

# 高校学生学业表现研究： 历史回顾、现实问题与未来路向

丁国勇<sup>1 2</sup> 程晋宽<sup>2</sup>

(1. 南京审计大学 江苏 南京 211815; 2. 南京师范大学 江苏 南京 210097)

**摘要:**在高等教育大众化以及教育大数据的双重背景下,高校学生学业表现的研究越来越突出以教育数据挖掘和大数据分析的方法,在数据事实与数据关系的基础上,建构促进学生学业发展、学业管理和学业支持的体系。通过对高校学生学业表现研究的历史回顾和现实问题的分析,高校学生学业表现的未來研究呈现出明显的转向:研究模式转向整合化,研究数据转向主客观数据相结合,研究方法转向统计学与教育数据挖掘的综合应用。高校学生学业表现研究需要强化学校环境、社会及人口统计特征、个人特征和学生投入等四个维度的研究框架和问题识别,采取数据理解、数据准备、探索性数据分析、模型建立与评价、模型应用等五个步骤的研究过程。

**关键词:**高校学生;学业表现;教育大数据;教育数据挖掘

**中图分类号:** G640

**文献标志码:** A

**文章编号:** 1003-2614(2019)10-0052-05

学业表现(Academic Performance)作为一个可测量的变量,可以定义为“在学习的课业中表现出来的可测量的行为结果”<sup>[1]</sup>。学业表现既具有普遍性,也具有相对性,学业表现的具体量化数值只在某个一定范围内,比如在同一所高校中,对于学习者之间进行比较时才有意义。近年来,有关高校学生发展的研究较多集中于学习投入和参与度,相关理论有派斯(Pace)的“努力质量理论”、阿斯廷(Astin)和库(kuh)的“学生投入理论”、齐克林(Chickering)的“七项最佳原则”等。以相关理论为基础,进行大规模的实证调查,国外有全美大学生投入调查(NSSE)、研究型大学本科生就读经验调查(SERU)等,国内有清华大学的“中国大学生学习与发展追踪调查(CCSS)”、厦门大学的“国家大学生学情调查(NCSS)”等。这些研究大多侧重于使用问卷调查方式获取学生学习投入、学习动机和学习行为等情况,关注于高校学生总体以及对学习投入与参与度进行校际比较。对于高校学生学业表现,因其不具有校际直接可比性,往往都归属于院校研究或校本研究的范畴,其影响力不如具有校际可比性的学生投入与参与度的研究。在“以本为本”、促进“四个回归”的前提下,高校需要突出关注学生个体的发展,探究学生学业表现及其影响因素,以制定切实有效的学生学业表现的管理制度和学业支持体系。因此,对于高校学生学业表现的研究更加具有现实和实践意义。

## 一、高校学生学业表现研究的历史回顾

### 1. 国外高校学生学业表现研究回顾

国外关于高校学生学业表现的研究起步开始于20世纪五六十年代,一般将影响因素分为认知因素和非认知因素。认知因素是与学业能力直接相关的,包括ACT分数、SAT分数、高中的GPA、高中的班级排名等,如潘蒂(Pantages)等人通过25年的研究发现高中成绩显著影响大学学业表现,杜利奥(Durio)、莱文(Levin)等人的研究也证明高中GPA(Grade Point Average,平均学分绩点)和班级排名可以预测大学中的学业表现<sup>[2][3][4]</sup>,但仅依靠认知因素对于研究高校学生学业表现是不够充分的,因此,学者们开始研究与学业能力不直接相关的非认知因素。非认知因素包括两大类:一类是入学前的个人特征和态度,主要包括性别、族裔和个人期望;另一类是入学后的相关因素,包括学习习惯、专业、与教师和同学的互动等。从20世纪70年代末起,国外学者对年龄、性别、种裔、家庭背景、助学金、学习支持等非认知因素对学业表现的影响做了大量的研究<sup>[5]</sup>。

近年来,国外研究主要集中在学生个人特征、心理特质、学习投入等方面。比如,帕帕耶奥尔尤(Papageorgiou)连续5年分析3075名会计专业学生第一年学业表现,发现年龄、性别、种族以及入学前的知识储备对学业表现都有影响<sup>[6]</sup>。卡拉尼(Kalani)分析家庭结构对大学生学业表现的影响,对88名夏威夷文理学院的大学生进行调查,发现来自离异家庭学生的GPA与来自完整家庭学生的GPA没有显著差异,但是来自离异家庭的学生更有可能出现课程不及格<sup>[7]</sup>。拉斯基

收稿日期:2019-07-04

基金项目:全国教育科学“十三五”规划教育部重点课题“教育大数据背景下高校学生学业表现建模研究”(编号:DIA180378)阶段性研究成果。

作者简介:丁国勇,南京审计大学教务处副研究员,南京师范大学教育科学学院博士研究生,研究方向:高等教育管理;程晋宽,南京师范大学教育科学学院教授,教育学博士,研究方向:比较高等教育。

(Laskey) 认为自我效能感、元认知和人格特质对大学高风险学生的学业表现有影响<sup>[8]</sup>。张文妮(Winnie Chang)认为,坚毅性(Grit)人格对大学生学业表现的提升有作用,性别、SAT分数、种族和坚毅性评价分能很好地预测学生第一年的GPA<sup>[9]</sup>。

随着学生投入理论的产生与发展,学生的投入、学校的政策环境对大学生学业表现的影响逐渐得到重视,这也得到一些研究结果的支撑。比如,莫妮卡(Monica)通过分析某高校的全美大学生学习投入度调查(NSSE)的数据发现,参加课程辅助活动的时间以及与教师、教育管理者的积极互动能够有效地影响学业表现<sup>[10]</sup>。科马拉(Komaraju)从美国中西部某中等规模公立大学抽取242名学生进行调查,通过师生互动量表、学业动机量表和学业自我概念量表等采集数据,结果显示学生的学业自我概念、内部动机、外部动机以及GPA都与师生互动的某些方面有显著的相关性<sup>[11]</sup>。

## 2. 国内高校学生学业表现研究回顾

与国外学者研究相比,国内学者对学业表现的影响因素的研究则更多地集中在不同群体比较方面,包括地域差异、城乡差异、贫困专项、自主招生、文理分科等。卢晓东等人对某部属高校城镇和农村户籍群体的大学生学业表现进行实证研究,结果发现,城镇学生群体学业表现大幅领先农村学生群体,这一差异将持续到大学毕业,基础教育的城乡差异在大学中延续<sup>[12]</sup>。王小虎等人对农村和贫困地区专项计划学生群体学业表现的研究发现,专项计划学生的高考成绩及大学GPA显著低于非专项计划学生,大学GPA差异主要来源于高考成绩差异,就读人文类学科相比社会类与理工类学科能够减少高考成绩差距对大学学业表现带来的持续负面影响<sup>[13]</sup>。

除了以上研究以外,国内学者还分析了家庭背景、家庭资本、性别差异等因素对学业表现的影响。高耀等人基于江苏省20所高校大学生的调研数据研究发现家庭资本会显著影响大学生学业表现<sup>[14]</sup>。李忠路对首都大学生成长追踪调查数据进行实证分析,发现读研学生的家庭背景和学业表现

都要显著地优于就业学生,在国内研究生教育机会的获得中,家庭背景通过影响本科就读高校的类型和学业表现来发挥作用,而国外研究生教育机会的获得则更主要受家庭经济地位的影响<sup>[15]</sup>。梁耀明和何勤英分析了性别差异对学业表现的影响,发现男生存在“后发优势”,学业表现在大二出现逆转并在大四超过女生,农村男生比城市男生的“后发优势”更加显著<sup>[16]</sup>。

国内学者对学习投入、心理学等学业表现影响因素的研究并不很丰富。邢全超基于某高校的“中国大学生学习与发展追踪研究(CCSS)问卷数据”进行分析,发现学习性投入和学业表现存在显著的相关性<sup>[17]</sup>。吴心灵等人的研究发现学生学业成绩与其人格特质中的自律性和稳定性成正比<sup>[18]</sup>。

通过对国内外研究的回顾,我们可以发现,对于高校学生学业表现的研究具有以下三个特征:一是研究内容聚焦化,主要是对学业表现进行差异分析和相关性分析;二是研究视角多样化,研究者从社会学、心理学、教育学等多个视角对学业表现的影响因素进行研究;三是影响机制复杂化,学业表现的影响因素数量众多,并且作用机制存在不确定性,同一影响因素在不同个体中的作用大小可能不同,作用方向甚至可能相反。正是这样一些特征表明高校学生学业表现及其影响因素具有较高的研究价值。

## 二、高校学生学业表现的现实问题

### 1. 高等教育扩张中学生群体的多样化与学业严重分化问题

截至2017年年底,我国有各类高校2631所,在校人数3779万,是1949年(11.7万)的323倍,是1978年(86.7万)的44倍,在校规模世界第一。2017年高等教育毛入学率已达到45.7%,已处于由大众化向普及化迈进的关键时刻。表1中列出2008-2017年全国本科招生数、在校生数、毕(结)业数等数据,可以看到,这10年间各类数据均呈现上升趋势,本科招生由297万人增加至410万人,本科毕业生由225万人增长至384万人。

表1 2008-2017年全国本科招生数、在校生数、毕(结)业数、授予学位数

年份	招生数	在校生数	预计毕业	毕(结)业数	授予学位	未毕业	未获学位
2008	2 970 601	11 042 207	2 324 882	2 256 783	2 082 558	68 099	242 324
2009	3 261 081	11 798 511	2 528 494	2 455 359	2 298 200	73 135	230 294
2010	3 512 563	12 656 132	2 662 960	2 590 535	2 435 867	72 425	227 093
2011	3 566 411	13 496 577	2 868 211	2 796 229	2 707 934	71 982	160 277
2012	3 740 574	14 270 888	3 113 087	3 038 473	2 966 148	74 614	146 939
2013	3 814 331	14 944 353	3 231 125	3 199 716	3 130 415	31 409	100 710
2014	3 834 152	15 410 653	3 497 450	3 413 787	3 338 323	83 663	159 127
2015	3 894 184	15 766 848	3 701 175	3 585 940	3 503 230	115 235	197 945
2016	4 054 007	16 129 535	3 879 845	3 743 680	3 659 686	136 165	220 159
2017	4 107 534	16 486 320	3 991 586	3 841 839	3 771 039	149 747	220 547

随着高校招生规模的持续增加和各类专项计划的实施,学生来源更加多样化、复杂化,学生接受的中等教育水平差异、学生个人学习能力的差异等也在不断加大,这也是导致学生学业表现严重分化的原因之一。由表1,我们也可以发

现,2008年至2012年不能正常取得毕业(结业)证书的学生人数,在基本维持在7万人左右,不能获得学位的由24.2万人下降至14.6万人,但2013年至2017年不能正常取得毕业(结业)由3.1万人增加到14.9万人,不能获得学位的由10

万人增加至 22 万人。按照《普通高校学生管理规定》,不能获得毕业(结业)证书意味着中途退学。虽然毕业率和获得学位率都在 95% 和 90% 以上,与西方大学的毕业率和获得学位率相比较,但由于我国高等教育总体量较大,因此不能毕业和不能获得学位的绝对数较大。

教育的主体是学生,提高高等教育的质量,学业发展是关键。学生学业发展的重要结果性衡量指标就是学生的学业表现,学业表现不良是学生不能毕业、不能获得学位的主要原因。

## 2. 高校学生学业投入不足与学业困惑问题

根据 2017 年发布的《中国大学生成长白皮书》,当代大学生普遍的问题是迷茫,95.7% 的大学生存在迷茫和困惑,其中,大一和大四学生更为明显,地区越偏远、入学成绩越低的大学生的迷茫和困惑程度越严重。大学生最困惑的十个主题是学业、职业、组织、日常、实践、自我管理、情感、精神与价值、烦恼、未来。其中,“学业”是大学生十大困惑之首,具体内容包括“专业课”“考研”“成绩”“挂科”“考试”“四六级”等。在大学生中普遍弥漫着“迷茫”的气氛,对大学生学习产生较大的影响,导致学生对学业的投入水平不高、投入方式单一、投入质量偏低。

学生对学业投入不足的表现是随意缺课、出勤率低、“不想学习”。根据笔者所在单位进行的一项调查,1% 的学生经常逃课,60% 的学生偶尔逃课,在逃课理由中,“参加学生工作或课外活动”排在第一位,占 54.9%,表明第二课堂活动确实在一定程度上冲击了第一课堂教学活动;其次是“对课程内容不感兴趣”,占 35.9%;“不满意教师的教学水平”,占 16.2%;“自身学习动力不足,不想学习”,占 15.1%;有 0.6% 的学生因为“沉迷游戏”逃课,1.2% 的学生“因为恋爱”而逃课;另有 10% 的学生选择“其他”原因。

学生学业投入的问题表现还包括普遍存在学业困惑,“不会学习”。进入大学以后,学生仍然延续高中阶段的学习方式,没有能适应大学教学和学习模式,学习主动性差,学习方法不当。同一项调查的结果显示,学生碰到的学习困惑,排在第一位的是“学习方法不当,不适应大学教与学”,占 26.8%,其次是“学习动力不足”,占 24.7%,“学习目标模糊”,占 21.9%,这三项之和占 73.4%。其他学习困惑还有学习氛围不好(11.8%)、缺少教师指导(8.2%)等。

大学生的学业目标模糊、学习动力不足、学习方法不当等学业投入方面存在的问题,其结果将直接体现到学生的学业表现中,是造成学生学业表现差异增大和部分学生学业表现不良的主要原因,直接影响了高校人才培养质量。

## 3. 高校对大学生学业的教学和管理支持体系构建的问题

在高等教育规模扩张前提下,相当数量学业基础不实、学习目标模糊、学习动机不足的学生涌入校园,学业表现不良学生群体出现的概率和群体数量较之以往自然呈现出显

著上升的趋势,这给高校传统的精英教育模式和教育管理制度带来极大挑战<sup>[19]</sup>。在应对这些学业表现不良学生群体时,高校采取的一种应对方式是严格管理,开除、劝退学业表现不良学生。

严格管理是只是硬币的一面,另一面,还需要建立和完善学业辅导与学业支持制度,进一步落实大学生权利保障,将对学生学业的事后管理转变为事中、事前管理,及时发现学生学业中存在的问题和遇到的困难,并给予适时适度的帮助。事后管理方式相对比较简单,根据学生所修读的学分情况、课程通过情况,依据相应的规则给予处理,但事中、事前方式相对就较为困难,事中方式需要掌握学生个体的全面情况,根据学生学业进程,依据个体情况制订针对性的方案,及时干预;事前方式还要能够进行学生学业表现的预测,及早干预。

## 三、高校学生学业表现研究的未来路向

### 1. 研究模式趋向整合化

传统的研究往往是对学业表现的单个或多个影响因素进行研究。在教育大数据背景下,应该设计一个总体研究框架,把诸多复杂的影响因素有效组织,使得研究模式从离散化转向整合化,这样才能保证研究设计更加具有逻辑性,研究结果更加具有可解释性。

高校学生学业表现的研究框架可以参考已有的相关研究框架,比如,阿斯顿 I-E-O 模型、乔治·库的大学生成功要素模型等。我们可以构建一个从学校环境、社会及人口统计特征、个人特征和学生投入等四个维度的高校学生学业表现研究总体框架(见图 1)。这一研究框架基于这样一个逻辑,学业表现是每一个学生个体的具体特征,因此应该先把每一位学生看成一个独立对象。社会及人口统计特征是类别特征,是可以被多个独立对象共享的特征,如性别、民族等。个人特征是保持每一个独立对象与其他独立对象相区别的静态特征,又包括入学前的学业能力、个体心理因素等。学生投入是每一个独立对象与环境互动的行为,表现为能动性特征。学校环境既是以上三种特征影响学业表现的工作背景,也可能通过与每一个独立对象三种特征的相互作用对学业表现产生影响。

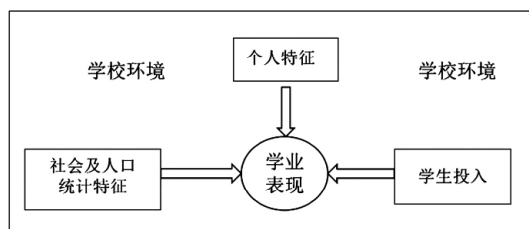


图 1 高校学生学业表现的研究框架

在这一个研究框架指导下,把学业表现的相关影响因素梳理、归类,可以使得研究设计、研究过程、研究结果更为清晰、明确(见表 2)。

表2 高校学生学业表现的影响因素

研究维度	影响因素
学校环境	学校类型、学校规模、学校文化、管理制度、招生政策等
社会及人口统计特征	性别、民族、生源地、城乡、社会经济地位、家庭背景等
个人特征	高考成绩、人格特质、成就动机、自我效能、身心状态、学生身份等
学生投入	参与第二课堂、参与社团、图书馆利用、师生互动、同辈关系、网络及手机使用等

2. 需要主客观相结合的数据支持

高校学生表现研究作为一项实证研究,数据是其重要资源,数据质量直接决定了研究质量。高校学生学业表现的研究数据可以分为客观数据与主观数据。客观数据是指可直接获得的反映现实世界的客观存在状态的数据,如受教育年限、学生的学业成绩、学生性别等。主观数据是指通过调查工具等方式获得的教师或学生的主体感知数据,如学生的就读经验、满意度等。尽管通过调查工作获得的信息带有个体主观评价色彩,但当众多个体主观信息汇聚在一起,并且结果具有统计意义上的显著性,它就能够反映现实世界的客观存在状态<sup>[20]</sup>。

高校学生学业表现研究的客观数据包括以下三种类型:一是学生自身的客观数据如高考成绩、高考科类、所学专业、所属学科等;二是与学生相关的客观数据,如学校类型、学校规模、班级规模、专业设置、课程设置、父母学历、家庭经济状态、同宿舍学生等;三是学生的客观行为数据,包括学生选课记录、学生考试成绩记录、图书借阅数据、第二课堂参与数据、社团参与数据等。客观数据的获取渠道主要通过教学管理信息系统以及其他相关的管理信息系统。高校学生学业表现研究的主观数据包括学生的人格特质、成就动机、学业自我效能、自我感知的师生互动、与同学的关系。主观数据的采集主要通过问卷调查的方式,问卷调查使用的量表,既可以是自行开发的,也可以是其他人开发的,但都应具有一定的信度和效度。在教育大数据背景下,可以构建整合型数据系统,将主观数据和客观数据结合起来,提高数据可信度和数据质量。

3. 注重多种分析方法的综合应用

充足的便于采集的主客观相结合的数据,为高校学生学业表现的研究提供了巨大的“金矿”。高校学生学业表现的研究方法应从单纯的传统统计方法转向注重统计学方法、教育数据挖掘等多种分析方法的综合应用,从注重因果关系转向为注重相关性,从注重精确性转向注重效率,将描述性统计、探索性分析和数据建模等多种研究方法结合运用。

教育数据挖掘遵循 CRISP - DM 标准,遵循着发现问题—分析问题—解决问题的流程。CRISP - DM ( Cross - Industry Standard Process for Data Mining, 跨行业数据挖掘标准流程)是1999年欧盟机构联合起草的数据挖掘建模流程标准,在各种 KDD(数据库中的知识发现)应用中占据领先地位。高校学生学业表现研究的教育数据挖掘流程分为问题识别与数据理解、数据准备、探索性数据分析、模型建立与评价、模型应用等步骤,见图2。

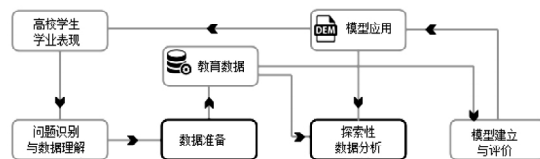


图2 高校学生学业表现研究的教育数据挖掘流程

问题识别与数据理解是“关键第一步”,明确高校学生学业表现研究的目的是建立聚类模型和分类预测模型,期望建模结果能够为高校改进教学管理、促进学生个性化发展提供数据依据和政策建议。在明确识别问题领域和建模目标后,需要进行的工作是数据理解,包括对存在数据的熟悉和初步的数据采集工作。数据准备阶段涵盖了从原始粗糙数据中构建最终数据集(作为探索性分析和建模的对象)的全部工作,包括收集原始数据,对数据进行装载,检验数据的质量,包括数据的完整性和正确性以及缺失值的填补等。这一阶段的任务主要包括采集数据、设计标准表、选择和转换数据、为适应建模工具进行数据清理等。数据准备工作有可能被实施多次,而且其实施顺序并不是预先规定好的。充分的数据准备,有助于避免“Garbage in, Garbage out(垃圾进,垃圾出)”的状况,数据准备工作的质量将直接影响后续探索性数据分析和数据挖掘建模的质量。

在数据准备工作完成后,进行探索性数据分析,通过统计学的、图形化的分析方法,深入了解数据集,校验数据间的相互关系,确定感兴趣的数据子集,对于预测变量之间可能存在的关系以及预测变量与目标变量之间可能存在的关系建立初步的构想。在模型建立与评价阶段,通过各种各样的建模方法对数据准备阶段建立的数据集进行分析。一些建模方法对数据的形式有具体的要求,有时需要回到数据准备阶段进行某些数据采集和数据清理。对于相同的问题也可以有多种建模方法选择使用,多种建模方法之间需要进行比较以确定最优的方法。模型建立阶段的结果是一组组的数学模型或规则模型。这些模型需要通过一定的评价方法来确定其效率与质量,只有达到预定质量的模型,才可以进行模型应用。

参考文献:

[1] Lavin, D. The Prediction of Academic Performance: A Theoretical Analysis and Review of Research [M]. New York: Russell Sage Foundation, 1965.  
 [2] TJ Pantages, CF Creedon. Studies of College Attrition: 1950 - 1975 [J]. Review of Educational Research, 1978(1): 49 - 101.

- [3] HF Durio. Mathematics Achievement Level Testing as a Predictor of Academic Performance and Retention in Engineering Students [J]. *Achievement Tests*, 1979(17).
- [4] J Levin, J Wyckoff. Effective Advising: Identifying Students Most Likely to Persist and Succeed in Engineering [J]. *Engineering Education*, 1988(11): 178 - 182.
- [5] Pati J. Fleck. Expectations, Academic Performance and Retention of Undergraduate Engineering Students [D]. Kent: Kent State University, 2000.
- [6] E. Papageorgiou. Accounting Students' Profile versus Academic Performance: A Five-year Analysis [J]. *South African Journal of Higher Education* 2017(4): 209 - 229.
- [7] B. Kalani Spain. Family Diversity: A Quantitative Study of the Impact of Divorce and Family Structure on the Academic Performance of College Students [D]. Capella University, 2008.
- [8] Marcia L. Laskey. Assessing the Influence of Self-efficacy, Metacognition, and Personality Traits on At-risk College Students' Academic Performance and Persistence [D]. Cardinal Stritch University, 2004.
- [9] Winnie Chang. Grit and Academic Performance: is Being Grittier Better? [D]. Coral Gables: University of Miami, 2014.
- [10] Monica Clayborne Scott. The Relationship Between Student Engagement Variables and Academic Performance among Students Enrolled at Historically Black Colleges and Universities in the United States [D]. Union University, 2014.
- [11] Komarraju M, Musulkin S, Bhattacharya G. Role of student-faculty interactions in developing college students' academic self-concept, motivation, and achievement [J]. *Journal of College Student Development*, 2010(3): 332 - 342.
- [12] 卢晓东, 等. 基础教育中的城乡差异是否在大学延续——高校城乡学生学业表现差异的实证研究 [J]. *高校教育管理* 2016(1): 56 - 60.
- [13] 王小虎, 潘昆峰, 吴秋翔. 高水平大学农村和贫困地区专项计划学生的学业表现研究——以 A 大学为例 [J]. *国家教育行政学院学报* 2017(5): 66 - 75.
- [14] 高耀, 刘志民, 方鹏. 家庭资本对大学生在校学业表现影响研究——基于江苏省 20 所高校的调研数据 [J]. *高教探索* 2011(1): 137 - 143.
- [15] 李忠路. 家庭背景、学业表现与研究生教育机会获得 [J]. *社会* 2016(3): 86 - 109.
- [16] 梁耀明, 何勤英. 大学生学业表现性别差异研究 [J]. *中国成人教育* 2017(6): 55 - 59.
- [17] 邢全超. 基于 CCSS 调查的本科生学习性投入与学业表现关系的实证研究 [D]. 重庆: 重庆大学硕士学位论文, 2015.
- [18] 吴心灵, 等. 医科大学生优秀学业表现的人格心理学分析及相关性研究 [J]. *社会心理科学* 2010(2): 50 - 54.
- [19] 孙建. 国外高校学业危机研究及启示 [J]. *首都师范大学学报: 社会科学版* 2015(6): 138 - 143.
- [20] 阎光才. 教育及社会科学研究中的数据——兼议当前的大数据热潮 [J]. *北京大学教育评论* 2013(4): 77 - 86, 187.

## College Students' Academic Performance: Historical Review, Realistic Problems and Future Directions

DING Guo-yong<sup>1 2</sup>, CHENG Jin-kuan<sup>2</sup>

(1. Nanjing Audit University, Nanjing 211815, China;

2. Nanjing Normal University, Nanjing 210097, China)

**Abstract:** Under the dual background of the popularization of higher education and the big data of education, the research on academic performance of college students is becoming more and more prominent by means of data mining and big data analysis. On the basis of the relationship between data facts and data, the system of promoting students' academic development, academic management and academic support is constructed. Through historical review and analysis of the practical problems of the research on academic performance of college students, the future research of academic performance of university students shows an obvious turn: the research mode turns to integration, the research data turns to the combination of subjective and objective data, the research methods turn to the comprehensive application of statistics and educational data mining. The research on academic performance of college students needs to strengthen the four-dimensional research framework and problem identification, including school environment, social and demographic characteristics, personal characteristics and student input, and adopt the five-step research process: data understanding, data preparation, exploratory data analysis, model establishment and evaluation, and model application.

**Key words:** college students; academic performance; educational big data; educational data mining