

高等教育研究通讯

〔2013〕第六期

本期主要内容

刊首语.....	2
中国高等教育质量保障：新形势 新理念 新制度（教育部高教评估中心副主任王战军）	3
推动优势转化，激发学术志趣，培养领跑人才（清华大学副校长袁嗣）	7
教育数据挖掘：现状与趋势（中央广播电视大学现代远程教育研究所魏顺平）	14
学习分析技术研究与应用现状述评（北京邮电大学李青等）	20
北农 8 号学生宿舍楼卫生情况影响因素的实证分析（北京农学院 11 级会计 1 班王婧）	27

主题词：教育评估；新理念；数据挖掘；学习分析；课程论文

北京农学院高教研究会秘书处、北京农学院高教研究所

2013 年 12 月 1 日印发

刊首语

10月27日、28日，“全球化时代大学生学习与发展研究”国际研讨会在清华大学召开，来自国内外高校的200多名专家学者聚集一堂，就大学生的学习投入、学习发展、学习收获、学习评价等进行演讲讨论，两天之内包括美国印地安那大学、哥伦比亚大学、匹兹堡大学、波斯顿学院、加州伯克利分校、马里兰大学、圣何塞州立大学、英国牛津大学、巴斯大学、澳大利亚墨尔本大学、南威尔士大学、日本九州产业大学、荷兰阿姆斯特丹大学、国内清华大学、北京大学、复旦大学、上海交通大学、西安交通大学、中国农业大学、南京邮电大学、中国海洋大学、中国教育科学研究院等共三十多个专家学者做了精彩发言并产生热烈讨论。现将会议特邀的教育部评估中心副主任王战军教授、清华大学副校长袁嗣教授两个报告录音整理成文，为开展新一轮教学质量评估、用教育发展新理念引领人才培养模式、深入开展教育教学改革提供新的内部学习材料。

我们已经进入大数据时代，云课堂、云数据、云计算已经不是一个陌生的词语。高等教育每天产生的海量数据如何处理，学校如何从“教育质量报告”和“教育评价数据”中挖掘有用信息，教师如何利用“学习分析”这一快速发展的信息化工具了解学生的学习成长过程，如何因人制宜、因势利导，帮助学生制定个性化学习发展计划并及时调整教学进度，“教育数据挖掘”和“学习分析技术研究”两篇文章介绍了这一教育信息化发展的来龙去脉、现状与技术应用发展趋势。

经济管理学院越来越重视对二、三年级本科学生课程论文的写作指导。近期发现一批学生论文写作水平已有很大提高。从规范化模仿、深层次分析、到独立创新写作，取得很大进步，这样的进步得益于任课老师的辛勤指导和严格要求，借一些师生的说法是经历了一场科学写作训练的艰苦“折磨”。本刊节选了会计专业三年级学生王婧同学“北京农学院8号学生宿舍楼卫生情况影响因素的实证分析”一文，希望能够引发思考。当网络快餐式学习成为当今社会主流的学习阅读与传播模式时，高校深层次科研训练如何引领，如何引导学生利用自己熟悉的生活题材，按照科研论文的写作规范，遵循数据科学处理和分析要求，撰写优秀课程论文，是很值得研讨并发扬光大的。看到越来越多这样优秀课程论文，不仅是任课老师的骄傲，是我们学生成才的希望，同时也是新一轮教育教学改革可以期盼的硕果之一。

中国高等教育质量保障：新形势 新理念 新制度

教育部高教评估中心副主任王战军教授

（在“全球化时代大学生学习与发展研究”国际研讨会上的演讲报告）

（根据现场录音整理，未经本人审阅，仅供内部学习参考）

女士们先生们：

大家上午好！非常感谢史静寰教授对我的介绍，我今天上午报告的题目是“中国高等教育质量保障新形势、新理念，新制度”。

昨天，各位专家学者对于高等教育的尤其是对本科生的学习和发展的评价做了非常好的报告，包括袁嗣校长昨天上午做了一个非常精彩的报告，那么我的报告两个问题。第一个问题是从外部评估的角度，来看如何保障和提高高等教育的质量，第二个问题是从宏观的角度，从院校外部怎么样对高等教育进行评估，来看如何提高和保障高等教育的质量。

先看看新的形势。现在中国高等教育新的形势，我想可能用三个“最”形容最合适，第一个就是中国的经济发展最快，第二个就是中国高等教育的学生规模最大，第三个就是中国提高高等教育质量最紧迫。

首先说中国的经济发展最快。由于经济发展最快，所以城乡居民发展性消费的支出已经比生存性消费的支出大了，这有详细的数据。中国经济发展最快，我有一组数据，从这组数据可以看到国际货币基金组织去年年底 GDP 的世界排名，我是把前五位和中国附近的一些国家的排名演示了一下。大家可以看到，中国经济的总量现在在世界排第二位，但只看排位还不行，我们要看一看占世界 GDP 的比重。中国占世界 GDP 的比重已达到 11.5%，在整个世界上经济上占的比重还是比较大的，当然最大的还是美国，美国占的比重是 21.9%，中国大致上和美国还差一倍左右。所以大家可以理解为什么中国到国外留学的学生在世界上比重是比较大的，经济增长越快，老百姓对优质高等教育资源的需求越来越强烈，这是第一个最。

第二个“最”是刚才所讲的，目前中国高等教育的总量已经达到了 3300 多万，详细的数据应该是 3326 万，这是中国高等教育总量，高等教育规模应该在世界上是最大的。那么，毛重学历呢，近几年的增长是非常快的，到去年年底中国毛重学历已经达到了 30%。

第三个“最”，中国的经济增长最快，学生的规模也最大，这样的话就带来一个问题，中国保障和提高高等教育质量最紧迫，为什么？昨天袁校长也讲了前一阶段高等教育有一个快速的发展，发展完了以后，提高质量就是最紧迫的问题，还有今天有很多国家的学者，国际上的学者对我国高等教学质量也会提出一些质疑，为什么？刚才通过演示大家也可以看

到，虽然我国经济总量最大，但是人均经济总量还是相差很多，总体上讲，我们还是一个发展中国家，那么发展中国家又承受了世界最大的高等教育，一方面总体上因为人均还是比较低，高等教育又是最大，怎样能保证质量，而且别的国际会议也讲中国还有重点建设，这样本来资金就不是很富裕，还要集中在某学校中的重点建设，这样的话高等教育的质量怎么保障。所以现在提高和保障高等教育质量确实是我们中国高等教育当前最紧迫的问题。

下面我们看，提高高等教育质量有哪些新的理念。

第一个新的理念，我们根据新的形势要以高等学校为主体，这是第一个理念。高等学校为主体就是要强调高等学校对于质量建设的责任，昨天袁校长也讲清华大学教学质量到底怎样，谁负责？清华大学负责。所以高等教育主体是高等学校，这是一个要强调的，同时我们也强化高等学校的自我评估，自我改进，就是这个质量不光要评估同时要改变，包括我们会议的主题，昨天有许多专家也提出这个问题，大学生学习发展研究，研究完了以后怎么去改进，怎样利用现有的研究成果去提高学生发展的质量，去提高教学的质量，强调以学校为主体，是当前一个新的理念，以前作为中国政府来讲，都是以政府作为主体做一些提高质量的工作，那么现在强调学校为主体，这是一个新的变化。以学校为主体，政府要求各个高校要发布各个高校的教学质量的年度报告，就是说每个学校每个学年要发布学校的质量报告，这个质量报告应该在 2010 年，就是中国 39 所“985”院校发布了年度的质量报告，这个报告因为是第一次发布，所以社会上有很多评价说，大部分学校的报告像招生宣传，还不像个质量报告，当然我们也采取了一系列的措施，提出一系列的要求，如我们提出来的 25 项数据，希望各个学校都要公布。这 25 项数据，其中一项就是学生的满意度，类似这样的指标学校需要公布。去年有 112 所“211”的学校已经发布了研究报告，今年实际上是要求所有的高校发布自己的研究报告，以后中国的高等院校每年都要发布自己的年度质量报告。根据这个报告，实际上也要求各级政府、各个区域甚至各个行业都要发布多种类型的质量报告，根据去年 112 所“211”学校发布的报告，我们评估中心和厦门大学教育学院合作形成了一本全国的“211 工程”高校的本科质量报告，这个报告目前已经寄到了“211”学校的校办，每所学校寄了 5 本，还有一些高校，比如说我们现在进行新建本科院校的合格评估，已经做了 2011、2012 年两个年度的合格评估报告，也发到了新建的本科院校。为了促进质量的提高，我们没有公开的出版，只是发给学校参考。这些报告里都会把全国整体的情况都会说到，包括一些统计的分析，也包括一些评估的报告。目前我们本科教学还存在哪些问题，都会告诉各个学校。同时，正在编写各种类型的质量报告，我们还编写了一本中国研究生教育的质量研究报告，2012 的这本报告已经出来了，2013 年这一版已经完成，正在联系科技出版社，

目前已经进入到出版过程，预计十月份可以出来。针对研究生的报告，我们还做了一个研究生的满意度调查，后面我会演示这个数据，所以强调学校这个主体地位就是学校要对自己的质量负责，这是第一点。

第二点我们强调以学生为根本。以学生为根本这个标题昨天袁嗣校长很好的诠释了以学生为根本，那么现在，从政府这个角度也在强调高校一切的教育教学活动都是为了学生的发展。第二点就是强化对学生学习成果的评价，特别重视学生对教学质量的评价，也是我们这次会议的主题。现在从整个中国政府来讲，也非常重视学生学习的成果评价，同时也非常重视学生对教学质量的评价，包括要求学校发布每个年度的质量报告，25项数据要公布，其中一项是学生对教学质量的满意度，中山大学已经做了两年了，每年会做一个学生满意度的调查。由于重视学生的满意度调查，所以也做了一个研究生教育的满意度调查。通过满意度调查我们也发现了一些问题，这个满意度调查呢，我们做了两个年度，2012年一共发了全国院校35所，2013年增加了两所，今年发放了14000份问卷，回收了10000多一点，回收率75.8%，我们的回收率相对来讲算比较高，去年研究生的满意度是63.1，今年达到67.4。最近几年，提高教学质量最紧迫，不管是从政府、还是从各个高校、还是从教学的教师、还是我们在校的学生，这几年对于这个问题的认识是非常清楚的，就是提高我们教育教学质量是当前高等教育发展的核心任务，我们上上下下，包括从政府也包括各个学校都采取了一定的措施，这些措施正在显现它的成效，显现的成效之一是对于满意度调查来讲，最近两个年度来看是有所上升的，总体上升，教学满意度上升的比较快，科研训练也有所上升，导师满意度也有所上升，还有一个管理与服务也都有所上升，我们在新的理念下应该说初见成效，另外还有一个非常相似的现象，通过我们学习效果的调查，师生互动是我们的一个短板、是中国高等教育的一个短板，也是研究生教学的短板，师生互动的满意度相对较低。近几年我在不断加强与学生互动，这是满意度调查。

第三个新的理念是第一以常态监测为手段。我们提出来的质量不是说只是一次调查、一次评估，现在非常强调进行常态的监测。常态监测首先强调要从周期性的评估转为常态的监测评估，常态监测评估是现在高等院校评估一种新的类型，我们也在加强这方面的研究。监测评估不同于以往的评估，以后这方面的成果还要逐渐地跟大家一起交流。第二个就是强化教育教学质量常态监测结果是及时的改进，常态监测发现问题以后要及时的进行改进，实际上建设了一个数据平台，目前这个数据平台有300多所高校的数据，11月份将对广东省、黑龙江全省进行采集我们本科教学的数据，根据学校不同，数据的量也不一样，像中山大学有十万左右的数据量，因为它的专业比较多。数据的采集点是800多个，这样会建设一个

数据的平台，会形成各个学校的数据分析报告，这个数据平台也可以形成全国的数据分析报告，还可以形成各省、自治区的质量分析的报告，我们现在强调常态监测，数据平台是按年度采集数据，将来通过技术的改进，希望将其周期缩短，以这种常态监测的方式来保障和提高教育教学质量，这是第三个理念。

中国高等教育质量保障目前有一系列新的制度，这个新的制度是在过去几十年发展的基础上，同时吸收了国际上各个发达国家的一些新的理念的基础上，从实践中总结出来的新的评估制度。新的评估制度分三个层次。第一个层次是有关研究生的质量监管制度，这个制度以这么几项为代表，一个是质量与约谈。如果通过各种类型的评估和常态监测发现问题的这些学校要进行质量约谈，由教育主管部门找学校进行质量约谈，去年已经对 8 所学校进行了质量约谈。第二个是进行学位论文的抽查。全国都要进行学位论文的抽查，而且各个学校都要进行博士论文匿名评审等等。第三个就是各个大学的学位授权点，要进行定时评估，6 年为一个周期的评估，这是关于研究生。

关于本科教学建立了“五位一体”的教学评估制度。“五位一体”包括院校的评估，包括学校自我评估，体现出以学校为主体，它不仅仅是为其他各项评估服务，同时有专业的认证和评估，还有国际评估。刚才讲的常态监测是这个制度下其中一部分内容。目前院校评估主要做的是两种类型，一种类型是合格评估，一种是审核评估。合格评估主要是指新建的本科院校进行合格评估，这个有指标体系，有详细的要求，但是审核评估的新理念，是没有评估指标体系，只有评估的范围，这个范围根据各个学校的标准不一样，各个学校用自己的标准评估自己，如教学目标，这是一个新的理念。

还有一个是中国高等职业院校数量比较多，有 1300 多所高职院校的评估是委托各自治区去做，目前已经评了 700 多所，新的制度主要是三个方面，一是研究生，二是本科生，三是高职。

这是我所跟大家报告和交流的，谢谢大家！

推动优势转化，激发学术志趣，培养领跑人才

清华大学副校长 袁驷教授

“全球化时代大学生学习与发展研究”国际研讨会上的演讲报告

（根据现场录音整理，未经本人审阅，仅供内部学习参考）

各位国内外嘉宾、各位同行：

我代表清华大学欢迎各位嘉宾的到来，这是一次很好的学习交流机会。我应邀来对清华大学近几年在人才培养方面做得一些工作与各位嘉宾进行一下交流和汇报。我今天演讲题目是“推动优势转化，激发学术志趣，培养领跑人才”。主要是清华大学近几年教育实践，还有其中的一些体会。想分四个方面跟大家做个汇报，第一是特色与理念，第二是反思与研讨，第三是实践与进展，第四是体会与经验。

第一部分特色与理念，是说清华大学的人才培养特色和我们的理念。清华大学已经有100多年的建校历史了，不同时期有不同的人才培养的特色，当前人才培养的特色可以概括为：厚基础，重实践，求创新。清华大学100年来始终秉承一个办学理念，我们把它归纳为一个根本，两个中心，多项职能。一个根本就是学校始终以人才培养为学校的根本任务，大学还有其他的多项职能，在这么多的职能当中一个根本是最重要的。如果把一所大学比作一棵参天大树的话，那么人才培养就是这棵大树的根和干，而两个中心和多项职能是这棵大树的茎和叶，根深才能叶茂。时代变了，千变万变以人才培养为根本任务始终没有改变，这几句话是我在学校经常和老师强调的，所谓人才培养是学校的根本任务，就是说学校要以人才培养为根、以人才培养为本、以人才培养为任，以人才培养为务，而且是第一要务，我这些话可能给翻译带来一些困难，都是中国语境下的表述。

学校的教学观，我们认为教是为了学，教学的本质是教学生学，教的质量要由学的质量来检验。要说你的教育教学取得了多大的成绩，质量有多高，不看你有没有多少名师，不看你有没有多少著名学科，也不取决于你有没有多少特色专业等等，应该用你学生学习的质量来检验，我们认为人才质量是教育教学质量的试金石和标尺。我们要以学生为主体、以学生发展为导向、以学生成才为目标，树立全周期的人才质量观。所谓全周期的人才质量观就是包括招生阶段的生源质量，在学期间的培养质量和就业之后的发展质量，都应该纳入到人才培养体系中来。今天我们讨论会的题目是“全球化”，人才质量观还有一个新的国际上的维度。这三句话是我经常对我们的学生讲的三句话我说：You are made in china, Made by Tsinghua, Made for the whole world.

第二部分反思与研讨。在我们清华大学百年校庆之前，百年校庆是 2011 年，在 2010 年的时候校友会做了一个全面的调研，校内做了一些研讨和座谈，呈现出了一些现象，我们根据这些调研的结果发现：改革开放 30 年之后在人才培养上有两个不明显。

第一个不明显就是学校人才培养优势未得以充分发展。清华大学过去由一个以工科为主的，以教学为主的大学转型为综合性的研究型大学，各方面的办学优势都得以形成和发展，但是回头一看，人才培养的优势不明显。因为当时还是在 2010 年的时候，我们国家最高级的党、政的领导班子还没有换，所以上一届，大家都知道我们国家主席，总书记都是我们清华大学培养出来的，很多 VIP 校友，包括吴邦国委员长也是我们清华大学走出来的，那么在现在的阶段，不知清华人还能否发挥当时的那种优势，所以当时新闻媒体觉得不明显，当然这是对于我们清华自己的高标准来讲。那么是不是改革开放 30 年来学校不重视人才培养呢？不是。多样化的教育教学改革 30 年不断线，不断有各种举措，各种措施出台，不断地有研讨会，但是回头一看，人才培养的成效不是很明显。那么我们做一个类比，就像我们国家发展的模式，多多少少也是先建设先污染，回头一看在治理，以牺牲资源、能源、环境为代价建设和发展，那么大学办学模式前若干年也是先建设先发展，回头再治本，这个本就是我之前所讲的根本任务，是以牺牲人才培养的优势为代价，其他的优势都上去了，我们的院士最多，我们的科研经费，发表论文，学科建设等等，各项指标都上去了，但人才培养的指标，没有确切的指标。有这么两个不明显，这让我们反思总结出两个经验，把这两个经验说的直白一点，第一条是就教学抓教学不行，第二条只抓教不抓学不行。我们学校的名师、精品课、特色专业教学评估各个方面指标都很好看，我自己也是我国首批的教学名师，但我也知道我们的教学，我们的人才培养，有很多的弱点，所以这是后来我们针对第一条经验，我们学校提出以优势转化为我们人才培养的基本战略。

针对第二条经验我们提出要以学生受益为检验。随后 2009 年 7 月 8 日到 2010 年 7 月 8 日整整一年，集全校之力召开了第 23 次教育工作讨论会。学校有这样的一个传统，每四年集全校之力，上上下下开展为期一年的教育工作讨论会。因为第 23 次讨论会是清华百年校庆之前，最后一次教育工作讨论会，所以我们把这个主题选为清华新百年人才培养的使命与战略，那么这个讨论会的详细情况我不展开来讲，简单地说这次讨论会至少形成了两大理论成果，第一是清华新百年还要不要以人才培养为根本任务，一开始大家不是很一致，通过这次讨论会之后，学校大家再次确立清华新百年仍然以人才培养为根本任务，尽管还有科研、社会服务、文化传承与创新等等其他的职能，但是人才培养为根本任务百年不动摇，新百年也不动摇，而且记录明确提出以优势转化为人才培养的基本战略。2011 年 3 月学校党委常

委会正式颁布了一个文件，这个文件是清华大学“关于进一步加强优势转化、提高人才培养质量的原则意见”，一共有 18 条，这个文件的基本思想就是加强优势转化。我用这样的一个人图表来给大家介绍所谓优势转化的基本战略。清华大学有综合性学科的优势、有高水平科研的优势、有高素质师资的优势、有国际化教育的优势、有优质生源的优势、有杰出校友的优势、有美丽校园的优势、有优良传统的优势，还有等等的优势。所谓优势转化就是要积极、主动、优先地将所有的这些办学优势转化为人才培养质量的优势，这就是回应就教学抓教学不行，这样一个历史上的经验。

第三部分实践与进展。由于时间关系，我只能简单介绍若干典型案例。前面第二条经验说只抓教不抓学不行，那么如何抓学生的学？第一要推动优势转化，第二要激发学生的学术志趣，我们经过讨论，我们培养学生就像要发射学术卫星，我们经常说培养创新型人才，有创新精神这是我们的终极目标，但是它是要有前提的，学术要创新首先学生要有学术挑战能力，有学术挑战的精神，要有学术挑战的精神就要有自信，要有自信就要有一定的学术能力，有一定学术能力就要肯学肯钻研，那就要有学术的志趣，没有学术的志趣，后边的发射各个阶段就都很难实现，所以我们学术培养就像发射卫星，发射卫星我们做的工作就是一级火箭加速，二级火箭加速，三级火箭加速，直到把它打入到太空的轨道当中，你再加速也加不上去了，它自己有自行的运转，那个时候就是学术创新自然而然就达成了。所以我们要从根本上抓，要抓学术志趣。一开始学生进入大学对某些学科有兴趣，你要激发他的乐趣，有了乐趣慢慢就形成了志趣，形成了志趣之后，毕业之后就成为他们的学术志趣。所以学术志趣激发了之后，学术志趣又是一个优势，优势转化来激发更多学生的学术志趣，所以我们有两句话：让优秀的学生领跑，把一些优秀学生的学术志趣激发出来让他们领跑，让他们带动所有学生优秀，形成一个你追我赶，创先争优的局面。

讲几个案例，第一个案例是我们有一个清华学堂人才培养计划，讲这个案例主要是要介绍为优势转化为学术志趣，我们做得很努力。学堂计划形成了六个班：数学、物理、化学、生物，基础科学还有两个计算机科学和力学班，他们是工科里边基础性的学科。每个班每年少而精。像四个理科班每年每个班不超过 20 个人，两个工科班每年每个班不超过 30 个人，他们实施开始的时间有先有后，是从 2009 年开始实施的，我们的理念就是优势转化，要有好的学生、好的教师、好的教材，好的环境，我们要用拔尖人才培养未来的拔尖人才，每个班我们设立了一个首席教授席位，都是请的学术大师来做首席教授，也是他们自己主动，像数学班的首席教授邱成桐先生，物理班的首席教授朱邦芬先生，计算机科学班的姚期智先生。经过几年的实践，我们自己梳理出来的一些经验就是如何培养拔尖学生。第一要创建一个理

念，创建一个领跑者的理念，第二是营造一个环境，这是学堂计划。清华大学的学堂计划是名符其实的学堂计划，名称是清华学堂培养计划，实施上学校下了很大的决心，把清华最早的建筑物——清华学堂还给学生，作为实施这个计划的专门场所。第三是实施优势转化战略，第四要优先一个环节，优先素质教育，时间关系我不展开讲，只是讲一些要领，在实施的过程中我们遵循三个个性，一个是学生的个性，一个是学科的个性，一个是教授的个性，不做统一规定，逐班量身定做，逐生个性培养，我们还要做到三个字，确实现在已经做到了，这六个班的首席教授还有教学团队，都是他们自发的、自觉的、自主的来实施，特别是我们的目标希望学生自主的来追求学术上的卓越。实施了几年下来，像姚期智的姚班，国际专家对姚班的评估第一句话是有最优秀的本科生，我还不以为然，因为我们很有自信，我们有最优秀的本科生，但是第二句话是姚班有最优秀的本科教育，这点我还感到非常欣慰。所以姚期智教授领衔的姚班经过几年的努力，在去年被评为北京市教学成果奖的特等奖，而且在第一轮投票中，12个委员都是北京市的校长，还有教育部的教育专家投了一个满票，我们倒不是要争取满票，我们是对社会所有对清华大学学堂计划的认同和认可、鼓励和支持感到很欣慰。再举个例子我们学校六个班的数学班，数学班2008级首届毕业生已经毕业了，数学班一共是20个人，当然数学系除了学堂班20个人外还有普通的学生，那么多数的学生都选择了基础学科为研究方向，他们的学术志趣被激发出来了，形成了他们的志向。过去我们数学系毕业的学生出来不搞数学，多数去金融去证券，去那些赚钱快的地方。毕业的时候，一个叫14个班内15个班外是什么意思，学堂班数学班班内有14名学生，数学系班外普通的学生有15名同学，同时都被欧美的著名高校全奖 offer 给录走了。这里面值得一提的是前面这两项，有两名同学被哈佛大学数学系录取，两名同学被哈佛大学统计系录取。哈佛大学数学系去年在全球一共只招12名博士生，清华大学进去两位，这在过去是没有的，我这儿还加了个*号，这个*号是什么意思，这两位同学一个是学堂班数学班的，一个是数学系普通班的学生，所以前边这个14+15，这14位是领跑者，15位是追赶者，确实追赶到了比肩而行，所以这个领跑者的效应开始显现出来了，我们不是就培养这些拔尖学生，我们是要用他们发挥引领示范作用。再说这个例子，到哈佛大学的一个学堂班数学班的叫朱一昂同学，他去年秋天去哈佛大学攻读博士学位，去后在一周内通过了所有数学科目博士生资格考试，这说明也是一个检验，我们在大学期间培养数学的功底是非常扎实的，也是很先进的，所以第一个例子就是要激发一些优秀的学生，让他们做领跑，把他们的学术志趣激发出来，再利用他们的学术志趣的这个优势转化为更多的学生学术志趣的激发。

第二个案例是我们的学生学术竞赛，讲这个案例是想给大家讲我们的体会，对学生你不

激发他就很“平庸”，当然我们的平庸是加引号的，清华学生没有太平庸的，你一激发他，他就优秀、就卓越。在我国国内有一个高校挑战杯，这是一个科技活动，号称是国内高校奥林匹克竞赛，学生的奥林匹克竞赛，科技的奥林匹克竞赛。挑战杯是由清华大学发起，后来推广到全国，推广到全国之后这个挑战杯就很少再回到清华大学，最早一次清华捧回挑战杯是2003年，从2003年到2009年之间，没有把挑战杯再捧回来过，捧不捧回来我不说，我不怕今天揭揭我们的短板，揭揭我们的家丑，我们的名次落到了二三十名，2009年是我们百年校庆前最后一次全国挑战杯，如果我们学生的成绩达到二三十名，我们很难向我们的校友交代，那抓什么，就抓学术志趣，激发学生的学术志趣，我们的学生是不是这么水，是不是这么怂，是不是这么笨？不是，我们都知道，只要一抓，2009年从第十一届的三十多名一下窜到了第一名，捧回了挑战杯，光一次不具规律性，也许你比较 lucky，其他的因素。每两年一次挑战杯，2011年我们是第二名，取得了优胜杯，今年刚刚结束，我们又是第一名，把挑战杯又捧回来了，所以清华到现在一共捧回了五次挑战杯，是全国高校捧得挑战杯次数最多的学校，这就是抓了学术志趣，还是这些教师，我们的学生表现出的优势。

第三个案例是学生国际化培养，国际化培养不是为了国际化而国际化，是为优势转化激发志趣，激发自信。我们的学生在校内，我们跟他们讲这个他们不爱听，在校外，国外的教授给他们说几句话，他们记一辈子。我们派出规模，以2012年为例，我们每年本科生招生3360人，2012年本科生派出1400多人，达到了大于40%，也就是说进入清华大学的本科生，在四年中，有40%的机率能够有至少一次出国交流的机会，给大家看这些线，快速增长，我见很多国外的大学，学生国际化是必谈的项目之一，开阔资源优势转化，我们有国际合作的义务，要让这些优势首先积极主动优先的转化为我们人才培养优势。我们有一个清华优秀新生海外研修计划，我们一些优秀新生，有一些奥数赛金牌获得者，状元等等，过去连续三年都去英国牛津大学做研修，为什么呢？因为去了一次之后，牛津大学就不放我们了，就形成了一个固定的项目，每年你都上我这来，每次去他们的校长，哈密尔顿校长都出来接见我们的学生，会见我们的学生，而且还给我们的学生致辞。本来有一小段视频我就不放了，牛津大学校长对清华学生提希望，我把它翻译成胆识、创新、挑战。我觉得牛津大学校长看到了我们身上的一些弱点，有针对性的给我们提的，也是我们追求的，激发学术志趣，激发学术自信，学术的志向，挑战创新。

第四部分，因为史静寰教授建议我把自己的一些土理论，土理念给大家介绍一些。有几条土理念，说出来可能有争议，还不成熟，供大家参考。

第一条我的土理念是人才公平。现在我们大讲教育公平，但是我觉得要区分对人的公平

和对才的公平，要区分起跑公平和终点公平。人才是由人和才组合起来的。对人的公平要有教无类，机会均等，教育均衡，有平等的接受教育的机会。但是对才的公平，我们要因材施教，人尽其才，才尽其能。说实话我们刚开始实施学堂计划的时候，挑一些好学生来做这样的强化培养，校内校外也是有一些不解和质疑的，但是我们还是坚定的去做。

第二条土理论就是螺旋式上升，螺旋式上升并不是我们的，在哲学上本来就有螺旋式上升。清华 98 年就有了数理基础科学班，03 年我们又开设了化生基础科学班，05 年又有了人文科学实验班和社会科学实验班，06 年又做了软件科学实验班，07 年又有了经济与金融科学实验班，09 年力学系航空航天学院又跟我们申请要成立钱学森力学班。那么这样一个散的摊子，我们不如把它整合一下，所以我们出台了人才培养计划，挑选了这些班，把他汇聚一堂，整合后就是一个升级的版本，外界说又开这班那班，其实我们觉得不是，你看像微软 office 他把 Word, Excel 等整合成一个 office，它就是一个升级，苹果也是一样，三大件整合成 iword，刚出来时完全免费，这些都是改进。

第三个理论叫馒头理论，说俗话说单个馒头凉得快，要是放在一屉里面它就凉的慢，这就是为什么我们要把那些零散的班汇聚一堂，做一个学堂计划，让他们互相有一个良性的感染和竞争。

第四个理论是赶羊和牵羊的理论。过去我教学的时候，我就是在后面赶羊，保及格率，后来才发现保及格率之外还有另外一个提高优秀率，后来我发现提高优秀率比保及格率更合算。追求保及格率的时候，国家提出检测一下我们学生，我们学生有 20%都 90 分以上，后来我转变了观念，我是提高优秀率，我让前边三分之一的学生，只要他们能接受，我就牵羊，结果我的优秀率，还用国家检测同样的难度，优秀率达到了 25%，达到了四分之一。这五个百分点的优秀率，就是让优秀学生领跑领出来的，赶羊没有把优秀学生的优势转化出来，优秀学生在前面没发挥，在那没动。

第五个理念就是领跑理念。学堂计划有一个问题，进入学堂计划的学生和在计划外的学生，我们是把进入学堂的学生定位为领跑者，所谓领跑者这有几句话：不是光环，是使命；不是荣誉，是责任；不是娇宠，是磨练；不是圈养，是放飞；不是孤傲，是引领。

第六个理论是优势转化理论。其实办大学我们实施优势转化理论，其实我们上一门课，教育教学，教书育人本质就是优势转化你教书就是把教师的知识能力，素质转化给学生，“教”转化为“学”，转化为“知”，转化为“行”。育人就是把老师的品德平品格品行转化给学生，言传身教、为人师表、行胜于言都是优势转化。

第七个就是情结理论。学堂挖掘学生，最后都被欧美高校录走了，这个好不好，也备受

攻击和质疑，那么我的想法是要在优秀的学生，特别是拔尖学生的心底深深的种下一颗情结种子，这颗种子就是清华的情结，国家的情结。所以我们对他们给予特别的关心，特别的培养，特有的激励，让他们心里有情结，用中文就是说事情一半是事，一半是情。英文的表达是 Home is where the heart is. 就是说你的人在哪，你的家在哪，真正你的家是在你心里。You are made in china ,Made by Tsinghua、Made for the whole world 就体现了我们要我们学生要有中国的情结，清华的情结，即使他们是具有国际竞争力的学生，要有这样的情结，所以今天的题目就是推动优势转化，即发生学术志趣，然后要让优秀的学生让所有的学生优秀。

请批评指正，谢谢！

教育数据挖掘：现状与趋势

摘自《第 13 届全国计算机模拟与信息技术学术会议论文》

http://d.g.wanfangdata.com.cn/Conference_7559888.aspx

作者 魏顺平（中央广播电视大学 现代远程教育研究所）

在过去的 10 余年，随着教育信息化工作的大力推进特别是数字化校园建设和网络高等教育的大力推进，在教育领域已经部署了众多的软件系统，并且这些软件系统中存储着海量的教育数据。如何利用这些教育数据，使这些数据转变为信息、知识，并为教育决策、教学优化服务，这便是教育数据挖掘所关注的内容。

近年来，教育数据挖掘（Educational Data Mining）在国内外得到了迅速发展，下面通过文献研究的方式，对教育数据挖掘的兴起、定义、价值、过程、方法及未来发展趋势进行概要介绍，用以对这新领域有一全面认识。

1. 教育数据挖掘的兴起

自 2005 年以来，国际上许多计算机应用相关会议均设置了有关教育数据挖掘（Educational Data Mining）的研讨会，这些研讨会虽然在不同时间、不同会议上举行，但是他们的宗旨都是共通的，即在各种 e - Learning 学习系统（如网络教学平台）广泛应用并产生了海量学习记录的背景下，聚集数据挖掘、智能导师系统（ITS）、适应性教育超媒体等领域的研究者，共同探讨如何应用数据挖掘技术来分析学习系统或教学实验产生的数据，并如何利用这些数据来改进学习系统的适应性和个性化。

由于“教育数据挖掘”这一研究领域得到了越来越多研究者的关注，这些研究者便组成了一个共同体——国际教育数据挖掘工作组（网址：<http://www.educationaldatamining.org/>）。该工作组创办了在线学术期刊《教育数据挖掘杂志》，并从 2008 年开始独立召开“教育数据挖掘国际会议”，而不再依附于其他国际学术会议。截止到 2011 年 11 月，“教育数据挖掘国际会议”已经举办四届。

2. 教育数据挖掘的内涵及其价值

教育数据挖掘可以看作是数据挖掘的下位概念。简单来说，数据挖掘是指从大量数据中提取或“挖掘”知识 [1]。许多人把数据挖掘视为另一个常用的术语数据中的知识发现或 KDD 的同义词。目前，国际教育数据挖掘工作组在其官方网站上给出了“教育数据挖掘”的定义，指出“教育数据挖掘是一个新兴的学科，该学科致力于开发出一系列数据挖掘方法，该方法可用于挖掘来自教育系统的独特数据，并用于更好的理解学生及其所在的教育系统” [2]。这里的“来自教育系统的数据库”既有师生在使用互动学习环境（如网络教学平台）时

产生的数据，又有学校开展教育教学管理过程中产生的数据。这里将“教育数据挖掘”上升为一个学科似乎为之过早，因为这一领域并未形成完备、独立的知识体系。而国际教育数据挖掘工作组在 2008 年教育数据挖掘国际会议论文集中给出的教育数据挖掘则更为恰当，他们指出，“教育数据挖掘是一个将来自教育系统的原始数据转换为有用信息的过程，这些有用信息可为教师、学生、家长、教育研究人员以及教育软件开发人员所利用。

教育数据挖掘可看作嵌入已有教育系统的一个新的模块，并与教育系统中的各种要素产生良性互动，最终实现改进学习的目的。

对于学习者而言，教育数据挖掘的作用体现在：向学习者推荐有助于改进他们学习的学习活动、学习资源和学习任务，向学习者推荐好的学习经验，向学习者推荐访问捷径或简单链接。这些建议基于这些学习者完成的行为以及与之相似的学习者完成的行为分析得来，等等。

对于教育工作者而言，教育数据挖掘的作用体现在：向他们提供更多客观的教学反馈，为他们评估课程内容结构以及课程内容在促进学生学习过程中的有效性，根据学习者的需要对学习者进行分组，发现最为常见的错误，发现用于改进课程适应性和个性化水平的信息，重构网站以优化课件的个性化水平，根据学习者的学习状态来组织教学内容，重构教学计划，等等。

对于专家学者和管理者而言，教育数据挖掘的作用体现在：获取相关参数以改进网站效率并使之适应用户的行为，生成量表以用于优化教学资源和教育产品的组织，增强新的网络教育方法在提供教育课程方面的有效性。

3.教育数据挖掘过程

教育数据挖掘过程与数据挖掘的一般过程是一致的，也由数据准备、数据预处理、数据挖掘、模式解释等几大部分构成。

(1) 数据准备：了解数据挖掘应用领域的有关情况。包括熟悉相关的背景知识，搞清用户需求。

(2) 数据选取：数据选取的目的是确定目标数据，根据用户的需要从原始数据库中选取相关数据或样本。在此过程中，将利用一些数据库操作对数据库进行相关处理。

(3) 数据预处理：对步骤 2 中选出的数据进行再处理，检查数据的完整性及数据一致性，消除噪声，滤除与数据挖掘无关的冗余数据，根据时间序列和已知的变化情况，利用统计等方法填充丢失的数据。

(4) 数据变换：根据知识发现的任务对经过预处理的数据进行再处理，主要是通过投影

或利用数据库的其他操作减少数据量。

(5) 确定数据挖掘目标：根据用户的要求，确定数据挖掘要发现的知识类型。因为对数据挖掘的不同要求会在具体的知识发现过程中采用不同的知识发现算法。如分类、总结、关联规则、聚类等。

(6) 选择算法：根据确定的任务选择合适的知识发现算法，包括选取合适的模型和参数。

(7) 数据挖掘：这是整个数据挖掘过程中很重要的一个步骤。运用前面选择的算法，从数据库中提取用户感兴趣的知识，并以一定的方式表示出来（如产生式规则等）是数据挖掘的目的。

(8) 模式解释：对在数据挖掘步骤中发现的模式（知识）进行解释。经过用户或机器评估后，可能会发现这些模式中存在冗余或无关的模式，此时应该将其剔除。如果模式不能满足用户的要求，就需要返回到前面的某些处理步骤中反复提取。

(9) 知识评价：将发现的知识以用户能了解的方式呈现给用户。

在上述步骤中，数据选取和数据挖掘算法的应用占据非常重要的地位，前者是整个挖掘工作的基础，后者主要是利用某些特定的知识发现算法，在一定的运算效率范围内，从数据中发现出有关知识，决定了整个数据挖掘过程的效果与效率。下面就来重点介绍这两个环节。

3.1. 教育数据挖掘的数据来源

根据 Romero 和 Ventura（2007）的研究，教育数据挖掘从数据来源来看可分为传统学校数据和远程教育数据[3]。这是从一个教育机构所采用的教学手段来划分的。一个教育机构一般有教学和教务管理两项业务。从不同业务的数据来源来看，教育数据挖掘可分为教学数据和教务管理数据，这两种数据在一个教育系统内是相互流动的。因此，可从教育机构的技术手段和业务范围两个维度来对教育数据挖掘的数据来源进行分类。

第 I 类数据——课堂教学数据。随着各种电子教具在课堂中的应用，课堂教学中也会产生并自动保存丰富的数据。如课堂互动反馈系统的应用将可以详细记录师生开展教学活动的过程。IRS(Interactive Response System)是基于交互式学习理论设计的一种允许学生在课堂上通过手持设备以选项或数字来回答问题的学生互动反馈系统。

第 II 类数据——远程教学数据，与本研究领域“e-Learning 学习过程中的数据挖掘”的数据几乎是一致的。远程教学可基于不同形式的数字化学习环境开展。这些数字化学习环境包括网络课程、学习内容管理系统（LCMS）以及智能与自适应网络教育系统。远程教学数据则来自这三类数字化学习环境所产生的各种日志，有在服务器上保存的日志数据，也有在客户端上保存的日志数据。

第 III 类数据——数字化教务管理系统的数据库，指的是教育机构所使用的数字化教务管理系统录入、保存和管理的数据，这类数据结构良好，可批量采集，是数据挖掘的理想对象。

3.2. 教育数据挖掘算法

教育数据挖掘仍然采用数据挖掘已有的各种算法，不过可以按照挖掘任务的相似性对这些算法进行分类。

综合 Romero 和 Ventura (2007) 和 Baker and Yacef (2009) 对教育数据挖掘方法的分类，这些方法可分为五类：统计分析可视化；聚类（聚类分析、离群点分析）；预测（分类、回归分析、时序分析）；关系挖掘（关联规则挖掘、序列模式挖掘、相关挖掘）；文本挖掘 [3][4]。

对 1995 年至 2005 年间发表的 60 篇有关教育数据挖掘的论文进行内容分析，分析结果表明：43% 的文献 (26 篇) 采用了关系挖掘方法，28% 的文献 (17 篇) 采用了预测挖掘方法，5% 的文献 (9 篇) 采用聚类挖掘方法。再对 2008 年和 2009 年两届教育数据挖掘国际会议的论文进行内容分析，不同数据挖掘方法的文献比例则有所变化，其中，9% 的文献使用了关系挖掘方法，40% 的文献采用了预测挖掘方法，15% 的文献采用了聚类挖掘方法。可见，预测挖掘方法在近年来得到了更多的应用。

从国内研究来看，笔者通过 CNKI 中国期刊数据库找到 1999 年至 2009 年发表的教育数据挖掘相关的文献 150 篇，其中只有 49 篇文献运用了数据挖掘方法对实际数据进行了分析。通过对这 49 篇所采用的数据挖掘方法的内容分析，结果表明：40% 的文献采用了关系挖掘方法（以“关联规则”为主），28% 的文献采用了预测挖掘方法（以“决策树”为主），可见这两类方法是目前国内教育数据挖掘领域的主要方法，并与国际上数据挖掘方法的应用情况较为一致。

下面逐一介绍前面提及的 5 类挖掘算法，并介绍应用了这些算法的研究成果。

(1) 统计与可视化。统计趋向于将基于假设的检验作为最终目标，可视化是将数据信息转化为有意义的、易于理解的图像的过程，虽然它们常常不被认为是数据挖掘技术，但是作为对数据的探测方法，它们可以处理一些通常由数据挖掘解决的问题。蒋玉兰等以宁波电大 2000 级金融专业为研究对象，采用统计和可视化图形分析学生流失的态势，旨在找出远程开放教育中学生辍学的原因[5]。

(2) 聚类（聚类分析、离群点分析）。聚类分析和离群点分析，二者看上去是矛盾的双方，离群点分析相当于去寻找聚类之外的类别，但是这二者判断聚类或离群的依据均是各个样本之间的距离，因此笔者将这二者归为一类。

聚类是一个将物理或者抽象对象的集合分成由相似的对象组成的多个类或簇的过程，

常见的聚类算法有 EM 算法与 k - means 算法，如刘钢等人（2011）采用 EM 算法对移动百科网站（www.hellobaike.com）的新注册用户进行聚类，以快速把握新注册用户的特征[6]。

数据库中可能包含一些数据对象，它们与数据的一般行为或模型不一致。这些数据对象是离群点（outlier）。离群点数据分析称作离群点挖掘（outlier mining），可以假定一个数据分布或概率模型，使用统计检验检测离群点；或者使用距离度量，将远离任何簇的对象视为离群点。Ueno（2004）使用在线离群点检测方法分析学习者非常规学习网络课程内容的反应时间数据，指导网络教学[7]。

(3) 预测（决策树、回归分析、时序分析）。决策树、回归分析、时序分析等算法均可实现某种预测功能，即由一些已知属性或记录来预知另一属性取值或新记录取值。决策树是用图形方式描述正在考虑中的某项决策以及选择这个或那个备选方案的潜在后果，在将来的某些情景或行动后果不确定时采用，它对每条由事件和决策构成的逻辑路径，都综合考虑相关概率和得失。黄勇等（2006）尝试采用决策树分类的 D3 算法，构造学习者学习能力决策树，对学习者的数据库进行分类，将学习者分成学习能力强和学习能力弱的两大类[8]。

(4) 关系挖掘（关联规则挖掘、时间序列模式挖掘、相关挖掘）关系挖掘，即可挖掘在某一空间上共现而产生的共现关系，也可挖掘在某一时间段上相继出现而产生的序列关系。关联规则挖掘技术用于从用户访问序列数据库的序列项中挖掘出相关的规则，能够揭示学习者访问一些内容的同时会访问哪些内容，借此找出具有相关内容的网页，可更好的组织课程页面和推荐页面，尽可能缩短相关内容的分布距离，或提供便捷的路径指引。如吴兵等人（2009）构建了一个基于关系演算的可变精度粗糙集模型，探索得到了开放教育学习者成绩的初步知识规则，为学校的管理工作指明了方向[9]。时间序列模式挖掘试图找出页面依照时间顺序出现的内在模式，能够揭示哪些内容能够激发对其它内容的访问，可以用来对学习者的浏览趋势分析，解决远程教育中针对各种层次学生进行因材施教等问题。如赵东轮等人（2008）基于隐马尔可夫模型（HMM），以实时交互言论的行为类型为观察序列，以小组讨论状态为状态集合，识别出了分享信息、协商冲突和共同创造等三类协作交互模式[10]。

(5) 文本挖掘。Web 文本挖掘主要是对 Web 上大量文本集合的内容进行总结、分类、聚类、关联分析以及运用 Web 文档进行趋势预测等，是针对非结构化或半结构化的数据集。Web 内容挖掘大多是基于文本信息的挖掘，这类方法相对更加困难、复杂，Dringus 和 Ellis（2005）使用文本挖掘策略对异步讨论区进行评价 [11]，国内黄荣怀教授领导的协作学习研究团队也采用文本挖掘技术对交互文本进行分析，从中发现交互内容和交互模式，并开发出了一款用于交互文本分析的工具[12][13][14][15]。

4.教育数据挖掘的未来研究重点

综合上述国内外已有研究成果,教育数据挖掘的未来研究重点将体现在教育数据挖掘方法和数据的标准化研究、开发易于使用的教育数据挖掘工具以及特定的教育数据挖掘技巧研究等三个方面。

(1) 教育数据挖掘方法和数据的标准化研究。标准对于任何系统都很重要,教育数据挖掘领域也需要整合网络学习环境的普适化的工具和方法。数据库之间采用不同的数据存储类型、对数据的不同定义等问题导致了不一致的现象,数据标准化的主要功能是消除变量间的量纲关系,从而使数据具有可比性。数据标准化、规范化是实现信息集成和共享的前提,在此基础上才能达到信息的准确、完整和及时,没有数据标准化,信息共享就无从谈起,没有信息共享则没有普适化的可以应用到任何教育系统的工具。因此,数据的标准化和预处理任务是必需的。

(2) 开发易于使用的教育数据挖掘工具。数据挖掘结合人工智能、统计学和数据库技术等多个学科的思想,数据挖掘技术是一种面向应用的复杂技术,应用难度很大。很多数据挖掘工具封装了挖掘算法、可视化技术等,例如应用于商业领域的 DBMiner、Clementine、Intelligent Miner 等,然而这些工具不是专门为教育领域而设计,对很多教育工作者来说设计的过于复杂,不易于使用,在挖掘方法和数据标准化的基础上,针对教育领域的特点,开发一些专门的教育数据挖掘、统计和可视化工具,设计更加直观和易于使用的接口,以帮助教育工作者对于不同层次的教学过程进行分析。

(3) 特定的教育数据挖掘技巧研究。这一领域的研究者群体中很大一部分是计算机专家,他们缺乏教育和心理学理论,即使通过数据挖掘能够获知学习者的行为,但是不能解释学习者产生这种行为的原因,提不出合理的预测和解决方案。教育领域有其自身的特点,某些具体对象的属性难以用数量方法描述,信息具有隐含性和模糊性,网络教学信息数量大,包括教师教的信息、学生学的信息及交互信息,这些信息很难进行挖掘。数据挖掘在教育系统中的应用,需要考虑教育情境做相应的调整,结合特定的整合教育领域知识的数据挖掘技巧,可以借助教育测评技术、教学理论和教育心理学理论等寻求突破口,在有效且易于使用的数据挖掘工具支持下,将特定的教育数据挖掘技巧整合到网络学习环境中,使得所有的数据挖掘任务都能成为一个应用,所获得的反馈和结果能够直接被应用到网络学习环境,更好地为教育教学服务。

(参考文献 略)

学习分析技术研究与应用现状述评

摘自 中国电化教育 (2012.8), 作者 李青、王涛
(北京邮电大学网络教育学院)

摘要: 学习分析是一类运用先进的分析方法和分析工具预测学习结果、诊断学习中发生的问题、优化学习效果的教学技术。随着教育信息化的发展和在线学习方式的普及, 学习分析已经开始被运用于高等教育的教学实践中, 并取得了一定的效果。为了解学习分析技术的研究和应用现状, 本研究运用文献分析法辨析了学习分析的基本概念, 讨论了其组成要素和应用模型。调查了学习分析技术的技术来源、分析方法和分析工具。该文还通过案例分析法, 搜集和整理了学习分析在国外高校中应用的一般情况和典型案例, 剖析了该技术的优势和不足, 以期对今后国内学习分析技术的理论研究和实践提供参考。

关键词: 学习分析; 文献研究; 教学评价; 数据挖掘

一、引言

学习分析 (Learning analytics) 是近年来教育技术领域内迅速发展的新热点, 它是运用先进的分析方法和工具预测学习结果、诊断学习中发生的问题、优化学习效果的一类教学技术的集合。虽然传统教学过程中也评估学生成绩、分析教学过程, 以提高教学质量, 但这些活动中采集的数据往往非常有限, 信息化程度不高, 而且分析结果用于干预教学的周期过长, 效果较差。随着教学资源越来越网络化, 以及基于网络的学习方式的普及, 我们能够获取的学习行为和学习结果数据也更加丰富。这就为学习分析技术的产生和应用奠定了物质基础, 并提出了迫切的需求, 使之进入人们的视野并迅速发展。“学习分析”的理念一经提出, 就受到教育界的广泛关注。美国新媒体联盟(NMC)与美国高校教育信息化协会(EDUCAUSE)合作的“地平线项目(The New Media Consortiums Horizon Project)”在其 2010 至 2013 年的年度报告中均预测学习分析技术将在未来的四到五年内成为主流^①。以“学习分析”为主题的“学习分析技术与知识国际会议”先后于 2010 年 2 月和 2012 年 4 月召开了两次, 会议涉及学习分析技术在技术、社会和教学等维度的整合, 以及知识建模和表征、知识工作和分析等领域的现状及发展策略等专题目。

二、学习分析的概念、要素和模型

目前, 学术界对“学习分析”尚未形成一个严密的定义, 不同的学者或研究机构对其涵义和外延有不同的理解。加拿大阿塞巴斯卡大学的 G. Siemens 教授认为: “学习分析是利用数据挖掘成果, 学习者产生的数据和分析模型探究信息和社会联系, 并且对学习做出预测和建议”。美国高等教育信息化推进组织 EDUCAUSE 则认为: 学习分析“使用数据和模型预测学生学习进程和绩效以及使用该信息进行干预”。新媒体论坛(NMC)在 2011 年度的地

平线报告中指出：学习分析松散地组合了一系列数据收集工具和分析技术，研究学生的投入情况、绩效和学习进展情况，以期及时将研究结果用于修订课程、教学和评估。虽然以上专家学者在对学习分析的定义上有一些差异，但在基本概念还是有较为一致的观点，认为学习分析具有以下特征：

(1) 多样化的数据来源。用于分析的数据大部分来自 LMS、CMS 和学生档案库等数据库。也有些来自于学生个人的非正式知识管理系统(如个人主页、博客、微博等)或是学生在传统环境下的数字化学习资料、作业、作品。数据的采集过程应该是自动化的，具有海量的数据规模。

(2) 采用现代化的分析工具和分析方法对数据进行加工、挖掘和分析，区别于传统的教学评估模型和评估方法。此外，数据的分析过程也是自动化的，并可根据需求实时反馈分析结果。

(3) 提供可视化的分析结果。学习分析的主要目的是预测学习结果和提高学习绩效，以可视化和直观化形式显示数据将便于学生和教师对自身情况做出判断。

(4) 面向学生和教师。学习分析的直接服务对象是教师和学生，就在线学习过程中发生各种数据提供分析和建议：通过对学习情况的反馈帮助教师提高教学质量、教学水平和职业技能；帮助学生提高学习质量和课程通过率。对于学习分析的组成要素和应用模型，不同的专家也有不同的见解。Malcolm Brown 认为学习分析是教育技术发展的第三次浪潮，其核心在于搜集和分析与学习行为相关的数据，包含数据收集、分析、学习、受益方和干预五个大要素。他强调学习行为是学习分析关注的重心，也是其区别于其他分析方法的标志。学习分析就是要了解学习者的学习情况，如他们在做什么，阅读什么内容，如何加工信息等等。分析可针对学习者个人或者某一群体，也可以两者兼顾。分析结果可以反馈给教师、学习者、管理者，并对教与学的活动进行干预。一般来说，教师和学习者在课程层面应用分析数据，而管理者则应用在更高的组织、部门、制度决策的层面上。

Tanya Elias 提出了一个更为复杂的学习分析应用模型。学习分析利用了计算机、人力、理论和组织机构四种技术资源。计算机用于计算处理分析数据，在分析过程中需要人的参与，分析模型以多种理论为指导，并通过组织机构实施决策支持。学习分析以上面四种资源为基础，分为数据收集、信息加工和知识应用三阶段和获取、选择、聚集、预测、使用、优化六种活动，这三个阶段在教学实施过程中不断循环。

G. Siemens 从另一个角度提出了学习分析的应用模型。他认为学习分析包括收集、分析、预测和调整几个阶段。从数据来源来看，学习者的数据主要有两类：第一类是学生自己

发布的数据，如在微博、Facebook、博客等个人学习环境中发布的状态、帖子及更新，在学习管理系统中的记录等；此外还有一些隐性数据，如用户在各种网站上的个人信息、资料等。另一类数据并不直接来自于学生，而是通过语义化和关联分析等手段，将课程数据、学习者的数据、资料信息都收集起来进行分析得出的再加工数据，例如通过分析可判断学生对于某一概念的了解情况、学习者解决某一问题的成功或失败的信号等。教师可以将这些分析的结果用于预测教学，并反馈到教学的实施中，指导和调整教学，使课程更加个性化，满足不同学习者的需求。

Wolfgang Greller 等从目标、数据来源、分析方法、约束条件、竞争力、利益相关方构建了学习分析的要素模型。Greller 认为学习分析的目的在于预测学习结果和帮助学习者反思，采用的分析方法涉及教育数据挖掘、机器学习和统计分析。学习分析的受益者包括学生、教师、学校和服务提供商，在实施过程中受隐私和伦理道德因素制约。

根据以上分析，我们可以看出学习分析至少可分为目标、对象、约束条件、数据来源、处理方法等要素，这些要素反应了学习分析得以实施的内部和外部条件。同时，根据以上学者的观点，我们也可以把学习分析的过程粗略地划分为数据收集、数据处理以及数据应用和反馈三个阶段，这三个阶段在实施过程中不断循环往复。

三、主要分析方法和分析工具

1. 技术来源

虽然学习分析在教育技术领域是一类新兴的技术，但是其使用的技术早已应用于商业和信息技术产业的多个领域，如商业智能（Business Intelligence）、网络分析(Web Analytics)、教育数据挖掘(Educational Data Mining, EDM)、学术分析(Academic Analytics)等。这些技术对海量数据进行分析和处理，并且具有很高的智能。可以说，学习分析是这些技术在教育领域中的继承和渗透，并与之密切相关。

以上技术中的商业智能和网站分析在互联网运营管理中广泛使用。商业智能以数据仓库、联机分析处理、数据挖掘等技术为基础，从不同的数据源中提取数据，将数据转换成有用的信息，为企业管理、策略制定提供了支持。网站分析是对互联网数据定量分析，包括跟踪、收集、测量、报告和分析数个流程+达到优化网站、发挥电子商务营销主动性等目标，通过分析网站访客的行为，评估和调整网站的运作。

和教育密切相关的技术包括教育数据挖掘和学术分析。前者应用数据挖掘技术从来自于教育系统的数据库中提取出有意义信息的过程，这些信息可以为教育者、学习者、管理者、教育软件开发者和教育研究者等服务。利用数据挖掘技术，可以发现数据中的教育模式和

规律，帮助管理者做出正确的教育决策，并且应用到教学中。学术分析是商业智能在教育领域中的应用，与教育数据挖掘旨在分析数据中的教育模式和规律的目的不同，其更加关注学习者的学习行为。

从技术发展的历史延续性来看，学习分析技术是计算机辅助教学管理与教育数据挖掘技术发展的产物；而从跨领域横向发展的角度来看，学习分析技术又是网络分析技术与学术分析技术综合应用的结果。根据这三者的特点，我们还可以进一步从数据加工的实时性、搜集数据的来源和面向的用户三个维度对它们作进一步区分。学习分析更加偏向个人、实时和多元数据分析，而教育数据挖掘则偏向非实时、机构应用。学术分析则偏向个人、非实时和多元数据。

2. 分析方法

除了以上几种普适的分析技术外，学习分析也在不断吸收和整合其他的技术和方法。包括社会网络分析法、话语分析法和内容分析法等。这些新分析方法的引入，大大丰富了学习分析的数据加工方法和策略。

社会网络分析法(Social Network Analysis)是由社会学家根据数学方法、图论等发展出的定量分析方法。以关系作为基本分析单位，描绘和测量行动者之间的关系及通过这些关系流动的各种有形或无形的东西，如信息、资源等。社会网络分析法最早应用于人类学研究，近年来，在社会学、心理学、教育学等领域都得到了广泛的应用。一些学者也开始利用这种方法来研究在线学习环境中的社会网络，分析活动者之间的关系。如 Palonen 和 Hakkarainen 利用社会网络分析工具分析学生在 CSILE 网络协作平台上的交互过程，并对学生网络协作学习情况与性别、成绩之间的关系进行研究。

话语分析法(Discourse Analysis)最早应用于语言学中。主要对话语结构形式、规则等进行研究，后经发展也开始运用到教育研究中。Arnseth 认为话语分析能够帮助我们理解知识是如何产生和建构的。网络学习过程中的话语不仅仅指面对面的对话，还包括文本内容和通过各种媒介实时或异步传播的交流内容。通过话语分析，我们可以发现学习者在对话中的知识建构与共享，意义表达等过程。

内容分析法(Content Analysis)是对传播内容作客观而系统的量化并加以描述。透过量化分析以及质的分析，以客观、系统的态度对文件内容进行研究和分析。分析内容包括传播中的各种语言和特性，以及传播内容对于整个传播过程所发生的影响，藉以推论产生该项内容的环境背景和意义。利用内容分析法，我们可以对学习者的学习过程与资源以及学习者学习反应的反馈信息进行定量和定性的分析，探索学习者的行为模式并进行预测。

3. 分析工具

按照专门程度来分，学习分析工具可分为专用和通用工具两类。“专用工具”是专门开发用于分析学习活动和评估学习者的工具。典型的专用工具有 Socrato、SNAPP 和 LOCO—Analyst 等。专用工具一般都是学习分析项目中的核心分析工具和开发成果，即专门针对某一项目的要求设计和开发，在项目实施中运用该工具搜集和分析学习者数据。并指导教学。而“通用工具”则是原来应用于互联网、可用性设计等其他领域内，被转用于教育情景下，用来分析学习者如何使用教学系统的工具，主要有 Mixpanel Analytics、Userfly、Gephi 等几种。

四、学习分析技术的典型应用案例

目前，国外多所大学已经开始开展学习分析的实践，包括开发学习分析系统和工具，收集学习者数据，分析学习行为模式，对学习进行干预和预测等。

下面我们还将以北亚利桑那大学的 GPS 项目和澳大利亚 Wollongong 大学的 SNAPP 项目为案例进一步介绍学习分析技术的应用情况。

1. 北亚利桑那大学的评价绩效状态系统

GPS(Grade Performance Status)是北亚利桑那大学研发的线上学生绩效评价系统，主要用于评估该校全日制学生的课堂学习绩效。系统能够收集学习者在课堂中的表现评级，并给予相应建议。然后通过邮件发送给学习者。GPS 系统邮件主要包括出勤情况、学习成绩和课业问题三类，学生收到邮件后应对相关问题给予反馈。

GPS 系统的工作流程如下：教师登录到 GPS 系统后可查询出勤率、学生评级等信息，也可以对学生的情况提供个性化的评价。邮件中除了教师评价外，还可包括重大活动、校历、事件的提醒以及校园网上资源的链接等。邮件发送后，系统会在学生个人门户网站上提醒查收，而邮件也会在责任教师的 GPS 页面中归档，使之能及时审查、更新和修改意见。教师可自主设定邮件发送频率，通常为一周一封。以避免频繁发送邮件导致学生疲于反馈。邮件可以针对单个学生发送，也可以对班级所有学生发送。学生通过邮件获得建议和资源，及时与指导教师沟通交流。GPS 系统通过这种形式帮助学习者提高他们的学习质量。

2. Wollongong 大学的学习网络可视化与评估项目

“学习网络可视化与评估项目”是由澳大利亚 Wollongong 大学领导，多个学校共同参与。项目基于学习分析的理念研发了学习网络可视化评估工具(SNAPP)。该软件可以从学习管理系统中收集学生的学习行为信息如在线时间、下载次数等；从论坛中提取学习过程中的交流互动数据，如发表帖子数、讨论内容、互动频次等。该项目面向在线学习的学习者，利

用 SNAPP 记录和分析其学习活动情况，使教师在学习的任何阶段都能确定学习者的行为模式，评估对学生的学习网络，并分析其对学习的影响。这些分析结果可以用于调整教学，为学习者提供指导，提高其学习能力。

SNAPP 可以搜集的论坛中的帖子总数目、单个用户的帖子数和回复数，绘制用户的社交网络图等。通过对这些数据的加工处理，系统可以分析出以下的信息：(1) 识别出远离网络的学习者，提示教师给予关注；(2) 识别出班级中的主要的信息传播者，他们是信息交换的节点；(3) 区分绩效良好和不佳的学习者，于学习评估前进行干预；(4) 预测学习社区在班级中的发展程度；(5) 通过“快照”提供学习“前”与“后”情况对比，观察各种活动对学习产生影响，从而优化学习活动设计；(6) 为学习者观察自己的表现提供依据。

五、结束语

学习分析技术涵盖了教育过程最主要的领域，在教学、学习、管理、研究等方面都有很大的潜力。当教育信息化成为教育机构的目标，以技术促进教育改革成为教育的使命之一，学习分析技术大有可为。对教育机构而言，数据和信息爆炸是一种挑战，而在这个数据和信息的泥沼中，学习分析理论为人们指出了一条出路，若能将学习分析理论与其它高效的新学习实践结合起来，就能为学习和教学的改善创造出无限的可能性。2011 年，盖茨基金会的“下一代学习计划”宣布将为利用科技解决方案应对学习挑战的教育机构提供两千万美元资助。第一批资助项目就包含学习分析在内。和传统方式相比，学习分析技术有很多的优势：

★为学习和评估方式创造出更多的可能性和灵活性，促进教学模式的改变和教学过程的优化。

★可通过对教学管理系统数据的挖掘，识别学习过程中出现问题的学生，并主动帮助或干预。

★为学习管理系统提供智能化和自动化的监测和分析手段，能减轻教师的工作负担，提高学习者的用户体验。

★能够引导学习者自我管理和自我激励，培养良好的学习习惯。

★通过数据挖掘和分析获得全面的分析报告，了解机构的优势和不足之处。

★可提高教育管理的决策效率和改善教育机构的资源分配方式。

虽然学习分析近年来得到了广泛关注，同时也面临着许多挑战，依然离大规模应用尚远。笔者认为其主要原因在于：(1) 学习分析要求收集并梳理不同来源的数据，且数据格式各不相同；(2) 学习分析涉及学生隐私和档案等问题，如何合理利用这些数据；(3) 学习分析的理论和实践还不成熟，不少研究者和开发者对于学习分析还处于贴标签和概念炒作的阶段：

(4) 尚未有较好的分析模型，现有的多个应用项目也仅完成对学习数据的表面挖掘，对实践未能产生革命性的影响；(5) 传统观念的改变是一个逐步的过程，必须要教育和引导用户，并且提供适当的培训和技术支持。因此，要推动学习分析技术的进一步发展应用，需要将信息技术更高效更全面地运用于课程和教学中，让教师和学生感受到学习分析的技术潜力及其对教育的积极作用。

北京农学院 8 号学生宿舍楼卫生情况影响因素的实证分析

(节录)

(11 级会计 1 班 王婧 201120441252 指导老师: 夏龙)

1. 引言

大学是人生最美好的阶段,是过渡到成人的关键时期,也是实现自我价值的理想场所。在这里,我们生活、我们学习、我们成长、我们成才。在这期间,锻炼自己的自理能力与积累自己的人生经验一样至关重要。自理能力仅仅是大学生活中的一部小插曲,但却充分体现着当代大学生的生活观念,铸就成我们将来面向社会的一种基本生存能耐。现代高校的学生宿舍是学生生活、学习、休息、娱乐、社交的重要场所,是学校开展思想政治工作的阵地。那么,围绕着宿舍卫生而展开的宿舍生活无疑成为一个可以聚焦我们自理能力的大战场。

良好的宿舍生活环境主要体现在拥有良好的宿舍卫生环境。据有关统计,高校学生每天在宿舍的时间超过 10 小时,有的甚至达到了 15 小时。宿舍是学生学习、生活、休息、社交的主要场所,从某种程度上说,学生宿舍就是学生在校期间的家。因此,宿舍卫生状况如何,直接关系着同学们的生活质量。众所周知,良好的学习环境和居住环境能够使人保持健康的心理和愉悦的心情,使人积极蓬勃向上,反之,则会影响人们的心态。宿舍环境状况不仅是一个生活习惯的体现,更是整个宿舍精神面貌的体现。因此,宿舍环境对学生身心健康发展至关重要。良好的卫生居住环境能够体现健康的审美情趣,陶冶情操,完善个人素质。同时,良好的宿舍环境有利于同学之间增进友谊和感情,有利于培养他们的主人翁精神和集体荣誉感,体现大学生精神面貌。而学生宿舍卫生管理如何更关系到学校形象,反映学校的学风、校风,是学校校园精神文明建设、和谐校园建设的重要组成部分与重要保证。

毋庸置疑,一个宿舍的卫生情况不仅与宿舍同学的生源地构成状况、家庭环境背景等学校随机分配所造成的外部环境、客观因素有关系,而且与宿舍同学的自身努力情况、奉献精神水平和责任感程度等个体特征、主观因素密切联系。它是学校随机分配组成的外部环境和宿舍成员自身的个体特征两个因素交织在一起共同决定的结果。本文将其置于一个统一的研究框架进行讨论,具体而言,是基于社会调查数据讨论宿舍成员来源构成、学习成绩、政治面貌、家庭背景和校内工作与宿舍卫生水平之间的定量关系,并对回归方程进行恰当的解释,以期对理解宿舍卫生水平的形成机制,并对高校宿舍卫生管理提出相关建议,对促进其卫生水平提高有所启发和帮助。

2. 研究框架

2.1 研究设计

本文的研究设计如图 1 所示：

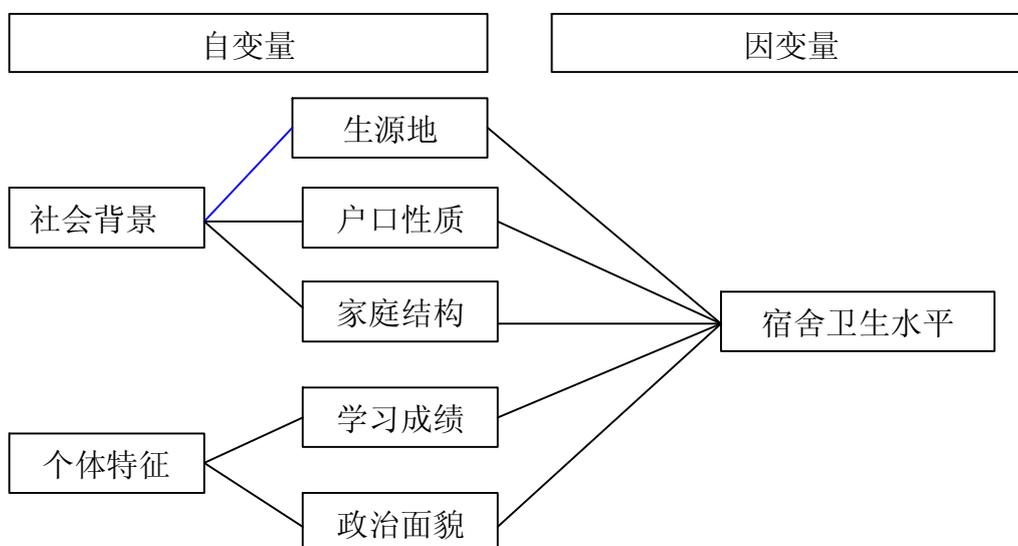


图 1：本文的研究设计

学生宿舍卫生与宿舍中的每一位同学都息息相关，每个人都需要对宿舍的卫生负责，而宿舍中的每一位同学对宿舍卫生水平都会产生影响。高校宿舍主要是学校主要采取随机分配的方式进行组合的，其社会背景构成在同学入校前就已决定，客观存在，无法更改。可以说，一个宿舍的卫生水平在学生进入大学开始学习前，从某种程度上说就已经决定了。

宿舍同学社会背景的不同，一方面原因是通常来自不同的省份，很少出现一个宿舍同学全都来自一个省份的情况。而不同的地区有不同的社会习俗，不同的社会背景对个体产生的影响不同，从而使同一宿舍来自不同省份的同学对待宿舍卫生的态度、习惯等都有所不同。这主要体现在三个方面：一是南方与北方的地域差异。一般说来，北方人豪爽大气，在生活上不拘小节，对待卫生相对而言不是太在意；而南方人精致讲究，追求生活情调，在卫生上要求较高。从这方面考虑，宿舍中南北方学生的数量比例对宿舍卫生应存在影响，宿舍中南方学生越多，宿舍卫生水平越高。二是生活用品多少的影响。外地学生远离家乡来到北京求学，一年四季的衣物及其他生活用品都放在了宿舍，而北京学生离家较近，可以每周都回去，所以只需带上一周所需要的生活用品即可，因此，外地学生越多，宿舍所放东西越多，相对而言宿舍会变得更加拥挤，从而显得更为杂乱。总而言之，宿舍中来自南方的学生越多，本地学生越多，宿舍卫生水平越高。

宿舍的同学社会背景的另一个不同是来自城市还是农村的不同。经济发展程度的差距、

人们生活水平的高低等因素必然造成居住在城市和农村的人形成不同的生活观念、生活习惯和生活方式,这其中就包括其对待家庭卫生的态度、摆放物品的习惯和打扫家庭卫生的频率、方式等。一方面,来自农村的同学所居住的房屋一般来说相对加大,这其中有房价的影响,同时,务农的劳动方式以及相对落后的经济发展水平使其房间与现代的生活方式必然有一定的差距。宿舍中来自农村的学生越多,宿舍卫生水平越低。另一方面,来自城市的学生大多家庭经济条件较好,其生活方式与现代高校宿舍相近,能够较快适应宿舍生活。但与此同时,较高的家庭经济条件也容易造成父母对于子女的过度宠爱,使其独立生活时,自理能力较差,如在家房间是父母收拾,进入宿舍也会使生活物品摆放杂乱,不会整理宿舍,相对农村孩子而言更为娇生惯养。但是综合城市孩子的两方面因素来看,后一方面的因素会和前一方面的影响相抵消,最后,我们认为,宿舍中来自农村的孩子越多,宿舍卫生水平越低。

家庭结构组成背景是宿舍同学社会背景的又一个不同,在家中是独生子女还是有兄弟姐妹都会对其生活自理能力产生影响,从而影响宿舍卫生。如果家中只有一个孩子,父母会相对而言更为宠爱一些,虽不至于溺爱,但所有的关注都在这一个孩子身上,必然会使孩子各方面都有人关心,造成孩子的自我约束、自我思考、自我生活能力较差,个性更为自我一些,对他人不太懂得包容和忍让,对宿舍关系和宿舍卫生这种需要一定的宽容和牺牲的方面有所影响。而若果家中孩子较多,父母的注意力就会分散,无法照顾到每一个孩子,从而是孩子不得不在父母照顾不周的地方学会自己成长,同时多个孩子共同成长使其性格会较为温和,对待他人会更为大度和宽容,更加懂得牺牲、奉献与合作,在日后的宿舍相处和宿舍卫生整理中这种影响会有所体现。相比较而言,宿舍中是独生子女家庭出身的学生越多,宿舍卫生水平越低。

除了宿舍学生的社会背景会影响到学生宿舍卫生水平外,宿舍学生的个体特征也是影响学生宿舍卫生水平的重要变量。个体特征是宿舍学生后天获得的可改变的个人属性和特点,比如学习成绩和政治面貌。虽然学习成绩不能代表生活独立自理能力,但毫无疑问,对于学生来说,个人学习成绩是最重要的个体特征。学习成绩越好,其学习能力、合作能力、责任心等方面相对而言会更强一些。我们将是否获得奖学金作为评判宿舍学生学习成绩好坏的标准,同时,又因为奖学金的评选是根据业务素质得分、身体素质得分、思想道德得分、发展性素质得分等各方面所计算出来的一个综合素质得分,与仅仅比较学习成绩更为全面和客观。获得奖学金不仅仅只需要较高的学习成绩,还要求身体素质、思想道德等方面的优秀。所以,我们认为宿舍中获得奖学金的学生人数越多,所占比例越大,其宿舍卫生水平也会越好。

中国共产党的宗旨就是全心全意为人民服务，可见，作为中共党员的首要前提便是奉献精神。那么，加入中国共产党、成为中国共产党党员的学生，其态度、精神、行动等方面在宿舍卫生水平上应该都有所反映。作为中共党员，在宿舍中应发挥模范带头作用，在自身搞好个人卫生的同时，督促宿舍其他同学整理好宿舍卫生。因此，我们认为宿舍中的学生党员越多，宿舍卫生水平越高。

2.2 变量选取

本文利用回归分析来验证本文的研究假设。其数据来自于北京农学院 8 号宿舍楼学生宿舍调查，该宿舍调查主要通过宿舍调查问卷的形式完成。为了深入了解宿舍卫生状况，掌握学生的切实需求，倾听同学们的心声，改善宿舍的管理工作，提高同学们的卫生管理意识，切实为同学们服务，使同学们能在一个舒适的环境下生活与学习，我们设置了调查问卷，并实地查看了部分宿舍卫生。问卷经过精心选择维度，合理配置题目，预测验，征集修改建议，重新集中研究，二次预测验，修改定稿等多个环节，保证了问卷的合理、科学、规范和可操作性。对于问卷调查，首先在班级中进行了小规模的抽样，对调查问卷进行检验，修正完宿舍调查问卷后，进行了大规模的北京农学院 8 号宿舍楼学生宿舍整体调查。共计发放问卷 50 份，其中有效问卷 50 份，调查对象采用了随机的形式，主要是北京农学院 8 号宿舍楼学生宿舍的宿舍长。第一次调查的宿舍卫生水平是以 2013 年 10 月份第二个星期的宿舍卫生评分作为标准的，考虑到这个标准只针对一天，并不能如实、全面的评判一个宿舍卫生水平，我们又找到 8 号宿舍楼楼长，要到了上一学年整个宿舍楼的卫生评分，取其平均分作为宿舍卫生水平的评判标准。

我们以上一学年宿舍卫生检查评分的对数平均分 ($\lg\text{grade}$) 来表示被调查宿舍的卫生水平，它是本文中的因变量，是被调查宿舍上一学年平均每星期的宿舍卫生评分 (grade , 分/星期) 的自然对数。

代表学校随机分配的生源地以 origin of student 表示，简写为 original 。我们把宿舍学生的生源地构成情况转化为外地学生占宿舍学生人数的比例，根据本文的研究设计，我们预期该变量与 $\lg\text{grade}$ 正相关，不过，也有可能为倒 U 型。被调查宿舍的户口性质和家庭背景构成情况分别以 $\text{registered residence}$ 和 $\text{family composition}$ 表示，简写为 residence 和 family ，统计被调查宿舍学生来自城镇人数和独生子女家庭人数，同样也计算其人数所占宿舍学生比例。根据本文的研究设计，我们无法明确确定户口性质与 $\lg\text{grade}$ 的关系，当来自城镇的宿舍学生人数越多，宿舍卫生情况越好时，符号为正，当来自农村的宿舍学生人数越多，宿舍卫生情况越好时，符号为负。同时，根据本文的研究设计，我们预期家庭结构背景与宿舍卫生

水平负相关。

表 1 变量的描述性统计量（略）

宿舍学生的学习情况以 *study* 表示，为了研究问题的方便，我们把宿舍学生的学习情况转化为是否获得奖学金的问题，统计被调查宿舍上一学年曾获得过奖学金学生人数，并计算其所占宿舍学生人数比例。按照本文的研究假设，我们预期学习情况与宿舍卫生水平正相关。宿舍学生的政治面貌以 *political party* 表示，简称为 *party*，它是以是否为中共党员（包括预备党员）为标准的。对于 *party* 这一自变量，我主要从两方面考虑，一方面是在统计被调查宿舍的党员人数后，计算出其所占宿舍学生人数比例；另一方面是调查宿舍是否为学生党员宿舍，即宿舍中是否有中共党员（包括预备党员）。按照本文的研究假设，我们依然预期它与宿舍卫生水平正相关。

我们所建立的一般性模型如式（1）所示，所有变量的描述性统计如表 1 所示（略）：

3. 实证研究

3.1 社会背景对宿舍卫生水平的影响

生源地对宿舍卫生水平的影响如式（2）所示（公式略）：

从整体上看，*origin* 在 1% 的显著性水平上显著，它的符号也符合我们的预期，并且能够解释整体因变量变异的 13.5%，有一定的解释力。式（2）说明，当宿舍外地学生人数每增加 1 人，即在 6 人组成的宿舍中外地学生所占宿舍人数比例每增加 16.67%，5 人组成的宿舍中外地学生比例每增加 20% 时，宿舍卫生评分减少 0.762 分。

户口性质对宿舍卫生水平的影响如式（3）所示（公式略）

在式（3）中，*residence* 在 5% 的水平上是显著的， R^2 为 0.102 说明 *residence* 可以解释整体因变量变异的 10.2%，对因变量的解释力度小于 *origin*。注意到 *residence* 的符号是正号，它说明宿舍卫生水平与户口性质正相关，城镇户口学生人数越多，宿舍卫生水平越高。宿舍中，城镇户口学生比例每增加 1%，宿舍卫生评分大约增加 0.048 分。

家庭背景对宿舍卫生水平的影响如式（4）所示（公式略）：

与 *residence* 一样，*family* 在 5% 的显著性水平上也是显著的，它的符号同样符合我们的预期， R^2 为 0.124 说明 *residence* 可以解释整体因变量变异的 12.4%。通过比较可以发现，*residence* 对因变量的解释力度小于 *origin*，但大于 *residence*。式（4）说明，随着宿舍中独生子女家庭出身的学生人数的增加，宿舍卫生水平也随之下降，即宿舍中独生子女所占宿舍学生比例每增加 1%，宿舍卫生评分将下降大约 0.056 分。

3.2 个体特征与宿舍卫生水平

为了进一步考察生源地、户口性质和家庭结构对宿舍卫生水平的影响机制，我们估计了式（5）和式（6）。

我们根据研究框架构造了式（5）来描述了学习成绩与宿舍卫生水平的关系（公式略）

可以看出，study 在 1% 的显著性水平上是显著的，同时，它的符号也符合我们的预期。 R^2 为 0.232，即有 23.2% 的整体因变量变异可以被 study 所解释，其解释力度在所有的因素中表现为最高，拟合度在所有自变量中是最好的。随着宿舍中奖学金获得人数的增加，宿舍卫生水平也在上升。宿舍中获得奖学金的人数所占宿舍学生比例每上升 1%，宿舍卫生评分将增加 0.065。

式（6）和式（7）也是我们根据研究框架构造的，描述政治面貌和宿舍卫生水平关系的方程式（公式略）。

上式的自变量 party1 却是不显著的，不具有统计学意义，这与我们的预期不相符合。这说明宿舍中学生党员人数的多少与宿舍卫生水平之间没有关系（公式略）。

上式的自变量 party2 的 P 值为 0.815，远远大于 α ，所以接受原假设，说明是否为学生党员宿舍与宿舍卫生水平不具有统计学意义，这与我们最初的假设同样不相符合。

综合上述两方面来看，两个方程的 P 值均未通过 t 检验，这说明宿舍中学生的政治面貌与宿舍卫生水平之间没有关系。当然，这也可能是由于样本量太少的原因所造成的，在 3.4 的模型分析中有其具体分析。

3.3 整体回归

我们共有五个自变量，根据单变量回归结果，式（2）、（3）、（4）和（5）是描述单个自变量与因变量的最优方程。而且，学习成绩对宿舍卫生水平差异的解释力度最大（23.2%），生源地次之，在次之是家庭构成和户口性质，而政治面貌对宿舍卫生水平没有影响。我们以学习成绩为基础，通过逐步回归法考察变量间的多元回归方程。

最终的估计结果如表 2 所示（公式略）

（变量增减、公式改进、假设前提的验证以及文字分析略）。

大致来讲，在其他条件不变的情况下，一个宿舍中的外地学生每增加 1%，宿舍卫生水平将降低 7.075%，而独生子女家庭出身的学生每增加 1%，宿舍卫生水平降低 4.152%，来自农村的同学每增加 1%，宿舍卫生水平上升 4.022%。

3.4 模型分析

对于以上模型的建立，我们进行了以下分析：

第一，在众多的因素中，学习成绩是最重要的因素。宿舍中获得奖学金的学生人数越多，所占比例越大，宿舍卫生水平越高。可见，与客观的社会背景相比，宿舍卫生水平的高低主要与宿舍学生自身的个体特征相关，尤其与学习成绩相关。大学仍然处于人生中的学习阶段，大学生主要还是以学习作为主要的日常活动，是否获得奖学金可以作为衡量一个学生的大学生活是否丰富、充实、富有规律的客观标准。同时，宿舍卫生成绩作为奖学金评定中的一项，也会反映出宿舍卫生水平的高低，宿舍学生获得奖学金的人数越多，说明宿舍各方面更为突出，宿舍卫生水平也就更高了。

第二，学生宿舍卫生水平与宿舍学生的政治面貌无关，这与我们最初的预测相反。我们原本预计宿舍中学生党员越多，宿舍卫生评分会越高，但调查数据表明，是否成为中共党员以及宿舍中中共党员的多少与宿舍卫生水平没有关系，不具有显著性。我们是否可以这样认为：学生党员在宿舍卫生方面并没有应有的模范带头作用，或者说，学生党员在宿舍卫生方面的贡献并没有我们想象的那么突出。更进一步，我们提出疑问：加入中国共产党的标准是否公平和全面？衡量学生的思想道德素质的机制是否健全和周到？另一方面，我们从另外一个角度进行思考，学生加入中国共产，成为中共党员后，是不是就放松了原本对自己的严格要求，入党不仅没有成为学生增强为同学们服务的奉献意识和督促自己更为努力的动力，反而成为放松生活要求，降低卫生标准的借口。因此，学校更应加强对学生党员宿舍卫生的检查，加强对学生入党后各方面表现的监督，促使学生党员真正发挥其在学习、生活等各方面的模范带头作用。

第三，以 $origin$ 作为 X 对模型 VII 进行求导，我们可以得到式 (9) (公式与分析略)

分析：当宿舍内获得奖学金的学生比例小于 50.892% 时，外地学生越多，宿舍卫生水平越低；而当 $study$ 大于 0.50892 时，式 (9) 大于零，为正数，即在 $study$ 大于 0.50892 保持一个数不变的情况下， $origin$ 对 $grade$ 的影响为正，且随着 $origin$ 的增大， $grade$ 的增长速度也越快，即当宿舍内获得奖学金的学生比例大于 50.892% 时，外地学生越多，宿舍卫生水平越高，生源地对宿舍卫生水平的影响越大。那么，为什么会出现这一现象呢？我们对此进行了思考与探讨。首先，根据式 (2) 我们可知，如不考虑学习成绩的因素，生源地与宿舍卫生水平负相关，但是式 (9) 表明，当宿舍获得奖学金的学生人数大约超过半数以上时，外地学生对宿舍卫生水平的影响为正。因此，我们这样认为，学习成绩能够消除生源地对宿舍卫生水平的负面影响，且随着获得奖学金人数的上升，这种负面影响消除的也越多，达到一定比例，即 50.892% 时，生源地所带来的不利影响被学习成绩全部消除，而超过这个比例后，生源地将给宿舍卫生水平带来正面影响，且这种正面影响随着学习成绩的上升也越来越大。

由此可知，虽然社会背景，比如生源地，会对宿舍卫生水平产生负面影响，但个体特征，如个人的努力等可以消除这种影响，并且随着比例的上升，影响速度越来越大，最后反过来提高宿舍卫生水平。可见，虽然宿舍卫生水平在学生入学前在某种程度上就已决定，但还是能够通过宿舍成员的个人努力而得到改变的。

4. 结论与启示

基于调查数据，本文实证研究了影响学生宿舍卫生水平的主要因素。基本结论为下：第一，在众多的因素中，学习成绩是最重要的因素，宿舍中获得奖学金的学生人数越多，所占比例越大，宿舍卫生水平越高。第二，学生宿舍卫生存在严重的地域差异，社会背景对一个人的生活习惯具有重大影响。在同等条件下，本地学生多的宿舍卫生水平比外地学生为主的宿舍卫生水平高，如果按户口，即来自农村还是城镇在进行区别，这种差异将更加大，来自本地城镇的学生宿舍卫生水平比来自外地农村的学生宿舍卫生水平要高。第三，学生的学习成绩能够消除生源地所带来的不利影响。第四，随着独生子女家庭出身的学生所占宿舍人数比例的上升，宿舍卫生水平在逐渐下降。

根据上述结论，我们提出以下建议：首先，学生应发挥自身的主观能动性，在整理好个人卫生的同时，对宿舍卫生进行管理；其次，学生党员应加强自身的思想觉悟，以高标准继续严格要求自己，发挥模范带头作用；最后，学校应加强学生宿舍卫生的检查和管理工作，同时，在学生宿舍的分配上应考虑到生源地、户口性质和家庭构成对宿舍卫生水平的影响，尽量均衡分配，各宿舍间不要差异太大。