间进行实时调整。

3. 研究设计与分析方法

在本案例中,混合学习呈现为每周安排一次线下面授,相关学习资料的查阅、作业和测试可在课后通过线上平台完成,由此收集的数据包括 Sakai 平台中的在线学习行为数据、不定期的作业表现数据以及作为总结性评价的期末考试数据。研究者定时对数据进行分析,借助 Tableau 分析工具得到可视化的诊断结果,并以邮件形式即时反馈给学生,进而帮助学生及时了解自身表现。整个干预过程在为期两个多月的课程学习期间持续进行,具体的邮件干预内容包括言语信息和图表信息(可视化分析)(见图4)。言语部分主要针对自我效能感设计劝导内容,依据成绩不同分别对应不同的鼓励言语。

研究采用自我效能感量表对实验组和对照组进行测量,以验证这种基于数据的干预是否会对学生自我效能感产生影响(简菁,2016)。案例选取学习“软件测试”课程的30名本科三年级学生,依据学生前一学期相关课程的表现及自我效能感前测结果,两组学生的初始能力和自我效能水平没有显著差异。研究设计利用实验法,采用学习分析技术对实验组(N=14)行为数据、测试数据进行分析反馈,对照组(N=16)则采用传统的言语激励提升自我效能感。研究人员每周收集在线数据和作业成绩,给实验组每位被试单独编辑和发送个人学习情况邮件,对照组无任何邮件反馈。

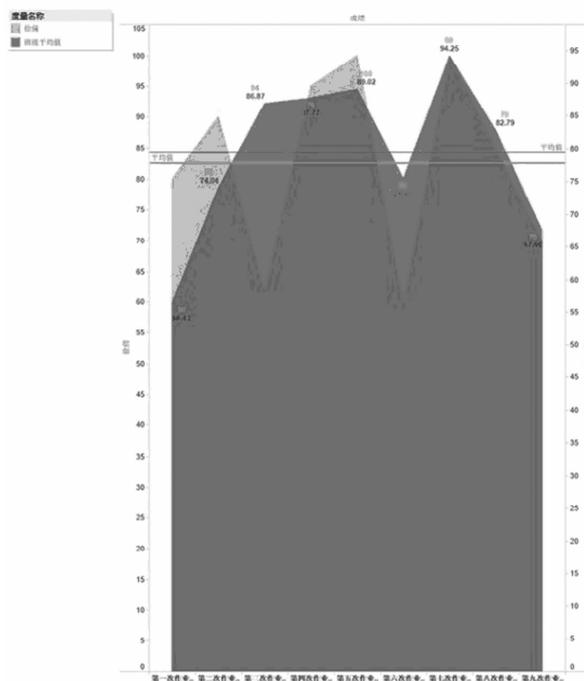
4. 结果分析与应用

研究结果表明,实验组的自我效能感得分显著高于对照组,且实验组学生提交作业次数逐步增多,提交时间提前,测试成绩也有所提升。该结果从一

《软件测试》学习情况

从作业的整体完成情况看,你的作业表现不太稳定,多数情况下能与班级水平持平,但最近这两次的作业表现一般。请继续努力,不要懈怠!

我的历次作业成绩与全班平均成绩对比



注:橙色为“我的成绩”,蓝色为“班级平均成绩”

橙色区域越多,说明我的成绩低于班级成绩

从左到右依次为:第1、2、3、4、5、6、7、8、9次作业成绩

2015 ECMO 华东师范大学 软件学院

图4 邮件反馈示例

定程度证实混合学习环境中对学生基于数据的学情反馈能有效提升其自我效能感,也为教师提升学生自我效能感提供了具体的教学改进意见。

上述三个实践案例,分别通过评估、预测和干预的方式为教学改进提供建议:案例一采用学习分析技术对问题解决能力水平进行分析,对问题解决过程序列进行实时记录,实现对信息加工层面的认知模式评估;案例二借助 Sakai 平台存储、收集学习过程数据,建构预测成绩的学生模型,并进行真实成绩与预测成绩之间的一致性比较;案例三利用学习分析技术对学习的阶段性结果进行可视化呈现,并以邮件形式反馈给学习者,进而提高其学习的自我效能感,最终实现以学习分析改进教学实践的目标。

这几个案例从不同角度进一步印证,全面、持续的数据能够使得教与学改进的指标更加精准和明确,为实践教学提供数据驱动的改进策略(顾小清等, 2012;魏顺平,2013)。

四、以改进教学实践聚焦学习分析研究

学习分析研究是一个运用复杂、多维的数据分析方法和工具,促进教学实践改进的新兴技术应用领域。美国高等教育信息化协会(EDUCAUSE, 2010)指出,学习分析关键应用之一是预测学生学业表现,为课程学习和学业困难的学生提供评估报告和干预措施(Johnson et al., 2011)。前文的案例同样以建模、评估和干预视角提出以学习分析改进教学实践的思路。综上所述,学习分析的应用主要体现在依赖庞大数据的分析技术,面向学习者学习行为过程,以评估、预测和干预为手段对其加以追踪、干预与改进(见图5)。

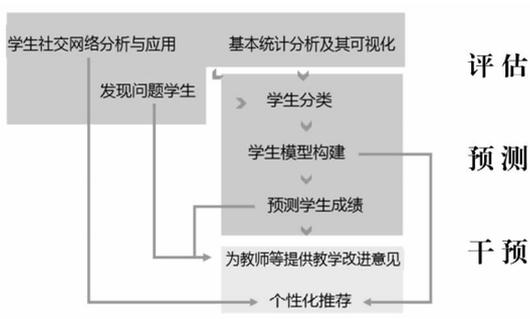


图5 学习分析的具体应用

(一) 实践应用

1) 评估指应用学习分析结果对研究对象的某些指标进行评价,具体应用包括基本统计、发现问题学生和社交网络分析与应用。基本统计方法能够突出、直观显示观测指标,利用可视化方式展示分析结果,帮助用户解读数据含义;应用异常点检测、关联规则、潜在反应模型等方法能对学习问题进行甄别,以提供个性化指导和检测活动中学生的异常行为;以学生社交网络分析为应用的实践,利用学生在线社交网络,能推荐学习资源或者分析社交网络模型,支持和辅助教育决策。

2) 预测指应用数据分析结果构建预测模型或为学业表现预测提供依据,重点关注学业成就或学业表现(West, 2012)。具体应用包括学生分类、学

生模型构建和预测学生成绩等。学生分类可运用聚类算法(Kizilcec et al., 2013)、分类算法(Dejaeger et al., 2012),针对关键指标对学生进行差异化处理,有助于个性化学习指导和组建协作团队。学生模型构建可利用贝叶斯网络、序列模式挖掘、关联规则和马尔科夫决策过程等,对学习者的知识、技能和学习风格等不同维度进行刻画。预测模型能对学生成绩、知识和技能掌握情况进行分层预测。

3) 干预指将以分析结果所发现的问题为依据进行有效的教学干预,具体包括为教师提供教学改进意见和个性化推荐。前者可采用关联规则、协同过滤和文本挖掘等方法,对数据建模并深度分析,为课程开发者、教师和管理者提供决策支持意见,以便更有效地组织教学资源、提高学生学习效率;后者一方面可基于序列模式挖掘、关联规则、决策树等方法,另一方面根据学生的历史行为数据和学习目标,向其推荐个性化的学习资源,提供个别化的帮助和指导。

4) 其他应用,包括课件自动开发、教育过程的设计和教学资源调度等。

(二) 关键技术

从上述应用可知,基本统计分析和学生模型构建贯穿整个学习分析过程。可视化的统计分析能够加强数据处理结果呈现方式的有效性,凸显学习过程中的问题。学生模型构建是学生分类和学习行为跟踪的依据,也是进行教学预测与干预行为的决策依据。因此,学习分析技术是利用数据挖掘工具、策略和方法,收集和分析学习过程中储存的信息与数据,对资源进行统计、运算、分析和可视化表征(见表一)。

在借鉴教育数据挖掘相关技术的基础上,学习分析的关键技术涉及系统建模、分类、聚类、关系挖掘、社会网络分析、话语分析及文本挖掘一系列方法。最常用的方式是分类,其次为聚类、回归(线性回归或者多元回归)和模式挖掘。正如表一所列,利用分类算法(神经网络、决策树、朴素贝叶斯等)可预测学生辍学率、通过率(Guo, 2010; Guruler et al., 2010; Kizilcec et al., 2013; Lykourantzou et al., 2009)。

综上所述,学习分析是一种应用技术,是一座连接提升学习者学业表现、预测成绩、挖掘潜在问题、推荐个性化资源和改进实践教学的技术“桥梁”,也

表一 学习分析关键技术

具体应用	数据分析技术	拟研究问题举例
基本统计分析	统计和可视化	统计最受欢迎的学习内容;学习时间分布模式;访问次数、点击次数
发现问题学生	异常点检测;关联规则;潜在反应模型;马尔科夫链	检测非典型的学习行为;发现学习能力较弱的学生
	分类算法(神经网络、决策树、朴素贝叶斯、支持向量机)	预测退学的学生数量;预测选课的学生数量
社交网络分析与应用	协同过滤	推荐学习资源;推荐学习路径等
	社交网络分析技术	向学生推荐学习伙伴;分析在线交流特征
学生分类	各种聚类算法	学习者特征分析
	社交网络分析技术	学业预警
学生模型构建	序列模式挖掘、关联规则、逻辑斯蒂回归	推断学习者的个性特征;自动构建学习者模型
预测学生成绩	回归(结合神经网络、支持向量机、决策树等)	预测学生成绩;预测学习时长;预测学生满意度
	基于规则的预测	根据日志等预测学生的成绩和表现,及时提供辅导
	贝叶斯网络	对学生知识结构建模,进而预测分数
提供教学改进意见	关联规则	发现学习行为间的关联性
	聚类;分类;协同过滤	估计测试题考察的能力和知识点
	文本挖掘	抽取网络学习环境中的语义信息
个性化推荐	序列模式挖掘	根据学生眼动设备检测注意力是否集中
	关联规则	根据上下文推荐适合的学习内容
	聚类;决策树;基于案例的推理文本挖掘	对学习内容文本进行聚类分析

是学习者、教师、教育组织者等之间的联接纽带,可应用不同关键技术将适合个人特征的学习资料、学习路径推荐给学习者,为教师反馈学生的学习行为分析结果和评测诊断结果,帮助教学组织、管理机构提高教育决策质量、优化课程评估流程。

五、结语

本文分别以评估、预测和干预取向的学习分析研究为案例,呈现了学习分析研究的问题提出、研究设计以及初步的研究发现。上述学习分析研究案例,均以改进具体教学实践为目的。在这些案例的呈现中,由于教学改进的落脚点不同,研究设计中所采用的学习分析方法与技术,以及分析的结果及其应用形式,相应地呈现出多样性。结合研究案例,本文试图进一步讨论学习分析技术对教学改进的实践应用,即以解决实际的教学问题为目的,所需要借助的学习分析技术,及其在“教育大数据”环境中数据

分析内容、关键技术等的应用。

随着“教育大数据”时代的到来,学习分析技术成为当前教育技术领域发展最快的技术之一。学习分析技术的研究目标是为了系统化监测与评估学习者在混合式学习环境下的学习过程,通过教学设计把创新技术和工具嵌入到教学过程中,改进教学实践。然而,作为一个相对前沿的领域,学习分析技术的发展与应用仍面临许多挑战和瓶颈:

1)数据来源问题。教育数据蕴含了巨大的分析价值,建立学生学习数据库和行为数据库存储学生与学习内容交互的过程,以及用户行为、个人特征信息,是开展学习分析的前提。然而,由于教育过程的复杂性以及学与教发生情境的多样性,获取完整、过程化的教育数据成为学习分析技术的重大瓶颈之一。

2)分析技术问题。如何挖掘半结构化、非结构化数据“背后”的含义,从“大数据”中利用分析技术

提取有价值的信息成为学习分析技术的另一个瓶颈。由于数据的劣构性和复杂性, 需要将多源异构的数据转化为可互操作的数据, 并通过复杂的、动态的处理和分析过程, 使得这些多源异构数据呈现出可读性和语义性。这一过程需要不断调整和优化数据的处理技术和策略, 是学习分析技术所面临的重大挑战。

总之, 学习分析是极具潜力和价值的教育研究领域, 其研究价值的充分体现, 还需要来自教学改进实践的声音和证据。如何解析出学习分析技术对教学实践的价值, 切实地将数据分析结果应用于教学过程, 有效地改进和指导学习, 要实现这些功能, 学习分析技术应用任重道远。

[参考文献]

- [1] Abdous, M., He, W., & Yen, C. (2012). Using data mining for predicting relationships between online question theme and final grade[J]. *Educational Technology & Society*, 15(3): 77-88.
- [2] Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief[J]. US Department of Education, Office of Educational Technology: 1-57.
- [3] Brown, M. (2011). Learning analytics: The coming third wave[J]. *EDUCAUSE Learning Initiative Brief*, 1(4).
- [4] Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thus, H. (2012). A reference model for learning analytics[J]. *International Journal of Technology Enhanced Learning*, 4(5-6): 318-331.
- [5] Chen, E., Heritage, M., & Lee, J. (2005). Identifying and monitoring students' learning needs with technology[J]. *Journal of Education for Students Placed at Risk*, 10(3): 309-332.
- [6] Dejaeger, K., Goethals, F., Giangreco, A., Mola, L., & Baesens, B. (2012). Gaining insight into student satisfaction using comprehensible data mining techniques[J]. *European Journal of Operational Research*, 218(2): 548-562.
- [7] EDUCAUSE (2010). Next generation learning challenges: Learner analytics premises [EB/OL]. <http://www.educause.edu/Resources/NextGenerationLearningChalleng/215028>
- [8] Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges[J]. *International Journal of Technology Enhanced Learning*, 4(5-6): 304-317.
- [9] 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.
- [10] 顾小清, 郑隆威, 简菁 (2014). 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[J]. *现代远程教育研究*, (5): 13-23.
- [11] Guo, W. W. (2010). Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction [J]. *Expert Systems with Applications*, 37(4): 3358-3365.
- [12] Guruler, H., Istanbulu, A., & Karahasan, M. (2010). A new student performance analysing system using knowledge discovery in higher educational databases[J]. *Computers & Education*, 55(1): 247-254.
- [13] He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining[J]. *Computers in Human Behavior*, 29(1): 90-102.
- [14] Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models[J]. *Computers & Education*, 61: 133-145.
- [15] 胡德维 (2013). 大数据“革命”教育[N]. *光明日报*: 10-19.
- [16] 胡艺龄, 顾小清, 赵春 (2014). 在线学习行为分析建模及挖掘[J]. *开放教育研究*, 20(2): 102-110.
- [17] 胡艺龄 (2016). 基于学习分析技术的问题解决能力评价研究[D]. 上海: 华东师范大学.
- [18] IBM (2013). Bigdata [DB/OL]. [2013-05-10]. <http://www-01.ibm.com/software/data/bigdata/>.
- [19] 简菁 (2016). 混合学习环境中以学习分析提升大学生自我效能感的研究[D]. 上海: 华东师范大学.
- [20] 金慧, 刘迪, 高玲慧, 宋蕾 (2016). 新媒体联盟《地平线报告》(2016 高等教育版) 解读与启示[J]. *远程教育杂志*, 34(2): 3-10.
- [21] Johnson, L., R. Smith, H. Willis, A. Levine, & K. Haywood (2011). The 2011 Horizon Report [EB/OL]. Austin, TX: The New Media Consortium. <http://net.educause.edu/ir/library/pdf/HR2011.pdf>.
- [22] Johnson, L., Becker, A. S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). The 2013 Horizon Report: Higher Education Edition [EB/OL]. The New Media Consortium.
- [23] Jovanovic, J., Gasevic, D., Brooks, C., Devedzic, V., Hatala, M., Eap, T., & Richards, G. (2008). LOCO-Analyst: Semantic web technologies in learning content usage analysis[J]. *International journal of continuing engineering education and life long learning*, 18(1): 54-76.
- [24] Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data[J]. *The Internet and Higher Education*, 27: 54-63.
- [25] Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses [A]. In *Proceedings of the third international conference on learning analytics and knowledge* [C]. ACM:170-179.
- [26] LAK (2011). 1st International Conference Learning Analytics & Knowledge [EB/OL]. <https://tekri.athabascau.ca/analytics/>.

- [27] LAK(2013). Dialectics in Learning Analytics[EB/OL]. <https://lakconference2013.wordpress.com/>.
- [28] LAK(2014). Intersection of learning analytics research, theory and practice[EB/OL]. <https://lak14indy.wordpress.com/>.
- [29] LAK (2015). Scaling up: Big Data to Big Impact [EB/OL]. <http://lak15.solaresearch.org/>.
- [30] LAK(2016). Enhancing Impact: Convergence of Communities for Grounding, Implementation and Validation [EB/OL]. <http://lak16.solaresearch.org/>.
- [31] Lee, M. J. W., Kirschner, P. A., & Kester, L. (2016). Learning analytics in massively multi-user virtual environments and courses[J]. *Journal of Computer Assisted Learning*, 32(3): 187-189.
- [32] Leong, C-K., Lee, Y-H., & Mak, W-K. (2012). Mining sentiments in SMS texts for teaching evaluation[J]. *Expert Systems with Applications*, 39(3): 2584-2589.
- [33] Lourens, A., & Bleazard, D. (2016). Applying predictive analytics in identifying students at risk: A case study [J]. *South African Journal of Higher Education*, 30(2): 129-142.
- [34] Lykourantzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques [J]. *Computers & Education*, 53(3): 950-965.
- [35] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity[DB/OL].
- [36] Nistor, N., Baltes, B., Dascalu, M., Mihail, D., Smeaton, G., & Truan-Matu, S. (2014). Participation in virtual academic communities of practice under the influence of technology acceptance and community factors. A learning analytics application[J]. *Computers in Human Behavior*, 34, 339-344. doi:10.1016/j.chb.2013.10.051.
- [37] OECD(2013). PISA 2012. Assessment and analytical framework: Mathematics, reading, science problem solving and financial literacy [DB/OL]. OECD Publishing <http://dx.doi.org/10.1787/9789264190511-en>.
- [38] Pulse, U. G. (2012). Big data for development: Challenges & opportunities[EB/OL]. New York: UN Global Pulse.
- [39] Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities[J]. *Computers in Human Behavior*, 31: 659-669.
- [40] Siemens, G. (2010). What are learning analytics [EB/OL]. <http://www.elearnspace.org/blog>.
- [41] Siemens, G. (2013). Learning Analytics: The emergence of a discipline[J]. *American Behavioral Scientist*, 57(10):1380-1400. doi: 10.1177/0002764213498851.
- [42] Siemens, G., Gaevi, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. J. D. (2011). Open learning analytics: An integrated and modularized platform[DB/OL]. Retrieved from <http://www.solaresearch.org/OpenLearningAnalytics.pdf>.
- [43] Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in learning and education[J]. *EDUCAUSE Review*, 46(5): 30.
- [44] Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics [J]. *Journal of Educational Technology & Society*, 15(3): 133-148.
- [45] 魏雪峰, 宋灵青 (2013). 学习分析: 更好地理解学生个性化学习过程——访谈学习分析研究专家 George Siemens 教授[J]. *中国电化教育*, (9): 1-4.
- [46] 魏顺平(2013). 学习分析技术: 挖掘大数据时代下教育数据的价值[J]. *现代教育技术*, 23(2):5-11.
- [47] West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards [EB/OL]. *Governance Studies at Brookings, US: Reuters*.
- [48] Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment [A]. *Proceedings of the third international conference on learning analytics and knowledge [C]*. 94-101.
- [49] Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment [A]. *Proceedings of the Third International Conference on Learning Analytics and Knowledge [C]*. 145-149.
- [50] 肖明, 陈嘉勇, 栗文超 (2010). 数据挖掘在学习管理系统中应用的研究进展综述[J]. *现代教育技术*, 20(9):127-133.

(编辑:魏志慧)

Linking Learning Analytics with Instruction Practices: Approach to the Data-enabled Research to Learning Enhancement

GU Xiaoqing^{1,2}, LIU Yan² & HU Yiling^{1,2}

(1. Shanghai Engineering Research Center of Digital Education Equipment, East China Normal University, Shanghai 200062, China; 2. Department of Educational Information Technology, East China Normal University, Shanghai 200062, China)

Abstract: *Learning analytics is fast growing as one of the leading-edge topics in learning science due to the dramatically accumulated data in education system. The wide spread ICT has resulted in the phenomenon of "big data" within various learning technology systems and may have great potential for education researchers to obtain an in-depth understanding of student learning behaviors and performances and to foster data-driven education improvement. One of the biggest challenges is how we can collect, analyze and present the appropriate evidence by integrating researchers' expertise and stakeholders' expectations.*

The purpose of this paper is to present a series of studies on learning analytics for improving learning outcomes in educational settings. These studies were designed to provide data-enabled learning enhancement with slightly different approaches. The first case was designed to enhance problem-solving outcomes, by a deliberately designed assessment approach. A three-stage approach was proposed to analyze problem-solving behaviors, abilities, and performances, and was applied to assess five hundred and fifty-four students' online problem solving in a primary school. The study reveals four clusters with distinctive problem-solving behaviors, abilities, and performances. The analysis method proposed in this study can also be used to provide interpretation of the problem-solving abilities for each and every student. The second case was designed to enhance online learning by developing a prediction model, from which the instructor can provide timely intervention to facilitate online students' learning behaviors. In this case, a comprehensive model of learning analytics was established, in which the explicit and implicit learning behaviors were both included as the main data source. Clustering analysis and decision tree analysis were applied to reveal students' characteristics and predict their performances. The research result would be of practical value to fully motivate learning and promote learning outcomes. The third case was designed to provide visualized results of learning analytics as a way to promote students' self-efficacy, given the premise that students' beliefs on their academic capabilities play an essential role in their motivation for achievement; as a result, the way that intervention was provided in this study could enhance learning by improving students' self-efficacy. As shown in the result, this study did hold the hypotheses that visualized learning analysis has an effect on promoting students' self-efficacy, and thus influence the learning enhancement.

By presenting the above research cases dedicated to enhance learning with learning analytics methods and techniques, this paper further discussed the possible ways to link the research of learning analytics with the practical needs of learning enhancement.

Key words: *learning analytics; education data; data-enabled research; learning enhancement*