

## 在线学习行为分析的维度探析——基于文献分析的视角

### Dimensions of Online Learning Behavior Analysis: Perspective based on Literature Review

戴静<sup>1\*</sup>, 石小恋<sup>1</sup>, 肖君<sup>2</sup>, 陈伟运<sup>1</sup>, 顾小清<sup>1</sup>

<sup>1</sup>华东师范大学教育信息技术学系

<sup>2</sup>上海开放大学

\*925613528@qq.com

**【摘要】**学习分析技术是构建适应性的、个性化的学习支持系统的重要手段，本文从学习分析的三大主要利益相关者——学生、教师、管理者的实际需求出发，基于对当前学习分析领域的文献进行的元分析，从“角色—问题—指标”三个维度对在线学习行为分析进行了探析，梳理出不同角色所关注的、可用学习分析技术解决的主要问题，归纳总结出解决相关问题的数据采集指标，并在已有的指标体系的基础上对指标进行了分类，以期为基于学习分析和教育数据挖掘的学习支持系统构建提供有力的指导。

**【关键字】**学习分析；角色；在线学习行为；数据指标

**Abstract:** Learning analytics is an important means of building the adaptive and personalized learning support system. Based on the demands of different roles and current literature review of learning analytics, this paper aims to analyze online learning behavior in three dimensions: roles, concerns and indicators, sorting out the major concerns of students, teachers and administrators during the process of online learning, summarizing the data indicators to address related issues. So as to provide strong guidance for the design of the learning support system.

**Keywords:** learning analytics, roles, online learning behavior, data indicators

## 1. 导论

随着在线教育的蓬勃发展和教育大数据时代的来临，人们对在线学习行为分析的关注度愈来愈高，试图发现和理解在线学习行为数据背后所隐藏的信息并进行有效利用。在此背景下，学习分析技术应运而生。从现有文献看，国外学者关于学习分析的研究主要集中在学习分析服务框架、分析方法、工具与可视化工具领域等；国内学者的研究则主要集中于综述和应用分析（王良周和于卫红，2015）。在综述性的研究中，主要涉及学习分析相关概念界定和历史溯源、国内外的研究现状梳理、应用领域和关键技术的分析等内容（吴永和等，2013）。应用分析类的研究多侧重于基于学习管理系统的信息采集与分析，学习分析系统的构建（马晓玲，2014），较少探讨学习分析理论和技术的实现框架来以用来指导构建学习支持系统的研究。本文主要从软件工程和实现的角度探讨基于学习分析技术的软件系统实现所需要考虑的用户实际需求的维度以及具体系统实现的数据采集的维度，用于指导后期基于学习分析的学习支持环境或者系统的具体实现。一方面，软件工程需求分析对产品起着决定性的作用，作为一种软件系统，基于学习分析技术的教学环境开发必须满足用户的要求，并根据用户的需求来完善系统的性能、功能以及设计。另一方面，需求分析更是对系统的后期开发具有一定的引导作用，可以让开发者明确好开发的方向并加以实施。通过合理的需求分析，才能更好的将系统的功能、性能总体概括出来成为具体的规格说明，为系统开发指明方向（王兰，2013）。因而在利用学习分析技术来构建学习支持系统的过程中也应充分考虑到不同用户的需求，提取出在线学习过程中不同角色所关注的主要问题，从而有针对性地利用学习分

析技术帮助他们解决这些问题，使学习分析技术更高效地发挥出其作用。

本研究主要从学习分析技术的实施主体和面向的用户所真正关注的主要问题及其实际的需求出发，去探讨学习分析可以为之解决的问题，以及要解决这些问题所要采集的数据指标。作为学习分析面向的用户，教学者、学生、管理者是学习分析的主要利益相关者（Brown，2011）。基于文献分析的视角，本文从“角色—问题—指标”三个维度对在线学习行为分析进行了探析，梳理出不同角色所关注的、可用学习分析技术解决的问题，归纳总结出解决相关问题的数据采集指标，并在已有的指标体系的基础上对指标进行了分类，以期为基于学习分析和教育数据挖掘的学习支持系统构建提供有力的指导。

## 2. 文献综述

目前，国内外已有学者通过不同的维度对在线学习行为进行分析，相关研究主要来源于两大领域——教育数据挖掘和学习分析。数据挖掘领域的相关研究主要从数据挖掘的技术（Romero & Ventura, 2007）、数据挖掘技术可解决的问题（Baker & Yacef, 2009）、教育系统（Romero & Ventura, 2010）等维度展开。学习分析领域相关研究的维度较为类似，主要有学习分析的框架（Siemens, 2010）、学习分析技术（孟玲玲、顾小清和李泽，2014）、学习分析可解决的问题（顾小清、张进良和蔡慧英，2012）、在线学习行为指标（杨金来，2008）等维度，而很少从软件工程设计的需求分析入手，对不同角色关注的问题以及解决这些问题所需采集的数据指标进行较详细的分析，因而在具体指导学习支持系统设计时存在不足。

## 3. 研究思路

本文将对该领域中高引用率的论文进行比较全面的归纳和分析，系统地梳理学习分析利益相关者所关注的、可用学习分析技术解决的主要问题，并对当前学习分析和教育数据挖掘领域中常用的数据指标进行了分类整理，形成从“角色——问题——数据”三个维度对在线学习行为进行探析的研究思路，可以使得我们在今后建构学习支持系统时更容易从整体设计的角度把握住问题的实质。本文的研究过程如图 1 所示。

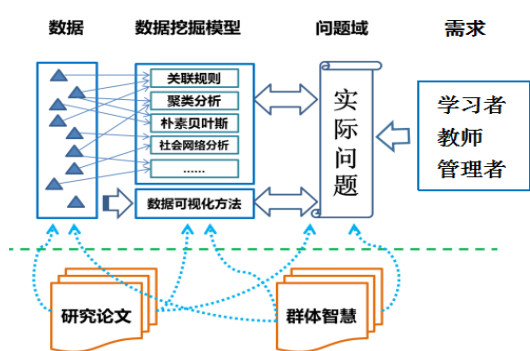


图 1 本文研究思路

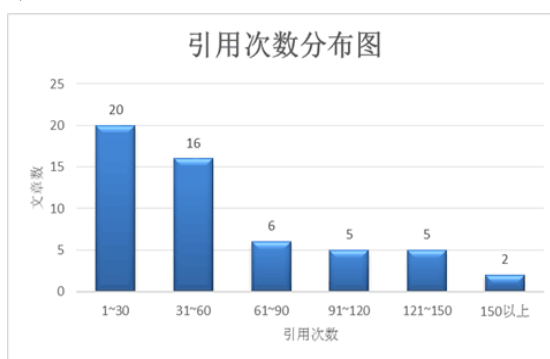


图 2 本文参考的文献引用次数分布图

本文回顾的论文主要来自三篇文献综述的参考文献（Romero & Ventura, 2007）（Romero & Ventura, 2010）（Peña-Ayala, 2013），共 54 篇论文，平均引用次数为 60 次，论文的引用次数分布图如图 2 所示。

## 4. 三大角色所关注的主要问题

表 1 三大角色关注的问题类别、对应的具体问题、关注相应问题类别的角色

问题类别	具体问题	角色	问题类别	具体问题	角色	
学习者特征	学习风格	教师	学习绩效	期中期末成绩	学生、教师、管理者	
	参与讨论的特征			个人知识状态		
	学生类别（根据对某学习对象的好恶程度或学生特征或浏览行为进行分类）			学生水平		
	学习动机			通过测试的可能性		
	学生的兴趣			预测期末成绩		
	认知方式		课程分析	课程分类（基于学生用户或课程内容本身分类）	教师、管理者	
	对待帮助的态度			课程资源的有效性		
	学生的积极程度			知识点之间的关联性		
互动行为分析	学生活动（或学习行为）类别以及频率	教师 （*备注：是否辍学也是管理者关心的问题）	教学技术	课程资源被浏览序列模式	教师、管理者	
	活动序列模式			课程的难点		
	小组讨论序列模式			教学技术有效性评估		管理者（*备注：有效的教学规律和模式也是教师关注的问题；个性化推荐也是学生关注的问题）
	小组任务完成模式			有效的网站建设		
	最常见的活动模式			网页之间的关联性		
	识别有意义的互动行为		有效的教学规律和模式*			
	不专注的行为		最优预测模型			
	鉴别学生在线测试中的欺骗行为		个性化推荐*			
	学生问题的层次					
	论坛话题类别（聚类）					
	当前学习状态					
	是否会辍学*					
各变量之间的相关关系，以推测学生的行为表现						

学习分析的利益相关者指的是学习分析生态系统的受益者，也可以认为是学习分析的实施主体、面向用户等（吴永和等，2013）。对教师、学生和管理者这三大角色的需求分析如下。

学习者主要关心的问题有：第一，知识点的掌握情况；第二，个性化推荐的内容，如学习活动、学习资源、学习路径等；第三，自己的行为表现，包括测试、作业、资源浏览、参与讨论、出勤、时间分配等，还有自己所学的知识之间的关系，即知识地图。

教师主要关心的问题有：第一，学生个体以及全班整体知识点的掌握情况，发现学生常犯的错误以及全班同学的学习难点，并给与相应的学生一定的帮助；第二，预测学生的表现，如期末成绩、测试表现等；第三，了解学生的兴趣、学习风格、学生的投入程度、学生的活跃程度、学生的进度、学生的浏览等；第四，关注课程内容的质量，一方面表现于课程内容本身的架构、知识点之间的联系、知识点与学习对象的联系等，另一方面表现于课程资源的利用情况以及通过课程资源的利用情况和学业成绩之间的关系得到课程资源的有效性。

本文的管理者包括校长等学校高层领导和网站管理员，校长主要关注的有：第一，学校整

体的进度、成绩、访问量等，可将不同班级、年级层次进行比较；第二，对课程、教师、学习者进行管理，这一点不容易体现；第三，制定计划与政策适应教与学并降低辍学率。网站管理员主要负责网站的建设，主要关注的有：第一，网站的架构、链接、预测模型的有效性；第二，自适应推荐的准确度，向学生推送合适的课程、学习路径、学习资源等。

结合以上需求分析，以学习者（特征-行为-结果）、课程、技术的视角出发，将三大角色所关注且学习分析可解决的问题整理为五大类别，分别是学习者特征、互动行为分析、学习绩效、课程分析和教学技术。问题类别、具体问题、关注相应问题类别的角色如表 1 所示。

## 5. 学习分析中的数据和指标

学习分析是实现个性化学习支持系统的先进手段，而数据和相关分析指标是进行有效学习分析的先决条件。系统化的学习行为分析指标能指导我们高效建立现有学习支持系统与已有成熟分析模型的关系，提高学习支持系统的智能化程度，提高构建智能化学习支持系统的效率。因此在分析不同角色需求的基础上，建立系统的在线学习行为指标体系是必要的。

目前国内比较著名的在线行为指标分类有以下几种：彭文辉等人提出的多维度、多层次的在线学习行为体系（彭文辉，杨宗凯和黄克斌，2006）。杨金来提出将在线学习行为信息模型描述为：Who Do What（杨金来，2008）。王丽娜把在线学习行为分成两大类——个性化交互行为、社会性交互行为（王丽娜，2009）。元帅将常见的在线学习行为分为面向学习结果评价的学习行为、面向协作学习的学习行为和面向平台使用的学习行为（元帅，2011）。李封把在线学习行为分为属性行为，静态行为和交互行为（李封，李婕和赵长宽，2014）。从学习分析的角度来说，以上这些已有的在线学习行为指标都存在着不同程度的问题。杨金来和王丽娜的分类从本质上来说是相似的，都将行为看成是一个系统，抓住了行为系统的各组成要素及其关系，然而却忽视了单个行为要素自身的特征和属性，也忽视了学习者的特征和属性。彭文辉等提出的多维度、多层次的在线学习行为体系中，每一类行为之间其实是相互联系的，没有明显的界限，不利于后期的学习分析。元帅提出的在线学习行为指标体系中，没有考虑到对学习者的特征、课程和技术属性方面的分析，而且面向协作学习的学习行为和面向平台使用的学习行为有很多重复的部分。而李封的在线学习行为属性模型中虽然考虑到了学习者，但是在线学习行为的归类不够完整和全面。

表 2 在线学习行为指标体系

在线学习行为	在线学习行为属性	在线学习行为	在线学习行为属性
静态属性行为	学习者特征	面向学习投入分析的行为	参与测试的行为
	课程属性		作业行为
	技术属性		交流讨论行为
面向学习结果分析的行为	作业成绩		资源利用行为
	测试的成绩		网站浏览活动
	期中期末考试的成绩		登陆行为

因此，本文在现有的在线学习行为指标体系的基础上，依据学习分析所要解决的问题，结合整理得到的在线学习行为数据，把在线学习行为分为三大类（如表 2 所示）：第一类是静态属性行为，包括学习者特征，课程属性，技术属性。第二类是面向学习结果分析的行为。第三类是面向学习投入分析的行为，包括参与测试的行为，完成作业的行为、论坛讨论的行为、资源利用的行为、其他学习活动的行为。

### 5.1. 静态属性行为

静态属性行为是指与行为主体和行为客体相关的一些基本属性，不涉及具体行为操作。这

些静态属性行为的分析是学习分析中的一个重要组成部分。包含以下几个部分：

### 5.1.1. 学习者特征

学习者特征是指学习者的背景、知识、兴趣、目标、个性特征等。学习者特征数据一方面可用于挖掘出学习者的特征，从而更好地指导学习者的学习，另一方面可用于对学习者的学习进行分组或者分类，从而预测每类学生的成绩，对每类学生进行个别化指导。分析学习者特征是实施个性化教学，促进个性化学习的关键步骤。

关于学习特征有关的数据有：学习者背景数据（年龄、性别、婚姻状况、有几个孩子、计算机能力、职业、数学成绩、工作是否与计算机有关等），学习风格类型的数据（学习风格数据可由量表和在线讨论数据估计得出），学习兴趣、学习起点等数据。

### 5.1.2. 课程属性

课程属性是指课程内容、课程资源、课程活动、课程作业、课程测试等课程本身的一些基本属性。课程属性会对学习者的学习结果和学习路径产生一定的影响，课程属性数据的分析有利于教师对课程活动以及资源进行有效的设计和组织的。

关于课程属性的数据有：课程类型，课程内容结构、完整性、难易程度，课程目标的合理性；每一个知识点对应的学习活动数量，学习活动的类型；本课程资源中文件数目，其他课程资源中文件数目，课程资源的可读性与易用性，课程资源类型（文本图像视频声音），资源的提供者；课程测试的有效性、针对性，有无测试项的答案解释；课程作业的要求等。

### 5.1.3. 技术属性

技术属性是指与在线学习平台，学习系统相关的功能特征。一个在线学习系统的功能会影响学习者在线学习的质量。技术的分析有利于获得在线学习系统功能的反馈，从而改进在线学习系统的设计与开发。

关于技术属性的数据有：网页类型（说明型、活动型、测试型），网页之间的关联性，页面设计质量，页面链接方式的合理性，用户界面响应的时间，安全机制的质量，反馈是否直接呈现，反馈推荐的强度，功能模块的类型，支持工具的类型等。

## 5.2. 面向学习结果分析的行为

面向学习结果分析是指依据教学目标和评价标准，采用一定的测量工具和方法对学生的学习结果进行描述与价值判断。对于不同的课程目标和课程内容，其学习结果评价的标准不同，所收集到的学习行为数据也很有差别。但是大体上包括作业、测验、期中期末考试三个方面的学习成果性数据。当然有的课程也将论坛发帖得分、回帖得分计入学习结果中。

表 3 面向学习结果分析的行为数据

一级指标	具体说明
考试成绩	在一门课程的中间或结束时进行的考试成绩，即期中或期末考试成绩。
测试的成绩	测试的成绩是指每次提交测试后获得的对应的成绩。
作业的成绩	作业成绩是指每次提交作业后得到的对应的成绩。

面向学习结果分析的行为数据有：作业成绩，测试成绩，考试成绩，如表 3 所示。一方面，从这三种学习成果性的数据中可以获得学习者知识点掌握情况，努力程度，在整个班级中的相对成绩等，有利于教师对学习者的学习作出评价和学习者的自我反思、学习策略的调整。另一方面，这些面向学习结果分析的行为数据可以与其他在线学习行为数据结合起来分析，如论坛参与度是否可以促进学习绩效、资源的利用是否影响学习成绩、在学习结果方面表现相同的学习者的作业行为有何异同，据此可以预测学习者的学习表现并有针对性地进行干预。

### 5.3. 面向学习投入分析的行为

表 4 面向学习投入分析的行为指标

一级指标	二级指标	具体说明	
登陆行为	登陆次数	学习者点击进入网站链接，账号登陆的次数	
	在线时长	学习者每次从账号登陆到退出登陆的时间间隔	
	登陆时间	学习者登陆的日期及具体时刻	
测验行为	测验提交量	学习者在线完成测试的数量	
	测验花费的时间	学习者每次完成测试花费的时间	
	测验通过率	学习者通过测验的次数占总测验次数的百分比	
作业行为	作业提交量	学习者完成教师布置的作业数量	
	浏览作业的次数	学习者点击指向具体作业内容的链接的次数	
	每次作业重复提交数	对于一次作业，学习者重复提交的次数	
	作业花费的时间	学习者每次完成作业花费的时间	
交流讨论行为	论坛讨论行为	论坛访问次数	学习者点击进入论坛的次数
		读帖数	学生阅读帖子数目
		回帖数	学习者回复他人帖子数目
		发帖数	学习者发帖子的数量
		被回帖数	学习者的帖子被回复数目
	聊天面板行为	在线交流的时间	学习者在聊天面板中在线交流的时间
		发言的次数	学习者在聊天面板中发言次数
	与教师交流的行为	与教师交流的次数	学习者与教师交流的次数
		交流的方式	学习者与教师交流的方式，如论坛、e-mail 等
	小组在线协作讨论行为	小组在线讨论的次数	学习者所在的学习小组在线讨论的次数
		每次小组讨论持续的时间	学习者所在的学习小组每次在线讨论的时长
	浏览行为	资源 ID	被学习者访问的资源标识号码
		资源访问时间	学习者从打开指向具体资源的页面，到关闭该页面的时间间隔
每个资源重复访问数		对于每个课程资源，学习者重复访问次数	
紧接着的页面内容		学习者关闭具体资源页面后，进入的下一个页面	

在线学习的过程中，教师与学生时空分离，在丰富的数字化资源以及各类学习支持系统的辅助下，学习者作为在线学习的主体具有很强的自主性——自定学习目标、学习内容、学习进度和学习方式。因此，教师、管理者、学生本人都对学习者的学习投入问题非常关注。面向学习投入的分析就是对在线学习情境下学习者进行自主学习的过程进行分析——如分析学习者选择的学习资源，学习活动参与度，作业完成过程中存在的问题，学习进度，学习路径等。由于在线学习的过程极其复杂，基于在线学习系统的功能，本节将面向学习投入分析的行为分为：登陆行为、测验行为、作业行为、论坛行为和浏览行为。具体分类如表 4 所示。

#### 5.3.1. 登陆行为

登陆行为数据是学生开始学习与结束学习最直接的表现，直观呈现了学生在线学习的出勤情况，以此了解学生登陆行为的特点如每天登陆时段偏好、每周登陆时间偏好、登陆时长等，是预测学习者的辍学率最直接的数据来源。

#### 5.3.2. 测验行为

参与测验行为数据是指学生在测验过程中能够一定程度反映学生状态及知识掌握情况的

Wu, Y.-T., Chang, M., Li, B., Chan, T.-W., Kong, S. C., Lin, H.-C.-K., Chu, H.-C., Jan, M., Lee, M.-H., Dong, Y., Tse, K. H., Wong, T. L., & Li, P. (Eds.). (2016). *Conference Proceedings of the 20th Global Chinese Conference on Computers in Education 2016*. Hong Kong: The Hong Kong Institute of Education.

行为数据，主要用于解决预测学生的期末成绩、了解学生的积极程度等问题（Pechenizkiy, Calders & Vasilyeva, 2008）

### 5.3.3. 作业行为

作业与测验相似，也是反映学生状态及知识掌握情况，测验更侧重于检测学生的学习水平，而作业更侧重于应用所学的知识或发现未理解的知识，一般作业行为发生的频率更高。作业行为的数据用于解决预测学生的期末成绩、辍学率、了解学生对待作业的态度及花费的精力等（Zafra & Ventura, 2009）。

### 5.3.4. 交流讨论行为

交流讨论是弥补在线学习中缺乏肢体语言与神情交流的重要方面，在线学习中的交流以异步交流为主。交流讨论行为数据可以反映学生的活跃程度，了解学习者特征如参与讨论的特征、话题偏好等，也作为预测期末成绩、了解学生目前知识水平的一个变量，还可从论坛互动情况了解学生合作信息等。笔者根据收集的数据，将交流讨论行为分为：论坛讨论行为，是在线交流讨论最主要的形式；聊天面板行为，是在线交流中的同步交流工具；与教师交流的行为，是获得有效反馈的重要方式，可用于探索多与教师交流能否提升成绩或能否提高学生学习的积极性等问题；小组在线协作讨论行为，可用于挖掘小组协作的特征或预测小组的最终表现等。

### 5.3.5. 浏览行为

浏览行为是面向学习投入行为的重要数据来源之一，通过浏览行为的分析可以解决的问题有：从浏览页面内容了解学生的兴趣；从学生浏览序列的代表路径或特点（如发现频率最高的活动模式）以及页面停留的时间，向学生推荐下一个浏览页面或浏览路径或一系列学习资源或确定个性化推荐的规则或页面之间跳转的概率（Wang, 2008）；登陆行为及浏览课程资源行为也是预测目前学生知识水平的一部分变量（Jong, 2007）；根据浏览过的、正在浏览的学习单元以及浏览学习单元的时间等来识别学生的学习风格（Chang, 2009）。总的来说，通过面向学习投入分析的行为数据，可以发现学习者在线学习的行为方式和规律，从而采取有效的措施促进在线学习高效率进行。面向学习投入的分析，有利于预测学习者表现，向学习者进行学习内容的推荐，检测学习者不良学习行为，进行学习者社交网络的分析；有利于依据学习者的行为特点和规律来更好地进行课程设计；有利于发现学生利用教育资源的行为方式和偏好，从而研究出满足学习需求的教育资源；有利于了解学习者使用在线学习系统的特点，为在线学习系统的设计和开发提供依据（Romero & Ventura, 2010）。

## 6. 总结

随着网络技术的发展，在线学习已经成为一种重要的学习形式。与传统学习相比，在线学习形式摆脱了时间与空间的约束，充分尊重学生的个性，有利于个性化学习的实现。个性化学习又依托于个性化学习支持系统，而新兴的学习分析技术是实现个性化学习支持系统的先进手段。在利用学习分析技术构建适应性的、个性化的学习支持系统时，充分考虑学习者、教师、管理者三大主要角色的需求，提取出在线学习过程中他们所关注的主要问题，有利于学习分析技术发挥出更巨大的优势。此外，系统化的学习行为分析指标有利于在线学习行为信息更精确、高效、实时的采集，从而可以提高后期分析的效率，促进主要问题的解决。本文从学生、教师、管理者三个角色的实际需求出发，以“角色-问题-指标”作为思路，对学习分析解决的主要问题和在线学习行为指标进行系统化整理，为今后基于学习分析和教育数据挖掘模型的学习支持系统的设计提供了较为系统的思考。

### 参考文献

马晓玲、邢万里、冯翔和吴永和（2014）。学习分析系统构建研究。《华东师范大学学报》（自

Wu, Y.-T., Chang, M., Li, B., Chan, T.-W., Kong, S. C., Lin, H.-C.-K., Chu, H.-C., Jan, M., Lee, M.-H., Dong, Y., Tse, K. H., Wong, T. L., & Li, P. (Eds.). (2016). *Conference Proceedings of the 20th Global Chinese Conference on Computers in Education 2016*. Hong Kong: The Hong Kong Institute of Education.

然科学版) , 2 , 002 。

元帅 (2011) 。在线学习行为分析评价及其应用研究 (硕士) 。华中师范大学。

王兰 (2013) 。提高软件需求分析质量的探讨。 *电脑知识与技术* , 23 , 026 。

王良周和于卫红 (2015) 。大数据视角下的学习分析综述。 *中国远程教育* , 3 。

王丽娜 (2009) 。网络学习行为分析及评价 (硕士) 。陕西师范大学。

李封、李婕、赵长宽和陈默 (2014) 。大规模网络课程中学习者学习行为分析。 *计算机教育* , 20 , 49-52 。

吴永和、陈丹、马晓玲、曹盼、冯翔和祝智庭 (2013) 。学习分析:教育信息化的新浪潮。 *远程教育杂志* , 04 , 11-19 。

杨金来 (2008) ) 。网络学习行为的实时监控研究与实践 (硕士) 。浙江工业大学。

孟玲玲、顾小清和李泽 (2014) 。学习分析工具比较研究。 *开放教育研究* , 04 , 66-75 。

顾小清、张进良和蔡慧英 (2012) 。学习分析:正在浮现中的数据技术。 *远程教育杂志* , 01 , 18-25 。

郁晓华和顾小清 (2013) 。学习活动流:一个学习分析的行为模型。 *远程教育杂志* , 04 , 20-28 。

彭文辉、杨宗凯和黄克斌 (2006) 。网络学习行为分析及其模型研究。 *中国电化教育* , 10 , 31-3 。

Brown, M. (2011). Learning analytics: The coming third wave. *EDUCAUSE Learning Initiative Brief*, 1-4.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining*, 1(1), 3-17.

Chang, Y.-C., Kao, W.-Y., Chu, C.-P., & Chiu, C.-H. (2009). A learning style classification mechanism for e-learning. *Computers & Education*, 53(2), 273-285.

Jong, B.-S., Chan, T.-Y., & Wu, Y.-L. (2007). Learning log explorer in e-learning diagnosis. *Education, IEEE Transactions on*, 50(3), 216-228.

Peña-Ayala, A. (2013). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4), 1432-1462.

Pechenizkiy, M., Calders, T., Vasilyeva, E., & De Bra, P. (2008). *Mining the Student Assessment Data: Lessons Drawn from a Small Scale Case Study*. Paper presented at the EDM.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146.

Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6), 601-618.

Siemens, G., & Long, P. (2010). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE review*, 46(5), 30.

Wang, F.-H. (2008). Content Recommendation Based on Education-Contextualized Browsing Events for Web-based Personalized Learning. *Educational Technology & Society*, 11(4), 94-112.

Zafra, A., & Ventura, S. (2009). Predicting Student Grades in Learning Management Systems with Multiple Instance Genetic Programming. *International Working Group on Educational Data Mining*.