

基于教务管理系统的数据分析工具的研究与设计*

邓文俊, 曹阳波, 刘英群, 程建钢

(清华大学 教育研究院 教育技术研究所, 北京 100084)

摘要: 基于大数据的学习分析受到越来越多的关注, 但能够直接用于教育教学实践的分析技术和工具则非常有限。随着教育信息化发展, 高校普遍应用了综合教务管理系统, 积累了大量的数据。该文收集并分析了教学管理者最迫切的数据分析需求, 选择了常用的数据分析方法, 设计并实现了一个基于教务管理系统的数据分析工具系统。该系统由统计分析工具集、数据计算方法模块、分析结果可视化模块三个部分组成, 其中统计分析工具集包含学生、教师、师生比、教师工作量和成绩分析等方面的七个常用工具。应用该系统对某高校教务管理系统中16年的数据进行分析, 结果验证了数据分析工具对于教务管理者的决策支持价值, 体现了嵌入教学系统的数据分析工具的优势。

关键词: 教务管理系统; 数据分析工具; 大数据分析; 决策支持

中图分类号: G434 **文献标识码:** A

一、引言

教育部《教育信息化十年发展规划(2011-2020年)》指出, 要加快学校管理信息化进程, 建立电子校务平台, 加强教学质量监控, 推动学校管理规范化与校务公开, 支持学校服务与管理流程优化与再造, 提升管理效率与决策水平, 提高办学效益, 支撑现代学校制度建设^[1]。教学管理在高校管理工作中居于核心地位, 教务管理信息化是教务管理信息化的重要组成部分, 能够促进教学管理过程的程序化、规范化和科学化, 保障教学工作稳定、协调、高效运转, 加快高校管理现代化的进程。

经过十多年的发展, 绝大多数高校都通过综合教务管理系统实现了教学管理信息化, 这些系统多年来积累了大量数据。这些数据涵盖了学生、教师、课程、成绩等多个方面, 不仅是教务管理的工作记录, 同时也是教务决策的重要依据。分析这些数据, 能帮助决策者更好地了解并明确教育实践的优势与缺陷, 从而确定需要改进的方向^[2]。

美国高校教育信息化协会(EDUCAUSE)也看好教学数据分析的前景, 其与美国新媒体联盟(NMC)合作的“地平线报告(Horizon Report)”预测, 教学

数据的分析技术将在未来两至三年内在教育领域广泛应用^[3]。但是数据分析技术的研究还处于起步阶段, 大部分的研究还是概念性的, 真正能供教学使用的分析工具还很少。数据分析工具分为两大类: 通用工具和专用工具。通用工具不是专门开发的, 而是将已有的分析工具直接拿来应用, 如SPSS等。这一类的工具需要将数据从系统导出, 并对数据做预处理, 这些都需要一定的专业技能, 不适合一般的教学管理者使用, 更适合研究人员; 专用工具是专门为分析教学数据而定制开发的, 这一类的工具能嵌入某个教学系统中, 与系统无缝连接, 可以让普通用户在系统中直接使用。目前这一类的工具非常有限。本文基于“清华教育在线”教务管理系统, 针对教育管理者最关心的问题, 设计并实现一个数据分析工具系统, 以便支持教务管理者对大量数据进行深度挖掘和分析。

二、数据分析的目标

设计数据分析工具的第一步是选择要分析的数据。而选择哪些数据, 关键在于数据的使用目的^[4], 因此设计数据分析工具集应从教务管理者关

* 本文系清华大学文化传承创新基金项目“高校教学质量信息管理信息化: 问题与策略研究”(项目编号: 2012WHQN016)成果。

心的问题出发。本文从应用“清华教育在线”本科教务管理系统的30多个院校提出的上百个需求中,提取出与数据分析有关的需求,整理出了若干教务管理者最关心的问题。

1. 学生情况

学生情况指从不同的维度对在校学生和已毕业学生的人数情况进行分析,例如民族、性别、生源等。按照不同民族、性别、生源地对在校学生人数分布进行分析,可以了解少数民族学生数、男女学生数、学生来源的构成比例等,从侧面反映教育公平情况^[5]。在已有维度的基础上,加上年级维度和时间维度,便能了解这些构成比例随时间的变化趋势。

2. 师资情况

师资情况指从不同的维度统计教师人数,分析结构分布。分析教师的年龄和职务,能了解各个层级教师的比例,分析教师结构是否合理;分析性别、民族的比例,能了解教师在性别和民族层次上的合理性;分析教职工总数中正式教师的数量与全部教职工数量的比,可以从中反映出教师队伍的稳定性。加上时间维度后,可看到这些结构随时间的变化。分析教师结构,对高等学校教师队伍素质建设具有一定的启发意义^[6]。

3. 教师工作量

大学教师工作量包含教学、科研和服务,它们共同构成了大学教师的学术生活。统计任课教师的上课时间,科学合理地计算大学教师工作量,能帮助解答“教学的数量和质量的关系如何”“教学与科研怎样结合”以及“报酬和声望如何分配”^[7]。一般以专任教师平均周课时作为工作量的计算方法,周课时数是衡量教师工作量的指标,应合理安排每位教师的工作量,以保证满足工作量的同时不使部分教师超负荷工作。

4. 师生比

师生比是指某特定的教育层次,在指定年份中的教师数与学生数之比。由于师生比表示平均一名教师所负担的学生数,所以人们往往把师生比作为衡量办学效益的指标之一^[8]。分析师生比的数量比,可以从所得比例数中清晰地了解是不是存在不合理的超编教师,或者教师数远不能满足正常教学的需求等情况,以便及时增减教师,满足正常教学的需要。

5. 成绩分析

成绩是评价高校教学质量的重要依据,也是评估学生对所学课程掌握程度的主要标志。对影响学业成绩表现的因素进行详尽分析,能在此基础上提出有针对性的教育质量改进与提高的政策建议^[9]。

成绩分析有两个层次,第一层是一般分析,例如课程成绩的平均分、方差,统计学生成绩分数段等;第二层是深入分析,分析学生某些特征与学习成绩之间的相关性,挖掘课程成绩之间的关联性,反映学生随时间变化的学习成绩特征等。

三、数据分析的方法及工具系统的设计

教育统计作为一种研究方法,是以数理统计学的一般原理和方法,对教育实践中所获得的数据进行整理、分析与解释的工作^[10]。统计方法是进行定量教育管理 with 教育定量研究的主要手段。本文归纳了高校教务管理者最关心的问题,从中提炼出分析这些问题所需的统计分析方法,作为设计数据分析工具的依据。常用的统计方法包括描述统计、皮尔逊积矩相关系数、t检验、方差分析、时间序列分析等。

(一) 系统的结构框架

本文将数据分析工具系统嵌入原有的教务管理系统中,成为其有机组成部分,其结构框架如图1所示。数据分析工具系统由统计分析工具集、数据计算方法模块、分析结果可视化模块三个部分组成,其中统计分析工具集包含学生、教师、师生比、教师工作量和成绩分析等方面的七个常见工具。数据计算方法模块完成对统计分析工具的技术实现,分析结果可视化将分析结果以图表的方式呈现出来。数据分析工具系统的数据来源于教务管理系统的数据库,与教务管理系统其它业务管理模块相互独立,修改业务管理模块,并不会影响数据分析工具系统。

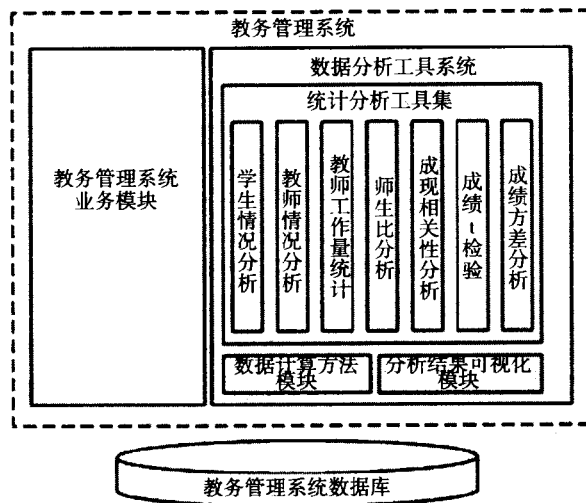


图1 数据分析工具系统的结构框架

统计分析工具集中包括七个分析工具:(1)学生情况分析:在一定范围内,按多种维度统计学生数量,例如按生源地、性别统计新生的数量等;(2)

教师情况分析：在一定范围内，按多种维度统计教师数量，例如历年来高级职称教师的比例；(3)教师工作量分析：在一定范围内，统计教师的周课时，例如按学院统计教师工作量的标准差；(4)师生比分析：在一定范围内，统计师生比的均值、方差等情况，例如统计某学院的师生比；(5)成绩相关性分析：在一定范围内，统计学生成绩绩点与特定因素的相关性，例如学院的师生比与该学院学生成绩的相关性；(6)成绩t检验：在一定范围内，统计两组学生的学生绩点是否有显著的差别，例如男女生成绩是否有差别；(7)成绩方差分析：在一定范围内，统计多组学生的学生绩点是否有显著的差别，例如来自不同地缘学生的成绩是否有差别。

数据分析工具的使用流程如图2所示。首先选择统计范围及数据。以学生情况统计为例，可以统计全校的学生，也可以统计某个学院的学生。其次选择维度，设置统计分组。以学生统计为例，可选的维度包括：院系、性别、民族等。再次选择内容，学生情况统计和教师情况统计不用选择内容，因为内容就是计算数量。成绩相关性分析、成绩t检验和成绩方差分析也不用选择，因为统计内容是固定的。师生比分析可以选择均值、标准差等描述统计的方法。最后选择结果呈现方式，有一维表、二维表、柱状图、饼状图等多种形式可供选择。

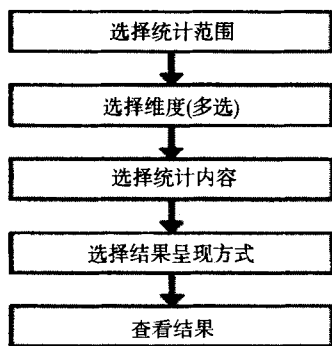


图2 数据分析工具的使用流程

(二)关键技术问题和解决方法

数据分析工具的技术开发涉及到两个关键问题，即选择数据计算方法及技术实现、分析结果可视化。

1.数据计算方法及技术实现

数据的计算方法包括平均数(Mean)、中位数(Median, Md)、皮尔逊积矩相关系数等的实现。计算方法有两种实现方式：数据库实现和Java实现。

本文中的教务管理系统使用Oracle数据库，它提供了分析函数。分析函数是Oracle 8中引入的一

个概念，为分析数据提供了一种简单高效的处理方式。分析函数共有32个，包含了描述统计中的所有方法、相关分析、方差、回归分析等多种统计方法^[11]。分析函数执行效率高，而且使用分析函数可以直接在数据库中计算平均数、相关系数等统计值，与Java的实现方式相比，不用传输原始数据，节省了大量时间^[12]。

Java有开源的统计方法库，例如Apache Commons Math、JStats等，这些方法库中包含了常用的统计方法。以Apache Commons Math为例，其包含描述统计、频率分布、二元回归、t检验和方差检验^[13]。与数据方法相比，Java的统计方法库支持更多的统计分析方法。

本文采用两种结合的方式，以Oracle分析函数为主，以Apache Commons Math为补充。若Oracle分析函数满足需求，则使用分析函数；若不满足，则使用Apache Commons Math。

2.分析结果可视化

分析结果可视化旨在借助于图形化手段，清晰有效地传达信息。教务管理系统一般采用Web页面展示数据，可视化结果也需在Web页面上呈现，并且需要设计成可交互的可视化数据，以方便对数据的动态处理。将统计结果以图表的形式呈现，有三种技术方案。

(1)Java方案。使用Java开源图表生成库，例如JFreeChart，在java中生成图表文件，然后在网页上呈现。这种方案的优点是运行效率高，图表的文件格式丰富，缺点是图表不能交互。

(2)Flash方案。利用Flash技术呈现图表，例如Open Flash Chart，优点是图表效果好，交互强，是图表表现力最强的一种技术。其缺点是，Flash需要的系统资源多，容易造成浏览器死机。

(3)JavaScript方案。使用JavaScript直接在浏览器中呈现图表，这是目前比较主流的做法。其优点是，绝大多数浏览器都支持，图表可交互，占用系统资源相对较少。而且，有很多开源的基于JavaScript的可视化图表库可供选择。

本文的工具采用JavaScript方案，设计了一组数据可视化API，将数据可视化的图形定义与图形呈现的实现解耦。在图形呈现实现部分，采用开源的数据可视化技术，既可以减少对某种开源技术的过度依赖，而且当出现新的数据可视化技术时，也可迅速切换。选择开源JavaScript可视化图标库时，选择了gRaphaë，因为其使用相对宽松的MIT许可证，库的体积小且满足绝大多数可视化需求。

四、案例分析与讨论

(一)案例分析

本文以北京市某高校为案例,对其教务管理系统中的数据进行分析。该校是一所以工科为主,具有鲜明工程实践特色的北京市属普通高等学校。教务管理系统中存有1997-2012年共16年的数据,每年数据都有变化,2012年的系统中包括26558条学籍记录数据,其中本科生25418人,专科生1140人。采用本文设计的数据分析工具集中的本科综合统计工具来分析高校本科教务的综合情况。

1.历年女生在总学生比重变化趋势

一般认为,多数理工科专业女生很少,男女比例严重失调。本文选取了1997-2012年学生数据,分析女生人数占总学生数比例的变化趋势。分析结果显示,女生比例从2002年起稳定在42%左右,只有三年出现了大的跌幅:2003年约33%,2006年约22%,和2008年约31%,表明该校男女生比例失衡的情况没有想象得那么大。

2.高考成绩和大学学习成绩相关性分析

高考成绩与大学成绩相关性的研究由来已久。随着招生规模的扩大,这一问题仍然具有重要的研究意义。而大学学习是一个较为长期的、动态的过程,只有对所有学年的分析才能揭示出其变化的规律^[14]。本文选取2006级全体学生高考成绩与大一、大二、大三及四年必修课平均学绩来分析相关性,分析结果如表1所示。

表1 2006级学生高考成绩与四年必修课学绩相关分析

学生范围	样本总数(个)	实际统计人数(个)	相关系数
四年	全校	3441	0.192***
	学院	290	0.245***
大一	全校	3441	0.271***
	学院	290	0.301***
大二	全校	3441	0.232***
	学院	290	0.168**
大三	全校	3441	0.231***
	学院	290	0.25***

注:*表示显著性在0.05水平,**表示显著性在0.01水平,***表示显著性在0.001水平。

结果表明,在学院和学校两个层次上,四年必修课的总学绩和高考成绩显著正相关;大一、大二和大三每年必修课的学绩和高考成绩也显著正相关。

3.性别对学习效果的影响

本文选取了一个工科学院和一个文科学院,前者男生多,后者女生多,分析结果如表2和表3所示。综合四年必修课总学绩和大一必修课学绩两种情况分析结果可以看出,女学生学习成绩整体好于男生。

表2 某两个学院四年必修课学绩性别差异显著性研究

	组别	算数平均值	学生数量(个)	t	sig.(双侧)
某工科学院	男	73.573	208	-9.324***	1.54E-17
	女	80.192	82		
某文科学院	男	73.876	143	-8.532***	3.03E-16
	女	78.259	268		

注:***表示显著性0.001水平。

表3 某两学院2006级学生大一学年必修课平均学绩性别差异分析

	组别	算数平均值	学生数量(个)	t	sig.(双侧)
某工科学院	男	72.215	208	-7.623***	5.56E-13
	女	78.537	82		
某文科学院	男	72.047	143	-5.045***	8.80E-07
	女	76.915	268		

注:***表示显著性0.001水平。

4.分数膨胀问题

所谓分数膨胀指学生的学业分数的增长超出其实际能力的增长。现实的中国高等院校是否真的存在分数膨胀?究竟在什么层面和程度上存在分数膨胀?这是需要实证检验的一个有趣的问题^[15]。本文选取某专业2001级到2008级学生四年必修课的平均分,结果如图3所示。

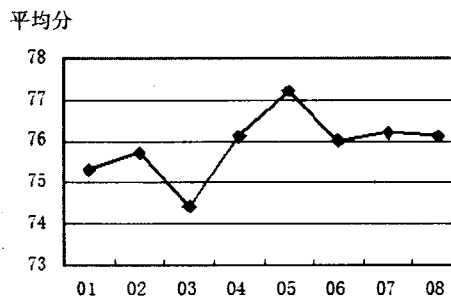


图3 某专业四年必修课平均分变化趋势

学生四年必修课平均分整体没有呈现明显上升的趋势。但是2003年到2005年整体呈现上升趋势。为了验证2003、2004和2005年学生四年必修课平均分是否有显著差别,对这三个年级再次进行方差

分析,如表4和表5所示。

表4 某专业2003、2004和2005级学生四年必修课平均分方差分析

	平方和	df	平均差	F	sig.(双侧)
组间差异	299.570	2	149.785*	5.813	0.004
组内差异	5076.570	197	25.769	—	—
总计	5376.140	199	—	—	—

注: *表示显著性在0.05水平。

表5 某专业2003、2004和2005级学生四年必修课平均分差异分析(Tukey HSD)

(I) 年级	(J) 年级	平均差(I-J)	标准误	sig.(双侧)
2003(83)	2004(58)	-1.766	0.868	0.107
	2005(59)	-2.880*	0.864	0.003
2004(58)	2003(83)	1.766	0.868	0.107
	2005(59)	-1.114	0.938	0.462

注: 1. *表示显著性在0.05水平; 2. 年份后括号内表示该年级学生人数。

由表4可以看出,该专业2003、2004和2005级学生四年必修课平均分存在显著差异。对三组样本进一步进行两两分析,结果如表5所示,2003级和2005级学生四年必修课平均分有显著差异,其他组无显著差异,因此该校并未出现分数膨胀的现象。

(二)讨论

1. 分析结果的讨论

通过对北京市某高校教务管理系统中16年全量数据进行挖掘分析,发现了一些有决策参考价值的信息,如:一个工科为主的院校十年来男女生的比例趋于均衡,四年间每年的必修课成绩与高考成绩显著正相关,在各个学院中女生学习成绩普遍好于男生等。

2. 数据分析工具的讨论

本文设计的数据分析工具系统能够满足教务管理者常用的数据分析需求,支持教务管理者进行科学决策,同时从使用性能来看,与SPSS等其它通用统计分析软件相比,有两大优势:

(1)对使用者统计知识和技能要求低。系统中的工具已针对特定教务问题深入定制,使用者无需太多的统计知识和技能,可像使用教务管理系统普通功能一样使用工具集。而SPSS等系统要求使用者有相对水平的统计知识和技能。

(2)使用简便。系统嵌入教务管理系统中,可直接使用。而SPSS等系统,用户至少要多做三件事:数据导出、数据预处理、数据导入。每一项工作都需要一定的技能,并且工作量很大。

五、结论

本文基于已有教务管理系统,提出了一种嵌入式教学数据分析工具的设计方法,构建一个数据分析工具系统,在大数据挖掘与分析工具方面做了一

定的探索。应用这套系统分析了某高校教务管理数据,验证了数据分析工具对于教务管理者的决策支持价值,体现了嵌入教学系统的数据分析工具的优势。下一步的研究将着力于该系统在其它院校中的推广应用,同时收集管理者的使用反馈和新数据分析工具的需求,进一步扩展教务管理系统中的数据分析方法。

参考文献:

- [1] 教育部. 关于印发《教育信息化十年发展规划(2011-2020年)》的通知(教技[2012]5号)[DB/OL]. <http://www.moe.gov.cn/rubicfiles/business/htmlfiles/moe/s3342/201203/xxgk—133322.htm>, 2014-02-04.
- [2] Wohlstetter P, Datnow A, Park V. Creating a system for data-driven decision-making: Applying the principal-agent framework[J]. *School Effectiveness and School Improvement*, 2008, 19(3): 239-259.
- [3] Johnson, L., Smith, R., Willis, H., & Levine, A., & Haywood, K. NMC horizon report: 2013 higher education edition[DB/OL]. <http://www.nmc.org/pdf/2013-horizon-report-HE.pdf>, 2013-12-23.
- [4] 顾小清,黄景碧,朱元锟,袁成坤. 让数据说话:决策支持系统在教育中的应用[J]. *教育科学文摘*, 2010, 29(6): 79-80.
- [5] 王善迈. 教育公平的分析框架和评价指标[J]. *北京师范大学学报: 社会科学版*, 2009, (3): 93-97.
- [6] 蒋伟. 高校教师队伍素质建设现状及改革对策分析[J]. *高等教育研究*, 2004, 25(3): 70-73.
- [7] 程晋宽. 美国大学教师工作量浅析[J]. *高等教育研究*, 2005, 26(6): 102-107.
- [8] 陈敏. 高校办学效益与师生比[J]. *高等教育研究*, 1991, (2): 62-76.
- [9] 张咏梅,郝懿,李美娟. 教师因素、学生因素对学生学业成绩影响的实证研究——基于大规模测验数据的多层线性模型分析[J]. *教师教育研究*, 2012, 24(4): 56-62.
- [10] 葛军. 教育统计在教育决策中的应用[J]. *统计与决策*, 2007, (247): 167-168.
- [11] Lorentz D, Jenkins B. Oracle database SQL language reference, 11g release 1 (11.1)[DB/OL]. http://docs.oracle.com/cd/B28359_01/server.111/b28286/toc.htm, 2014-01-02.
- [12] Chatziantoniou D. Using grouping variables to express complex decision support queries[J]. *Data & Knowledge Engineering*, 2007, 61(1): 114-136.
- [13] Commons A. Commons-math: The apache commons mathematics library[EB/OL]. <http://commons.apache.org/proper/commons-math/>, 2014-01-02.
- [14] 丁澍,缪柏其,叶大鹏. 高考成绩与大学成绩的相关性分析[J]. *中国大学教学*, 2008, (11): 29-31.
- [15] 何山,丁小浩. 从对分数膨胀现象的分析看高等院校应对劳动力市场的压力[J]. *清华大学教育研究*, 2004, 25(6): 35-42.

作者简介:

邓文俊: 博士后, 研究方向为数字化教育教学管理的研究与实践(edanwade@163.com)。

曹阳波: 在读硕士, 研究方向为数字化教育教学管理的研究与实践。

刘英群: 博士, 高级工程师, 研究方向为数字化教育

教学管理环境的理论与实践、基于知识管理的数字校园研究与实践。

程建钢：博士，研究员，博士生导师，研究方向为

数字校园理论与实践，网络教学理论、模式与系统构建，信息社会与高等教育，知识管理在高等教育信息化中的应用。

Research and Design of Data Analysis Tool System

—Based on Educational Management System

Deng Wenjun, Cao Yangbo, Liu Yingqun, Cheng Jiangan

(Educational Technology Institute, Tsinghua University, Beijing 100084)

Abstract: Learning analysis for big data has attracted more and more attention, but practical analysis tools are very seldom. With the development of information technology in education, more and more universities implement educational administrative management system, and these systems have collected a large amount of data. In this paper, the authors collect and analyze the most important requirement of educational data analysis from education administrators, design and implement a data analysis tool system based on an educational administrative management system. The system consists of a set of statistical tools, a set of analytical methods, and a data visualization system. We analyze last 16 years education data of a college with the system, and the result proves the value of data analysis tool system for education decision making supporting, and the advantage of embedded data analysis system.

Keywords: Academic Affairs Management System; Statistical Tools; Big Data Analytics; Decision Making Supporting

收稿日期：2014年2月10日

责任编辑：马小强

.....
(上接第129页)

应用。

丁树峰：馆长，高级讲师，研究方向为区域信息技术

张晶：工程师，研究方向为远程教育与网络技术

应用。

Design and Application of Competition System Based on e-Learning Platform

Zhang Baojun¹, Zhang Jing², Ding Shufeng¹

(1.Tianjin Educational Technology Center, Tianjin 300074; 2.Tianjin Open University, Tianjin 300384)

Abstract: With the rise of interactive network application based on Web 2.0, such as blog, micro-blog, etc, the research in e-Learning platform focused more on the learners' interaction methods. The purpose is to promote learners' participation. Learning activities, which is in the form of competition, has the high involvement feature. However, as the information tool for organization and management of competition, network competition system is widely used in recent years, but most competition systems still be developed and deployed independently and difficult to exchange data with e-Learning platform. This paper takes the system which is applied in the Tianjin Municipal Education Commission Erudite Park calligraphy competition for example, describes a kind of competition system based on the e-Learning platform. The system combined the e-Learning platform and competition system. On one hand, it provides one-stop user experience. On the other hand, interconnected data provide the possibility of deep binding and data mining for two systems. The construction of the system, not only solves the shortage of the calligraphy teacher, but also promotes the transformation of the study method.

Keywords: Network Competition; e-Learning Platform; .NET; Distance Education

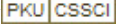
收稿日期：2014年3月26日

责任编辑：马小强

基于教务管理系统的数据分析工具的研究与设计

作者: [邓文俊](#), [曹阳波](#), [刘英群](#), [程建钢](#), [Deng Wenjun](#), [Cao Yangbo](#), [Liu Yingqun](#), [Cheng Jiangang](#)

作者单位: [清华大学教育研究院教育技术研究所, 北京, 100084](#)

刊名: [中国电化教育](#) 

英文刊名: [China Educational Technology](#)

年, 卷(期): 2014(5)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_zgdhgy201405023.aspx