

基于在线评教的教学效能研究

——以 RateMyProfessors.com 为例

(申请清华大学应用心理硕士学位论文)

培 养 单 位： 社会科学学院

学 科： 应用心理

研 究 生： 刘明春

指 导 教 师： 何吉波 副教授

二〇二〇年六月

Teaching Effectiveness of Online Teaching Evaluation

—Take RateMyProfessors.com for Example

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Master of Applied Psychology

by

Liu, Mingchun

Thesis Supervisor Associate Professor He, Jibo

June, 2020

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：(1)已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；(2)为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容；(3)根据《中华人民共和国学位条例暂行实施办法》，向国家图书馆报送可以公开的学位论文。

本人保证遵守上述规定。

(保密的论文在解密后遵守此规定)

作者签名：_____

日 期：_____

导师签名：_____

日 期：_____

摘要

学生对于教师的整体评价可以反应教师的教学效能，因此学生对于教师的教学效能评价显得尤为重要。目前通过在线评教平台评价教师的教学效能已经非常流行，这些平台的评教结果可以帮助学生了解教师教学特点，也为教师提高教学效能提供一定的指导。而针对线下课程及在线课程，教师的教学行为存在明显不同，因此，本文将进行两个研究，分别探究教学效能的影响因素、教师在线下课程及在线课程的教学效能差异，以下是本文的主要发现：

研究一：本研究从 RateMyProfessors.com 网站获取了 919,750 名国外教授的评价数据。结果显示，第一，教授的教学效能综合得分保持不变，助理教授、副教授与教授的教学效能综合得分尽管存在差异，但效应量较小，可以忽略。教授的教学效能综合得分要高于助理教授及副教授；第二，授课难度较低并且给学生高分的教授，往往会获得更高的教学效能综合得分。第三，在 RMP 上的 12 门学科的教学效能得分方面，线下课程的教学效能综合得分均高于在线课程。

研究二：本研究基于《大学教师教学效果评价问卷》获取并分析了 375 名中国学生的教学评价数据。结果显示，第一，国内教师的在线课程及线下课程教学效果得分不存在显著差异；第二，副教授、教授在线课程及线下课程的教学效果得分不存在显著差异，但讲师在线课程的教学效果得分显著高于线下课程；第三，多维度对比国内教师的教学效果发现，在线课程的功课量要显著高于线下课程。

总之，本文揭示教师教学效能的影响因素，对于教师而言，通过关注学生成长、积极给予学生良好的反馈，可以帮助教师提高其教学效能。

关键字：教学效能；在线评估；RateMyProfessors.com；大数据；教学评估

Abstract

Students' overall evaluation of teachers can reflect teaching effectiveness. Therefore, students' evaluation of teaching effectiveness is particularly important. It is very popular to evaluate teaching effectiveness through online evaluation platform. The results of evaluation can help students understanding the characteristics of teachers' teaching, and provide guidance for teachers to improve teaching effectiveness. Therefore, this paper will carry out two studies to explore factors of teaching effectiveness, and the differences of teaching effectiveness between offline courses and online courses. The following are the main findings:

Study 1: this study obtains the evaluation data of 919,750 foreign professors from RateMyProfessors.com. The result shows that the overall score of professors' teaching effectiveness remains unchanged. Although there are differences in the overall score of teaching effectiveness between assistant professors, associate professors and professors, the effect size is small, which can be ignored. The overall score of professors' teaching effectiveness is higher than that of assistant professors and associate professors. Secondly, professors who taught less difficult courses and who were lenient with grades received a better rating from student. Thirdly, in terms of the teaching effectiveness of 12 subjects on RMP, the overall score teaching effectiveness of offline courses are higher than online courses.

Study 2: this study obtains and analyzes the teaching evaluation data of 375 Chinese students based on the Questionnaire of University Teachers' Teaching Effectiveness Evaluation. The results show that there is no significant difference for domestic teachers in the score of teaching effect between online and offline courses. Secondly, there is no significant difference in the score of teaching effect between associate professors and professors, but the teaching effect score of online courses for lecturers is significantly higher than that of offline courses. Thirdly, the result of multi-dimensional comparison of the teaching effect of domestic teachers shows that the workload of online courses is significantly higher than that of offline courses.

In a word, this paper reveals the factors of teachers' teaching effectiveness. Teachers should pay attention to the growth of students, and actively give students good feedback, which can help teachers improving their teaching effectiveness.

Keywords: Teaching effectiveness; Online evaluation; RateMyProfessors.com; Big Data; Teaching Evaluation

目录

第1章 引言	1
1.1 选题背景	1
1.2 研究目的	2
1.3 研究意义	3
1.3.1 理论意义	3
1.3.2 现实意义	3
第2章 文献综述	5
2.1 教学效能研究	5
2.1.1 教学效能的概念及界定	5
2.1.2 增值性评价模型	5
2.1.3 教学效果内隐理论	7
2.1.4 多层次增长模型	7
2.2 教师教学效能在线评价研究	8
2.3 在线课程研究	9
2.4 终身教授教学效能研究	10
2.5 大数据研究	11
第3章 问题提出	13
3.1 研究问题	13
3.2 研究假设	13
第4章 研究方案	15
4.1 数据获取	15
4.2 变量说明	16
4.3 数据处理与分析	18
第5章 研究一：教师教学效能的影响机制研究	19
5.1 目的	19
5.2 方法	19
5.3 结果	19
5.3.1 助理教授、副教授与教授教龄的描述统计	19
5.3.2 高分教授与低分教授教学特点及综合得分分析	20
5.3.3 助理教授、副教授与教授的教学特点分析	22
5.3.4 助理教授、副教授与教授的综合得分与教龄的关系分析	23

5.3.5 课程难度、再次选课比例与教授综合得分的关系分析	24
5.3.6 学生成绩、出勤与教授综合得分的关系分析	25
5.3.7 教授的线下课程及在线课程的综合得分及课程难度分析	27
5.3.8 不同学科的线下课程及在线课程的综合得分及课程难度分析	29
5.4 讨论	31
第 6 章 研究二：线下课程及在线课程教学效能的差异研究	34
6.1 目的	34
6.2 方法	34
6.2.1 研究对象	34
6.2.2 研究工具	34
6.3 结果	34
6.3.1 国内学生给予教师线下课程及在线课程教学效果得分的描述统计	34
6.3.2 国内教师线下课程及在线课程的教学效果得分分析	35
6.3.3 不同职称国内教师的线下课程及在线课程的教学效果得分分析	36
6.3.4 国内教师线下课程及在线课程的教学效果各维度分析	37
6.4 讨论	38
第 7 章 总讨论	40
7.1 主要研究发现	40
7.2 研究贡献及启示	40
7.2.1 理论意义	40
7.2.2 现实意义	41
7.3 研究不足	42
7.4 未来展望	43
参考文献	44
附录 A 大学教师教学效果评价问卷	52
致谢	55
声明	56
个人简历、在学期间发表的学术论文及研究成果	57

第1章 引言

1.1 选题背景

《礼记·学记》里曾提到过：“学然后知不足，教然后知困。知不足，然后能自反也；知困，然后能自强也。故曰：教学相长也”。由此可以看出，教师的教与学生的学是一种相辅相成的结合，教师通过教学反馈而自我勉励，最终获得教师教学能力的发展。

对于教师的教与学生的学都可以通过教学评价来客观衡量。一直以来，对教师的教学评价是学校教育的重要内容，通常情况下，对教师的教学评价是直接在课堂上，学生填写纸质问卷对教师进行评价，也有的是在课程将要结束后，学生在学校内部的教务系统里对教师的教学进行评价。但是传统的教学评价局限较多，第一，传统的教学评价方法相对比较费时，因为需要给学生发放问卷，然后再收集问卷，录入系统，程序较为繁琐；即使是通过教务系统进行内部评教，但因为教务系统不开放评估结果，学生无法得知教师的教学评估结果。第二，学生不能随时随地反馈教师教学表现。第三，传统的教学评价多在学期末进行评价，在整个学习过程中教师没有获得及时反馈。因此，教师想要获得来自学生及时反馈，传统评教方式存在明显不足。

近年来，随着互联网的普及，在线评教平台兴起并得到迅速发展，这些在线评教平台也受到学生、教师及高校的高度关注。与传统评教方式相比，这种在线评教具有诸多优势，比如维护及运营成本更低，学生更容易评价教师和发表个人对教师的评论。此外，学生可以随时随地评价教师，而教师可以随时查看学生对自己的评价，并改善自己的教学方法。总之，在线评教的突出优点是方便和高效。这些平台是传统评教方法的重要补充，多样化的评价平台越来越受到学生们的欢迎，这些平台上所展示的结果也对教师的教学表现起着越来越重要的作用。国外著名在线评教网站有 RateMyProfessors.com, RateMyTeachers.com 等，国内的在线评教网站有学邦网(www.xuebang.com.cn)、导师评价网(www.mysupervisor.org)等，学生是这些网站的使用者，通常在结课后学生对本门课程的教师的教学内容、教学效果等方面进行评价。以往对教师评价的研究表明，通过问卷或网络进行评价的不会显著改变评价结果(Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003; Le et al., 2018)。不过相比于传统问卷，基于在线评价，学生往往作答的可能性会低(Kim & Schwartz, 2013; Joines, Standish, & Gallagher, 2019)。因此，尽管存在一定的局限，但是通过 RateMyProfessors.com 等平台对教师的教学进行在线评教具有可行性。

RateMyProfessors.com 吸引了学生、教师和大学的关注，由于历史悠久，RateMyProfessors.com 上拥有众多教授的评价数据，其中包括助理教授及获得终身职位的副教授和教授。终身教授的职位授予来自终身教职制度。终身教职制度是美国的一项

教育制度，尽管终身教职制度历史悠久且取得了巨大的成功，但终身教职制从一开始就存在着许多分歧和争议。从积极的一面来看，它能让教授在有额外保障和收入的情况下更有效地教学，教授们有更多的精力从事教学及科研工作。但另一方面，终身教职制度也保护教授免于被解雇。因此，终身教职制度保护了没有能力满足教学要求的教师(Rothgeb, 2014)，在这种情况下，教师的教学质量会受到影响，教授在教学上没有能力满足学生学习的需要，也会影响到与学生的关系。近年来，美国的威斯康辛州、肯塔基州、阿肯色州和田纳西州等在内的几个州都制定了减少终身聘任制的法律。一般认为，教授的教龄越长，教学经历越丰富，教学水平越好，但是在终身教职制度下，教授的教学质量从长期来看是提高了还是降低了，这是一个值得研究的问题，也是本文的关注点。教授的教学特点及课程类型与学生对教授的综合得分存在相关关系(Constand, Clarke, & Morgan, 2018)，而且学生的反馈意见是一种重要的教师评估工具，教师通过获得学生的反馈，可以完善他们的教学，提高教学效能(Mart, 2017)。因此，通过学生的视角来了解教授的教学很有必要。随着在线课程的发展，对于教授的教学能力提出了新的要求，教授的在线课程及线下课程的教学表现在学生眼中存在怎样的差异也是本文关注的问题之一。

互联网的普及让我们可以轻松获取网络数据，通过爬虫技术可以获取 RMP 上海量的教授评价数据，也为开展大数据研究提供了一条可行的途径。大数据研究方法提供了对现实世界行为的洞察，我们可以获取到丰富多彩的行为数据(Chen & Wojcik, 2016)。大数据的出现为心理学的研究提供了新的研究方法。传统教学评价研究多采用的是问卷法，所能获取的样本数量有限，而大数据出现使得样本即总体成为可能，突破了传统心理学中样本—总体的二维特性(喻丰, 彭凯平 & 郑先隽, 2015)。

因此，为了探究教学效能的影响因素与教师在线下及在线课程的差异，本文将通过大数据技术探究教师教学效能与教龄等因素的关系、教学效能的影响因素；通过问卷调查法，研究教师在线下及在线课程的教学表现差异，并基于此提出改善教师教学效能的策略，同时也为提高线下课程及在线课程的教学效能提供一定的参考依据。

1.2 研究目的

首先，本文将基于 RMP 的数据进行大数据分析，探讨教师教学效能与教龄的关系，同时也将探讨教师教学效能的影响因素，探索提高教师教学效能的策略，其次基于国内调查数据，分析教师的教学效能在线下课程及在线课程的差异，为提高教师的教学效能提供一定的依据。因此本文的研究结构是：

研究一：基于 RMP 上的学生评教数据，探讨教学效能的影响因素。主要关注的问题有教授的教学效能随着教龄增加的发展变化是怎样的，获得终身职位与未获得终身职位的教授的教学效能的是否存在差异。

研究二：基于《大学教师教学效果评价问卷》，分析国内教师的在不同授课形式(线下及在线课程)下的教学效能的差异，多维度评估教学效能，探寻教师改进教学策略的方法。

1.3 研究意义

1.3.1 理论意义

增值性评价模型已经关注到教师对学生的影响，但是单从学生的成绩来评价教师的教学效能难免存在一定的局限，增值性评价模型的不足之处在于缺少从学生评价角度看待教学效能的内容，而学生对于教师教学效能的反馈起着重要作用，学生评教作为教师教学反馈的重要一环能为教师提供反馈，帮助教师提高教学能力(Marsh & Roche, 1993)。而教学效果内隐理论也指出，学生对教师的反馈是对教学效能的整体反馈，因此，通过学生评价来衡量教师教学很有必要，而且增值性评价模型具有良好的扩展性，可以将学生评价数据加入模型内，本文的研究将可以丰富增值性评价模型的评估维度，为该模型解释教师对学生的影响提供一定的理论意义。

本研究也将扩展多层次增长模型在大数据方面应用。多层次增长模型已经加入学生评教的内容，但是传统上采用的是小样本研究，本文为多层次增长提供了大样本的数据支持，可以弥补了该模型研究中的小样本的不足。

1.3.2 现实意义

RMP 作为一个公开在线评教平台，数据丰富且应用广泛。本研究具有较大的应用价值：

第一，本文为评估教师教学效能提供一定的参考依据。本文基于 RMP，利用大数据技术，探讨了教师教学年龄与教学效能之间的关系，有利于帮助教师了解自身的教学效能发展情况。例如，如果教师教学效能随着年龄增长保持不变，那么教师应通过其他教学实践不断改善教学策略，促进专业发展，提高教学效能，以帮助学生学习的。

第二，本文为终身制度的实施标准提供一定的参考依据。本文基于 RMP，探讨教师教学效能及终身职位之间的关系，了解获得终身职位与未获得终身职位的教授之间教学效能的差异，为高校制定终身制度的标准提供一定依据。例如，高校可以将学生评价纳入评价指标，在学校层面建立共享评价平台，教师可以及时了解学生对教学的反馈，从而督促教师提高教学效能。

第三，本文为提高教师教学效能提供一定的指导。本文基于 RMP，探讨教师教学效能的影响因素，为教师提高教学效能提供一定的策略指导。

第四，本文为学生选择在线课程及线下课程提供一定的依据。例如，学生会考虑在线及线下课程的互动模式及课程难度，根据需要选择对应的课程。

第2章 文献综述

2.1 教学效能研究

2.1.1 教学效能的概念及界定

教学一直是高校教师的核心工作，而针对教师的教学效果，一般使用教学效能作为评价内容。教师的教学效能对学生的学习有重要作用，也是影响学校教学质量的关键因素。关于教学效能，不同时期、不同的研究者有不同的界定，其内涵目前呈现多元化的解释。

20 世纪 80 年代左右，有研究者从教学能力角度出发，指出教师的教学效能，是教师在指导学生后，学生在标准化测验的结果(Good, 1979)。教学效能由充分备课、礼貌对待学生、清晰授课、课后与学生交流等 12 个方面组成(Tang, 1997)。还有研究者认为，教学效能是一种多元评价，通过大学生评价教师教学效果问卷 SEEQ(Students' Evaluations of Educational Quality)可以测量教学效能，包括学习价值感，教学热情与组织，功课量等 9 个维度(Marsh & Hocevar, 1991)。

近年来，关于教学效能的认识趋于一致。对教师的教学效能评价是一种形成性评估，目的是为了帮助教师改进教学(Jahangiri, Mucciolo, Choi, & Spielman, 2008)。教学效能是教师的一项综合教学能力，是指导、支持学生达到学业目的的程度，如获得成就、提高满意度(Gorsky & Blau, 2009)。教师的教学效能就是学生的学习结果(Marsh & Roche, 2000)。教学效能是教师对学生结果(课堂行为、教学结果)的影响，可以通过课堂观察、问卷调查和定性方法实现对教师教学效能的测量(Muijs, 2006)。

总之，从以上有关教学效能的概念及组成成分上看，教学效能并没有完全一致的概念，主要原因是各教育机构出于不同的目的，导致对教师教学效能的理解不一致(Jahangiri et al., 2008)，但是近年来的研究对教学效能有了更为清晰的界定：教学效能是教师的一项教学能力，是帮助学生达成学习目的的程度。教学效能在教师日常的教学工作中通过多个方面展现出来，对学生的学习及行为产生一定的影响，并且这种教学效能可以通过问卷、课堂观察、课后评测等手段进行测量。

2.1.2 增值性评价模型

增值性评价模型是目前应用相对广泛的评估教师教学效能的模型。增值这一词，最早可追溯到经济学中增值(value-added)的概念，其基本含义是指投入产出的差值(边玉芳 & 孙丽萍, 2015)。增值性评价模型一般以学生成绩作为评价指标(Kupermintz, 2003)，该模型是一种形成性评价，能够以动态、更加灵活的方式来考察教师的教学效能，其评价主要关注的是学生的学习结果(Hanushek & Rivkin, 2010)。

增值性评价模型是通过获取教师所教学生在一段时间内2次或者是多次成绩,运用一定的分析方法来分析教师所影响到学生学业成绩的变化,即教师对学生成绩的增值(Wright, Horn, & Sanders, 1997)。经典的增值性评价模型主要有获得分数模型、协变量校正模型等,这些模型目前已被应用到实际的教学评价中,如田纳西州建立了基于田纳西模型的教育增值性评价系统(Sanders & Horn, 1994)。但是经典模型不断受到质疑,寻找更适合的模型成为研究者的关注的重点。目前已经有一些增值性评价模型的变式(Sass, Horn, & Sanders, 2014),这些变式从单维增加到多维,从一年效能增加到几年效能(Mariano, McCaffrey, & Lockwood, 2010),而且增值性评价模型具有良好的扩展性,目前已经加入了学生的历史学习成绩、学校因素、同事因素、家庭等因素(Hanushek & Rivkin, 2010)。

增值性评价模型具有一定的优势。基于当前的增值性评价,可以区分出教师对学生成绩的影响和其他因素对学生学习的影响,因此相比于传统评价而言,有利于促进教师评价的公平性、引导教师关注所有学生,加强教师专业发展(边玉芳 & 孙丽萍, 2015)。基于增值性评价模型制定相关的政策,可以提高教师质量(Winters & Cowen, 2013)。

尽管增值模型比较好地评估教学效能,可以区分出教师的单独贡献,但是教师身处于学校环境下,实际上难以区分出学校对学生的影响,因此指标是否公平且可信尚存在争议。也有调查显示,有三分之二的教师是无法进行增值性评价(Papay, 2012),所以大部分教师不能使用该模型进行教学效能评价,这为继续推广该模型带来一定的困难。

基于以上发现,不难看出增值性评价模型在一定程度上较好地评价教学效能。如果无法获得学生的成绩或者两次考试测验不等值,教师在实际教学中难以进行增值性评价,这对于应用增值性评价模型有一定的困难,而且严格意义上,增值性评价也难以区分出教师单独对学生的贡献,因此,增值性评价模型存在明显不足。

增值性评价模型是从学生成绩的角度来评价教师教学效能,但是就目前而言,鲜有使用学生对教师的评价作为教学效能的指标。教学效能不仅仅从教师课堂表现,学生测验结果上评价,更应该从学生角度的看待教学效能。以往的研究发现,学生的评价和教师的评价在9个维度上存在显著正相关($r=0.45$),学生可以有效评价教师的本科课程及研究生课程(Marsh, 1982),学生评价教师是评估教学效能的有效手段,学生的评价可以为教师提供反馈,进而帮助教师提高教师教学质量(Marsh & Roche, 1993)。学生的课堂参与也能够预测教学效能(Burns & Ludlow, 2005),学生对教学效果的评估是一种形成性评估,这种评估可以帮助教师提高其技能(Hobson & Talbot, 2001)。但是,需要注意的是,学生的反馈对教学评估也存在一定偏见(Chitre & Srinivasan, 2018)。近年来,不断有研究者证明学生评价教师的意义,作为一种形成性评估,可以帮助教师提高教学水平,实现课程的交付,而且最终帮助大学等教育机构有效评估教师(Chen & Hoshower, 2003)。学

生对教师的评价和学生对课程评价显著正相关,教师在教学上的改进能够吸引学生参与教学评估(Yueh, Chen, Chiu, Lee, & Wang, 2012)。

因此,学生评价教师可以反应教师的教学效能,而增值性评价模型内鲜有该内容,对于无法进行增值性评价的教师,学生的反馈显得尤为重要,这对帮助教师获得教学反馈,不断改进教学方法及技能,达到提高教学效能的目的有重要作用。

2.1.3 教学效果内隐理论

Whitely 等人在 1976 年提出了教师行为的内隐理论,内隐理论是指学生对教师行为的整体评价,可以反应出学生对教师行为各个维度的评价,即当教师的行为 A 与教师的行为 B 存在关联时,学生对行为 A 评分较高,那么自然而然对行为 B 也评价较高,即使行为 B 并没有出现。内隐理论认为教师的行为对学生做出反馈有影响,如果在同一门课学生做出相同的反馈那么就出现了内隐理论(Marsh, 1984)。教学效果内隐理论认为,学生对于教师的教学行为存在固有信念,而且学生认为教师的教学行为是整体性、系统性的,因此,学生对教师的教学效能的整体评价是有效的(Harrison, Ryan, & Moore, 1996)。通过对学生评分的分析发现,学生的内隐理论是建立在以往教师教学经验的基础上,因而学生的内隐理论对教师教学效能的评价有显著影响(Whitely & Doyle, 1976; Larson, 1979)。不管学生是否随机分配到不同的班级,学生的反馈都会受到内隐理论的影响。由此可知,内隐理论是普遍存在,对个别学生具有特殊性的内隐理论以及随机带来的变化,也会在班级层面上相互抵消(Larson, 1979)。教师的课程准备和组织,清晰易懂、课堂水平和对学生的关注会影响到学生对教师的总体评价(Feldman, 1988)。因此,教学效果内隐理论普遍影响到学生对教师教学效能的反馈,学生对于教师的整体评价可以反应教师的教学效能,教师应当重视学生对教学效能的评价。

2.1.4 多层次增长模型

Marsh 在 2007 年提出了教学效能的多层次增长模型,使用学生评价的数据来探讨了教学效能随时间的是增加、减少或保持稳定这一问题。该模型对 195 名教师的 6024 门课程做了多层次建模,发现教师的教学效能随着时间的增长并没有发生变化。教师的教学效能存在本科和研究生水平的课程上保持稳定,尽管教师间的教学效能存在显著的个体差异,但随着时间的推移,教学效能保持稳定(Marsh, 2007)。即使教师的数据在数据集中只出现一次,多层次增长模型依然能分析教师的教学效能(Singer & Willett, 2003)。

有研究者继续探讨教师的性别、种族、时间因素、是否获得终身、教学模式(线上、线下)、课程水平(本科、研究生课程)、学科与教师教学效能的关系,Carle 在 2009 年使用多层次增长模型发现,教授的教学效能评分比较稳定,在研究开始时获得很高评价的教授,在研究结束时得到了相似的评价。获得终身的教授,在五年的时间里,他们的平均教学效能得分提高了 0.12 分,而对于在线课程及线下课程的分析发现,少数族裔的教

师在线下课程的教学效能得分比线下课程的低,但是总体上线下课程及在线课程的教学效能是不存在差异的(Carle, 2009)。

2.2 教师教学效能在线评价研究

对教师进行教学评价的目的是为了衡量教师的教学效能,传统上,对于教学评价从作用上可分为诊断性测评、形成性测评及总结性测评,这三类评价适用于不同的阶段,诊断性测评适用于教学前,形成性适用于教学中,总结性适用于教学后,其中总结性评价对以后的学习情况有预测作用(陈琦 & 刘儒德, 2011),其中形成性和总结性评价可以应用到对教师的教学效能的评价。学生对教师的评价则一般是在学期末结束后,学生依据学校的问卷进行填写,进而反馈教师这一学期的教学情况,教师根据学生的反馈建议改善教学方式,达到提高教学效能的目的。由此可以看出,学生评教是评估教师教学有效性和教育质量的重要方式。由于传统教学评价方式的数据难以公开,因此在线公开的评教平台获得极大关注,其中最具代表性的便是 RateMyProfessors.com。

RateMyProfessors.com(RMP)成立于 1999 年,是最具影响力的在线评教网站之一,是专为本科生和研究生提供评价教授和选课指导的平台。目前该网站已经有 1900 多万个评论,170 万名教授和 7500 多所学校的信息,作为最受欢迎的在线评教网站,RMP 每月的访问量超过 400 万次(RateMyProfessors.com[RMP], 2020)。由于本网站历史悠久且教师数据众多,因此可以作为公开的数据收集平台。以往的研究发现,在 216 名学生中,有 80% 的学生访问过 RMP,30% 的学生给教授打分并发表评论(Davison, Elizabeth, & Price, 2009)。另一项调查研究表明,197 名学生中有 84% 的学生访问 RMP 来选择他们想要的课程,194 名学生中 23% 的学生对教授的教学表现做出评价(Bleske-rechek & Michels, 2010)。虽然该网站数据量大、使用频繁,但 RMP 作为一种评教手段的有效性一直受到研究者的质疑。研究表明 RMP 的问卷存在偏见,该网站的评价问卷是一种态度量表,而不是教学绩效量表(Clayson, 2014; Legg & Wilson, 2012)。教授的整体教学水平与教授的吸引力呈显著正相关($r = 0.168$)(Bonds-Raacke & Raacke, 2007)。这表明,长相好看或漂亮的教授比长相不那么漂亮的教授将获得更高的评价(Theyson, 2015)。其他研究人员也质疑 RMP 未能真正反映了教师的教学效果(Felton, Koper, Mitchell, & Stinson, 2008)。但也有研究发现,RMP 上的综合得分能够预测教师教学质量。例如,RMP 上的综合得分与传统学生教学评价的结果之间存在显著的正相关($r = 0.68$)(Coladarci & Kornfield, 2007; Otto, Sanford, & Ross, 2008)。教授的综合得分与学生的 GPA 呈显著正相关($r = 0.41$)(Sonntag, Bassett, & Snyder, 2009)。另一些调查研究显示,58% 的学生在 RMP 会诚实地评价教授,RMP 评教能够预测是传统评教的结果 ($\beta = .264, p < .05$) (Brown, Baillie, & Fraser, 2009)。除了学生在选择课程和评价大学及教授时经常访问 RMP,教授

和大学机构也对在线评教网站感兴趣。教授们感兴趣地是如何提高学生对他们的评分,比如,通过提供难度较小的课程或更好的成绩来实现。另一方面,学生们关注的是如何找到好的教授和课程。而当下,在线教育的发展为教授的教学提供了更广阔的平台,也为学生选课提供了更多的机会。

2.3 在线课程研究

在线课程平台的迅猛发展,对于教师的教学提出了新的要求。目前,多数大学或机构也开始开设自己的在线课程平台,例如国外著名在线课程网站 Coursera 是由美国斯坦福大学两名计算机科学教授创办,旨在与世界顶尖大学开展合作推广在线课程,国内的在线课程网站有中国大学慕课、学堂在线等,这些平台的出现为教师提供了丰富多彩的授课渠道、为学生提供了丰富多彩的课程选择机会。

在线教育需求的增长,使得教师需要改进他们的方法,以确保学生理解授课内容,在线下课程方面,团体互动,以前的软件经验、教师指导和模拟工具易用性会影响学生的学习,而在线上课程,只有教师指导和工具易用性会影响到学生的学习(Riley, Ellegood, Solomon, & Baker, 2017),说明教师在两种不同课程的教学上存在一定差异。少数族裔的教师在线上课程得到更低的教学效能得分,但是在线下课程却不存在显著差异,总体上教师的教学效能在线和线下不存在显著差异(Carle, 2009),学生在线下课程及线下课程的学习结果是一样的(Lou, Bernard, & Abrami, 2006; Means, Toyama, Murphy, Bakia, & Jones, 2009),这意味着对于学生选择在线或者下线课程来说都是一样的,学生给予教师的教学表现的评价可能也是一样的。但是由于退课的学生不参加期末考试,因此在线课程线下课程两种课程模式下学生的学习结果受到一定影响(O'Neill & Sai, 2014),进而会影响学生对教师的评价。在线教育在空间上将教师与学生分离,对于如何保证在线教育的教学质量,人们进行了大量的讨论,基本共识是在线教育对课程质量保证与传统教育的方式基本相同(Stella & Gnanam, 2004)。而在线及线下课程的教学及课程设置上是否存在不同,对于学生选课来说至关重要,因为当学生知道在线课程及线下课程是在学分相同、学费相同且在同一学期上课时,学生倾向于选择线下课程,因为他们相信线下课程学地会更好(O'Neill & Sai, 2014)。残疾学生和非残疾学生的在线下及在线课程上的成绩基本一样,因此他们对于选择线下及在线课程是没有差异的(Richardson & John, 2016)。大量的在线课程和线下课程的实证比较发现,在线学生的学习表现和线下学生一样好或更好(Tucker 2001)。当在线课程在教学设计上采用更合理的方式,这种方式可以提供同样有效的学习环境(Driscoll, Jicha, Hunt, Tichavsky, & Thompson, 2012)。

RMP 另一个重要的作用是为学生提供选课指导, RMP 可以区分出线上课程与线下课程。对于学生的选课,学生考虑的因素有很多,有研究指出,学生在选择课程时会考

虑授课教师的性格特点，教师的外向性能够预测学生对教师的评价($\beta = 0.76$, $p < .001$)(Radmacher & Martin, 2001)，学生还会考虑到课程本身的特点，例如课程的难度，学生一般倾向于选择较低水平的简单课程(Babad, Darley, & Kaplowitz, 1999)，但如果课程评价较好或者课程很有价值时，学生也会选择难度高的课程(Coleman & McKeachie, 1981)。学生认为RMP上的内容是可信且有效的，因此选课依赖与本网站多于询问朋友、其他同学(Hayes & Prus, 2014)。而且学生倾向于选择在RMP上得到积极评价的教授，因为学生在课程结束后容易获得高分(Reber, Ridge, & Downs, 2017)。通过教学评价，可以改善教师和学生的教学活动(Umar et al., 2016)，针对于学生的课程选择，以往学生对教师的评价会影响到学生的选择。但多数研究都是测量学生在两种教学模式下的学习成绩，鲜有研究直接涉及在线和线下的课程的教学效能的评价(Carle, 2009)，这对学生选课及教师改进教学策略减少了重要的反馈途径。毕竟，学生评价教师是出于期望教师不断改进教学的目的(Chen & Hoshower, 2003)。

2.4 终身教授教学效能研究

RMP已经积累了170万的教授数据，其中包含众多助理教授、副教授、教授的评价数据。大学的终身制和晋升制度可能会影响教授的教学表现。在1900年，美国实行终身制，以促进学术自由和更好的教学表现。受美国终身制的影响，中国大学在2000年左右开始采用终身制来激励教授(Gonzalez, Liu, & Shu., 2012)。终身职位是一种无限期的学术职位，除非教授存在不称职的行为，例如无法履行工作职能，有能力执行工作但选择不做，以及存在欺诈和犯罪活动的不道德的行为，否则就不会终止终身任期。授予终身需要考虑研究教授的已发表的论文、教学绩效和教学服务方面的表现(Boyer, 1996)。博士研究生在毕业后可以担任助理教授，然后可以晋升为副教授并获得终身教职。从助理教授到副教授平均需要6年的时间。然而，如果助理教授在7年内未能获得终身职位，那么其获得终身职位的机会将随着时间的增长而急剧下降(Lutter & Schröder, 2016)，想要获得终身正教授的职位可能需要14到16年，而副教授平均需要6年才能成为教授(Cornell University, 2020; University of Houston, 2020)。

尽管终身教职制度有着悠久的历史，但该制度从一开始就存在着许多分歧和争议。不利的一面是，终身制度很可能保护那些不满足教学要求的不称职的教师(Rothgeb, 2014)。但从积极的一面来看，终身制还可能使教授在获得额外保障和收入的情况下更有效地进行教学。关于终身制度的观点存在分歧，结果导致威斯康星州，肯塔基州，阿肯色州和田纳西州等多个州制定了削弱终身制度的法律。根据美国教育委员会的公告，现在有11个州将教学绩效作为做出裁员决定的主要考虑因素，而不是终身的决定因素，并且有16个州将教授的绩效评级作为授予终身的依据(Thomsen, 2014)。终身制对教授

的教学表现没有显著影响(Cheng, 2015)。与非终身教授相比,终身教授具有同等的教学效果(Seiler et al., 1999)。教授为大学提供了有效的教学服务(Gentry & Stokes, 2015),由此看出,获得终身职位的教师的教学效能可能会降低或者没有变化,这对于继续推行终身制度带了一定的影响。教授的教学表现是一项重要的指标,而教师的教学表现欠佳可能会让教授失去终身职位。因此,对于终身制度是否影响教授的教学效能这一问题,依然存在研究的不足,本文将基于 RMP 探讨终身制度对教授教学效能的影响。

2.5 大数据研究

传统上,对于教师的教学效能评估是在小样本的问卷调查及个案研究的基础上进行的,这种研究存在样本量小的局限性(Kasten, 1984)。而 RMP 已经积累了大量教授数据,这为我们进行大数据研究提供了新的途径。大数据的出现让我们在看待研究的方式上产生了重大转变,大数据研究方法的特点是更具代表性,可实现对总体的直接研究;生态性,收集更加真实的行为数据;高效性,实现对大量数据的快速处理与分析;瞬时性,对人类心理及行为采集的速度更快(朱廷劭, 2016)。因此,大数据方法对现实世界行为的洞察并捕捉到各种各样的行为,可以克服传统方法的不足(Chen & Wojcik, 2016)。大数据研究呈现出海量、速度、多样性三大特征:海量意味着更大的数据容量;速度是指产生数据的速度;多样性是指各种数据类型,如网页、音频、视频(Laney, 2001)。大数据研究涵盖数据获取、数据管理、数据分析和数据分析,常见的大数据分析技术有监督学习,非监督学习,文本分析,多媒体分析。对于文本数据来说,需要删除数据采集过程中产生的无关元素,例如 HTML 网页中的标签(Chen & Wojcik, 2016)。常见的大数据方法有监督学习和无监督学习,常见的监督学习算法有分类和回归,大数据里的回归分析与非大数据中的回归分析基本一致。常见的非监督学习算法有聚类分析等,聚类分析利用样本之间的相似性和差异来发现数据中隐藏的模式和结构。大数据优势是,它使我们能够了解人们的行为和偏好。更重要的是,大样本是多维的,更具代表性。但是同样的,大数据也带来一定的局限,例如大数据引起了新的伦理问题,通过广告可以帮助消费者有更好的选择,它也利用一个人性格中的弱点(Matz & Netzer, 2017)。

综上所述,教学效能是教师的一项教学能力,是帮助学生达成学习目的的程度。以往对于教师教学效能的评价多使用学生的学业成绩,其中增值性评价模型最具有代表性,但是该模型存在的明显不足,近三分之二的教师无法进行增值性评价使其在应用上存在很大的局限,针对于评价教学效能,目前广泛使用的是学生评教数据,而且根据教学效果内隐理论可知,学生对于教师的整体评价反应出教师的教学效能,因此用学生评教作为教师教学效能具有可行性。多层次增长模型便是使用学生评教数据作为研究内容,在推广性上明显要优于增值性评价模型。但是,多层次增长模型使用的依然小样本量的学

生评价数据，这与传统上用问卷调查或者个案研究来评价教学效能在方法上都存在样本量小的局限，而当前 RMP 在线评教的广泛应用使得利用大数据的方式探究教师教学效能的影响因素成为可能。因此，本文将基于教学效果内隐理论，使用 RMP 的公开评教数据，利用大数据技术来探讨教师教学效能的影响因素，丰富发展多层次增长模型，为其提供大样本的支持，也为增值性评价模型增加新的研究内容提供一定的启发。

第3章 问题提出

3.1 研究问题

根据教学效果内隐理论可知,教师的教学行为会影响学生对教师教学效能的整体反馈,因此,学生对于教师的整体评价就是教师的教学效能。以往对于教授的教学效能的研究发现,教授的教学效能与教学年龄、终身制度存在一定的关系,但是对于这些关系,前人的研究并没有给出统一的答案,毕竟传统上对于教学效能的评价多采用小样本的问卷调查或者个案研究。尽管有研究者使用了多达 6024 门课程来分析学生对教授教学效能评价(Carle, 2009),但不同的研究者选择的样本及课程不同导致结果的不一致。针对于此问题,本文将通过大数据来试图解决这些问题,具体问题如下:

问题一:学生评价是目前教师教学效能常用的方式,学生对于教授总体评价就是教授的教学效能(Harrison, Ryan, & Moore, 1996)。所以,在学生眼中,教授的教学效能与教龄的关系是怎么样,教师的教学效能随着教龄的增长是增加还是降低了?

问题二:RMP 已上积累 170 多万教授的评价数据,其中有获得终身职位的教授,也有未获得的终身职位的教授。这两类不同职称的教授在教学效能评价是否存在差异,获得终身职位的教授的教学效能是否好于未获得终身职位的教授?

在线课程的兴起为教师的教学提出了新的要求、也为高校等机构提供了新的授课方式。有研究者认为,为了减少开支,管理者可以考虑用在线课程取代线下课程(O'Neill & Sai, 2014),但在线课程是否要好于传统的线下课程吗?目前,鲜有研究直接涉及在线和线下的课程的教学效能的评价(Carle, 2009),国内也鲜有研究教师在线下课程及在线课程的教学表现差异,因此具体问题如下:

问题三:国内教师在线课程与线下课程的教学效能的总体评价是否存在差异,在线课程与线下课程哪个教学效能更高?

3.2 研究假设

针对问题一,有研究指出,大学教师的教学效能与年龄成呈现负相关($r = -0.42$)(Robert & Harry, 1996),这意味着随着教师年龄增加,教学效能越低。教师的年龄和经验都会影响到教师的教学效能,31-40 岁的教师、41-50 岁的教师、51 岁以上的教师之间两两比较后教学效能存在差异,随着年龄的增加,教师的教学效能越高(Ismail, Arshad, & Abas, 2018),Zafer 和 Aslihan(2012)也发现,41 岁及以上的教师的教学效能及课堂管理能力比年轻教师要高,说明教学效能与教师的年龄存在一定的关系,因此,假设一是:教师的教学效能与教龄是正相关,随着教学年龄的增长,教学效能不会下降反而是上升的。

针对问题二,近来的研究表明,对于非终身职位的教授来说,获得终身职位的教授教学效能的评分得到了提升,而且两者教学效能不存在差异(Carle, 2009),但是利用增值性评价模型发现,非终身教授和终身教授在各个学科中增值得分存在差异,终身教授的得分显著好于非终身教授(Figlio, Schapiro, & Soter, 2015),因此,假设二是:终身制度会影响教授的教学效能,助理教授、副教授、教授之间的教学效能存在显著差异,教授的教学效能好于副教授、副教授的教学效能好于助理教授。

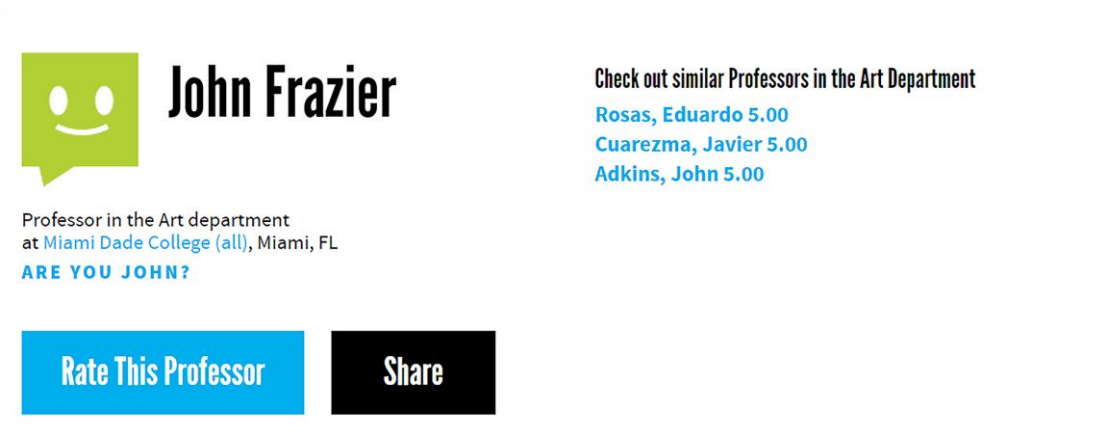
针对问题三,有研究指出,教师的教学行为在不同类型(线上、线下)上存在多个维度的差异,例如在线课程在互动模式等方面会影响到学生对知识的理解(Riley, Ellegood, Solomon, & Baker, 2017),而且在线课程上,学生的后测成绩和最终期末成绩要比线下课程高(Tucker 2001),再者,教师的行为和成绩会影响到学生对教师的评价,因此假设三是:教师的在线课程及线下课程的教学效能总体评价存在显著差异,在线课程的教学效能要高于线下课程。

第 4 章 研究方案

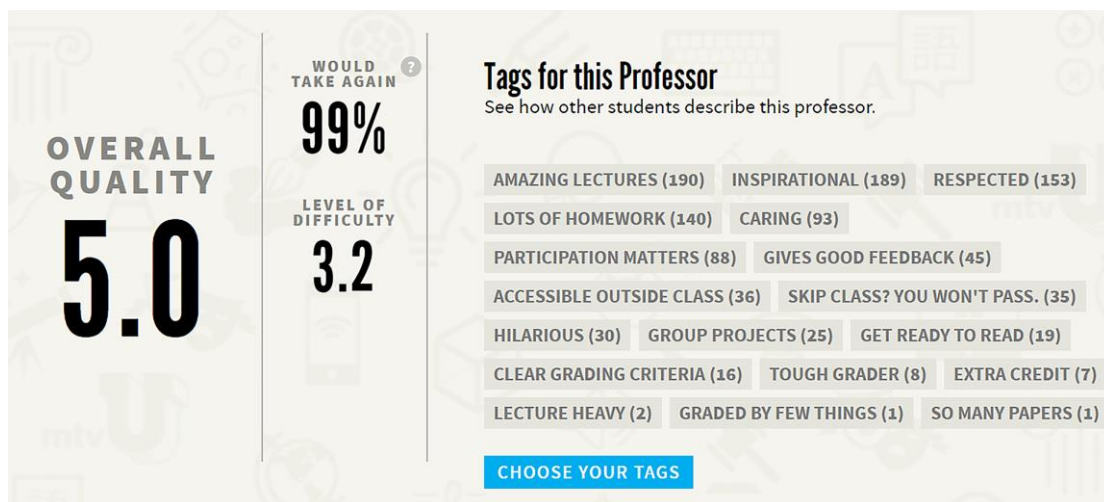
4.1 数据获取

使用 Python 网络爬虫技术抓取 RMP 原始网页，共计收集了 180 万个网页。因为原始网页是 HTML 格式，因此需要使用 Python 对原始数据进行清理。在 RMP 上，教授可以申请删除他们的个人页面；如果学生的评论违反网站指南或者教授不再在该大学任教，RMP 也会删除该教授的个人页面；而且 RMP 曾在 2015 年更新，因此造成 RMP 存在较多空白页。经过数据清理之后，RMP 数据集共包含 9,543,998 行数据和 919,750 位教授的数据。下图是 RMP 教授评价页面的示例图：

A



B



C

RATING: ALL **CLASS: ALL** **COMMENT**

09/16/2019

AWESOME

5.0 OVERALL QUALITY

2.0 LEVEL OF DIFFICULTY

IDH1001

For Credit: Yes
Attendance: Mandatory

Textbook Used: No
Would Take Again: Yes
Grade Received:
Audit/No Grade

CARING ACCESSIBLE OUTSIDE CLASS INSPIRATIONAL

This is my 1st semester in college & I have already spent more hours talking/planning my future w/ Professor Frazier than all of my former teachers. This weekend he spent nearly 3 hours just helping me create a proposal for the Clinton Global Initiatives Challenge. He is amazing and brilliant and caring. He rocks if you want to get involved

0 people found this useful 0 people did not find this useful

report this rating

图 4.1 RMP 教授评价页面示例图

注：RMP 网页包含三个部分：A 教授的个人信息，包括姓名、学校、部门和大学所在的州；B 教授的整体评价，包括综合得分，是否再次选课的比例、课程难度和学生给予的标签；C 每个学生对教授教学的评价信息，包括综合得分、课程难度、课程名称、是否强制出勤、是否为了学分、成绩、文本评价。

4.2 变量说明

从RMP中提取的数据集有20个变量。表4.1是RMP的样例数据，以下是数据简表及每个变量的简要说明：

表 4.1 Ratemyprofessors.com 样本数据集示例

A

教授姓名	学校	院系	所在地	州	教龄
Leslie Looney	University of Illinois at Urbana-Champaign	Astronomy department	Champaign Urbana	IL	11
Jans Wager	Utah Valley University	English department	Orem	UT	11

注：(1)教授姓名：被评教的教授姓名；(2)学校：教授任教的大学；(3)院系：所属院系；(4)所在地：大学所在地；(5)州：大学所在州；(6)教龄：教授的教学年龄，从第一次被学生评估的时间到2019年的时间差。

B

综合得分 1	再次 选课	课程 难度	标签	发布日期	线下 课程	在线 课程
4.7	Null	2	Hilarious(2) Gives good feedback(1) GROUP PROJECTS (2)	06/27/2017	NUL L	ASTR 122
4.5	83%	3	Gives good feedback (6) Participation matters (3) Respected (3) Caring (2) Hilarious (3)	12/12/2017	NUL L	ENG L101 0

注：(7)综合得分 1：教授的综合得分，3.5-5 分为好，2.5-3.4 分为一般，1-2.4 分为差，此得分是所有学生给予教授综合得分的平均分；(8)再次选课：希望再次选择这门课程的学生百分比；(9)课程难度：这门课的难度等级，1 分是最容易，5 分是最难，该得分是所有学生给教授课程难度评分的平均分；(10)标签：学生选择描述教授的标签；(11)发布日期：学生对教授评估的日期；(12)线下课程：线下课程的名字，包含课程缩写及课程号；(13)在线课程：在线课程的名字，包含课程缩写及课程号。

C

综合得分 2	课程 难度 2	是否 强制 出勤	为了 学分	再次选 课	成绩	评价
5	3	Manda tory	Yes	Yes	B	This class is hard, but it's a two-in-one gen-ed knockout, and the content is very stimulating.
5	2	Manda tory	Yes	Yes	A	I LOVE Prof. Wager. She's hilarious and dedicated. I took this as the hybrid block and I loved it.

注：(14)综合得分2：每个学生给教授的综合评分，3.5-5分为好，2.5-3.4分为一般，1-2.4分为差。不同于变量7，此得分为每个学生给予老师的得分；(15)课程难度2：每位学生评价教授课程难度的得分，1分是最容易，5分是最难，不同于变量9，此得分为每个学生给予教授课程难度的评分；(16)是否强制出勤：课程是否为强制参加(是或否)；(17)为了学分：学生选择本课程是否是为了学分(是或否)；(18)再次选课：学生是否愿意再次选择本课程(是或否)；(19)成绩：学生

在本门课获得的成绩, 如A+、A、A-、B+、B、B-、C+、C、C-、D+、D-、F(不及格)、WD(退课, withdraw)、INC(未完成, incomplete), Not(未报告), Audit/No(旁听生); (20)评价: 学生对教授的文本评价。

4.3 数据处理与分析

在研究一中, 首先调查了伯克利加州大学教授的教龄, 以验证以往对于助理教授、副教授、教授教龄的判断, 其次计算高分组教授(综合得分3.5-5.0分)及低分组教授(综合得分1.0-2.4分)的标签的数量, 并进行卡方检验以判断标签之间是否存在差异, 同时对标签进行词云可视化, 词云是一种在文本分析中展示词频的工具, 每个词的大小随其重要性和频率而大小不一。再者, 计算不同职位教授标签并进行卡方检验, 并可视化教授的教学质量综合得分随教龄的变化趋势。同时计算教授是否选课比例、课程难度及综合得分的关系, 获得不同期末成绩的学生给予教授综合得分的差异, 比较为了学分而选课的学生与不是为了学分而选课的学生在给教授评分上差异, 比较强制出勤与非强制出勤的学生在给教授评分上差异, 最后通过分析学生的文本, 对比不同学生对教授的评价是否存在差异。

在研究二中, 通过《大学教师教学效果评价问卷》调研国内学生对教师的教学评价, 对比分析在线课程及线下课程中的教师的教学表现, 其次基于RMP区分不同难度的课程及不同学科, 对比以上课程及学科在线课程及线下课程的教学差异。

由于大数据研究样本量大, p 值很容易有统计学意义, 因此针对于大数据的统计结果, 更多关注效应量, 而非 p 值。Chen和Wojcik (2016) 建议报告效应量。*Cohen's d*和*Cohen's w*是心理学研究中的常用的效应量, 其效果的大小可以解释为小、中、大, 相应的*Cohen's d*值分别为0.2、0.5和0.8。*Cohen's w*值分别为0.1、0.3和0.5(Cohen, 1988, 1992)。

第5章 研究一：教师教学效能的影响机制研究

5.1 目的

本研究基于 RMP 的进行大数据分析,探讨国外教授的教学特点、教学效能与教龄的关系、教授的教学效能的影响因素,为教授提高教学效能提供一定的参考依据,同时探讨教授的教学效能不同课程类型(线下课程、线上课程)上差异,帮助教授改进授课方式,也为学生选课提供一定指导。

5.2 方法

在本研究中,首先基于教授姓名、学校、部门和州对教授进行去重,然后根据综合得分将教授分为高分组(综合得分3.5-5.0分)和低分组(综合得分1.0-2.4分),分别计算高分教授和低分教授的标签频率并进行卡方检验,卡方检验的结果可以反映出高分教授及低分教授的教学特点和学生对两类教授的不同评价。本研究根据教龄来区分助理教授、副教授、教授,首先计算三类教授的标签频率,其次通过卡方检验来探讨不同类型教授的教学特点,分析助理教授、副教授、教授三类教授的教学效能综合得分。最后对教授的课程难度、再次选课比例与教授小学效能综合得分作相关及回归分析,同时针对得到不同期末学习成绩的学生,分析他们给予教师的综合得分是否存在显著差异。

由教学效果内隐理论可知,学生对于教师的整体评价可以反应教师的教学效能,因此本研究将学生对于教授的综合得分作为教师的教学效能。

5.3 结果

5.3.1 助理教授、副教授与教授教龄的描述统计

本研究首先调查了伯克利加州大学10所学院(历史系、社会学系、生物工程系、土木与环境系、古典文学系等)的443名在职教授,其中有115位教授的简历有效。男性教授77名,女性教授38名。本研究中的教龄是指从该教授博士毕业后,从事博士后研究工作或者开始当讲师时到被评为助理教授的时间,副教授教龄及教授教龄依此类推。从表5.1中可知,助理教授($M = 2.46$ 年, $SD = 2.33$ 年)到副教授($M = 7.75$ 年, $SD = 2.77$ 年)需要5.29年,而从副教授($M = 7.75$ 年, $SD = 2.77$ 年)到教授($M = 13.33$ 年, $SD = 4.77$ 年)需要5.58年。根据文献综述可知,助理教授一般需要6年才获得副教授职位,副教授需要6年时间被提升为教授,而教授的教学经验则是超过12年(Lutter & Schröder, 2016; Cornell University, 2020; University of Houston, 2020),由此可见,本调研的结果符合教授教龄的判断。

表5.1 助理教授、副教授、教授教龄的描述统计

教师职称	N	助理教授教龄		副教授教龄		教授教龄	
		M	SD	M	SD	M	SD
助理教授	20	3.35	2.28	--	--	--	--
副教授	16	2.80	1.93	9.07	2.67	--	--
教授	79	2.15	2.37	7.49	2.74	13.33	4.47
总计	115	2.46	2.33	7.75	2.77	13.33	4.47

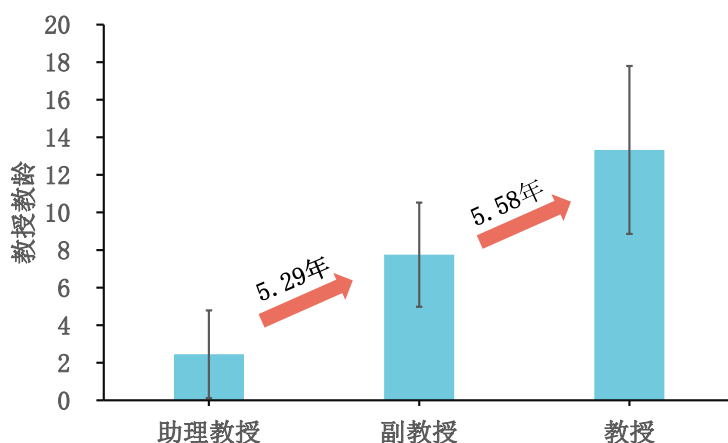


图 5.1 助理教授、副教授、教授的教龄

5.3.2 高分教授与低分教授教学特点及综合得分分析

在RMP中,教授的综合得分是1.0-5.0,其中1.0-2.4是低分教授,而3.5-5.0是高分教授。非参数检验的结果显示,高分教授($N = 610152$, $M = 4.35$, $SD = 0.48$)和低分教授($N = 110092$, $M = 1.79$, $SD = 0.47$)的综合得分具有显著差异(Mann-Whitney $U = 67172853984$, $p < .001$, Cohen's $d = 5.38$)。高分教授和低分教授的标签频率见表5.2,表中列出的20个标签中均存在显著差异。高分的教授通常得到“给予良好反馈(Gives good feedback)”(9.54%),“关怀(Caring)”(8.57%)和“受人尊敬(Respected)”(8.56%)等标签,而出现频率较低的标签有“测验多(Test heavy)”(1.83%),“太多论文(So many papers)”(1.50%)等。相比之下,低分的教授通常会得到“给分低(Tough grader)”(14.67%),“逃课?你会挂科(Skip class? You won't pass.)”(9.66%)和“课前准备阅读(Get ready to read)”(9.33%)等标签,出现频率较低的标签有“滑稽的(Hilarious)”(1.93%),“鼓舞人心的(Inspirational)”(1.58%)和“精彩演讲(Amazing lectures)”(1.22%)等标签。

表 5.2 高分教授和低分教授的标签频率及卡方检验

标签	高分教授标签	低分教授标签	Chi-Square (df = 1)	p	Cohen's w
1.Gives good feedback	↑ 9.54%	3.85%	396134.20	< .001	0.74
2.Caring	↑ 8.57%	3.21%	425095.16	< .001	0.77
3.Respected	↑ 8.56%	2.10%	429633.89	< .001	0.77
4.Participation matters	6.86%	6.54%	462407.00	< .001	0.80
5.Clear grading criteria	6.65%	3.36%	480192.10	< .001	0.82
6.Skip class? you won't pass.	6.23%	↑ 9.66%	470660.13	< .001	0.81
7.Amazing lectures	6.06%	↓ 1.22%	506682.51	< .001	0.84
8.Inspirational	6.02%	↓ 1.58%	506554.17	< .001	0.84
9.Tough grader	5.68%	↑ 14.67%	472094.74	< .001	0.81
10.Hilarious	5.57%	↓ 1.93%	519195.02	< .001	0.85
11.Get ready to read	5.26%	↑ 9.33%	501930.26	< .001	0.83
12.Lots of homework	4.32%	9.10%	533139.82	< .001	0.86
13.Accessible outside class	4.26%	2.32%	559866.32	< .001	0.88
14.Lecture heavy	4.19%	9.00%	537966.92	< .001	0.86
15.Extra credit	3.31%	2.33%	591466.24	< .001	0.91
16.Graded by few things	2.32%	4.98%	615816.08	< .001	0.92
17.Group projects	1.99%	3.35%	633691.38	< .001	0.94
18.Test heavy	↓ 1.83%	5.22%	632362.24	< .001	0.94
19.So many papers	↓ 1.50%	3.54%	650505.36	< .001	0.95
20.Beware of pop quizzes	↓ 1.26%	2.70%	662529.21	< .001	0.96
总计	100.00%	100.00%			



图 5.2 高分教授(左)和低分教授(右)的标签词云图

5.3.3 助理教授、副教授与教授的教学特点分析

助理教授通常具有不到6年的教学经验，副教授具有6年以上、12年以下的教学经验，而教授具有12年以上的教学经验。通过比较助理教授，副教授和教授的标签发现，助理教授通常得到“给予良好反馈(Gives good feedback)”(8.31%)，“关怀(Caring)”(7.48%)和“尊重(Respected)”(6.72%)等标签，副教授通常会得到“给予良好反馈(Gives good feedback)”(8.66%)，“给分低(Tough grader)”(8.33%)和“尊重(Respected)”(7.66%)等标签，教授通常会得到“给予良好反馈(Gives good feedback)”(8.92%)，“尊重(Respected)”(8.41%)和“给分低(Tough grader)”(8.17%)等标签。

表 5.3 助理教授、副教授和教授的标签频率及卡方检验

标签	助理教授	副教授	教授	Chi-Square (<i>df</i> =2)	<i>p</i>	Cohen's <i>w</i>
1.Gives good feedback	8.31%	8.66%	8.92%	572378.96	< .001	0.91
2.Caring	7.48%	7.48%	7.81%	539656.01	< .001	0.89
3.Respected	6.72%	7.66%	8.41%	563020.95	< .001	0.91
4.Skip class? you won't pass.	6.72%	7.25%	6.94%	551002.77	< .001	0.90
5.Participation matters	6.66%	7.12%	6.74%	547704.01	< .001	0.89
6.Tough grader	6.59%	8.33%	8.16%	598028.89	< .001	0.93
7.Clear grading criteria	6.44%	5.86%	5.18%	516475.93	< .001	0.87
8.Get ready to read	5.70%	6.56%	6.43%	551637.61	< .001	0.90
9.Amazing lectures	5.22%	4.86%	4.87%	523297.66	< .001	0.87
10.Lecture heavy	5.13%	5.26%	5.35%	532296.82	< .001	0.88
11.Lots of homework	5.03%	5.50%	5.11%	541702.25	< .001	0.89
12.Inspirational	4.95%	5.21%	5.35%	535702.82	< .001	0.88
13.Hilarious	4.87%	4.73%	4.81%	529840.23	< .001	0.88
14.Accessible outside class	4.63%	2.91%	3.76%	519931.31	< .001	0.87
15.Extra credit	3.59%	2.88%	2.08%	553367.62	< .001	0.90
16.Graded by few things	3.14%	2.31%	2.81%	561260.59	< .001	0.90
17.Test heavy	2.90%	1.95%	2.10%	571858.26	< .001	0.91
18.Group projects	2.36%	2.09%	2.11%	585635.40	< .001	0.92
19.So many papers	1.90%	1.95%	1.78%	599972.96	< .001	0.94
20.Beware of pop quizzes	1.67%	1.42%	1.27%	611998.65	< .001	0.94



图 5.3 助理教授(左)、副教授(中)、教授(右)的标签词云图

5.3.4 助理教授、副教授与教授的综合得分与教龄的关系分析

图 5.4 是教授的教学效能得分与教龄之间的变化趋势。结果表明,教授的教学效能与教龄是正相关($r = 0.01, p < .001$),但是相关系数非常低,随着教授教龄的增加,教授的教学效能基本是不变的。助理教授、副教授与教授的综合得分尽管存在差异($Kruskal-Wallis H = 2570.45, p < .001, \eta^2 = 0.00$),但是效应量很小。经过事后检验发现,教授的综合得分($N = 50261, M = 3.77, SD = 0.77$)高于助理教授的综合得分($N = 609193, M = 3.75, SD = 1.10$)和副教授的综合得分($N = 260284, M = 3.75, SD = 0.83$)。具有 16 年以上教学经验的教授的标准误相对较大,这是因为随着教龄的增加,教授的人数在迅速减少,仅有 156 位超过 17 年教学经验的教授。因此,由于样本量小且差异大,建议忽略教龄超过 16 年的教授。

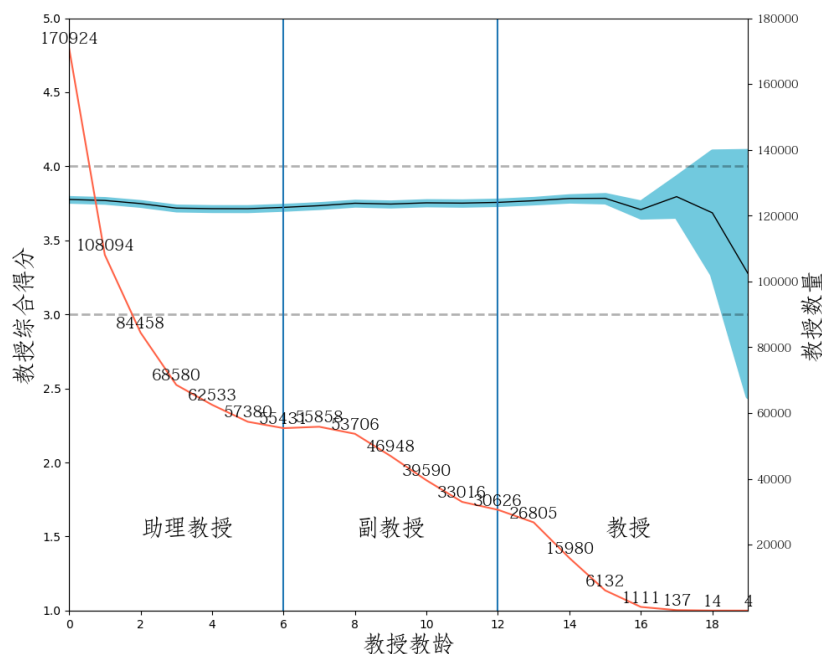


图 5.4 教师综合得分与教授教龄的关系

注：红线是不同教龄的教授数量，两条蓝线分别是助理教授与副教授、副教授与教授的教龄分隔线，黑线及蓝色阴影分别是教授教学效能综合得分平均分及标准误(95%置信区间)。

5.3.5 课程难度、再次选课比例与教授综合得分的关系分析

课程难度($M = 2.86, SD = 0.91$)和教授的综合得分($M = 3.75, SD = 1.01$)之间的呈显著负相关($r = -0.45, p < .001$), 即课程难度越高, 教授的综合得分越低。教授的综合得分($M = 3.82, SD = 0.81$)和再次选课($M = 0.71, SD = 0.27$) 呈显著正相关($r = 0.80, p < .001$)。教授综合得分($M = 3.82, SD = 0.81$)与是否再次选课($M = 0.7, SD = 0.27$)呈显著正相关($r = 0.80, p < .001, R^2 = 0.64$), 即综合得分越高, 学生再次选择课程的可能性越大。

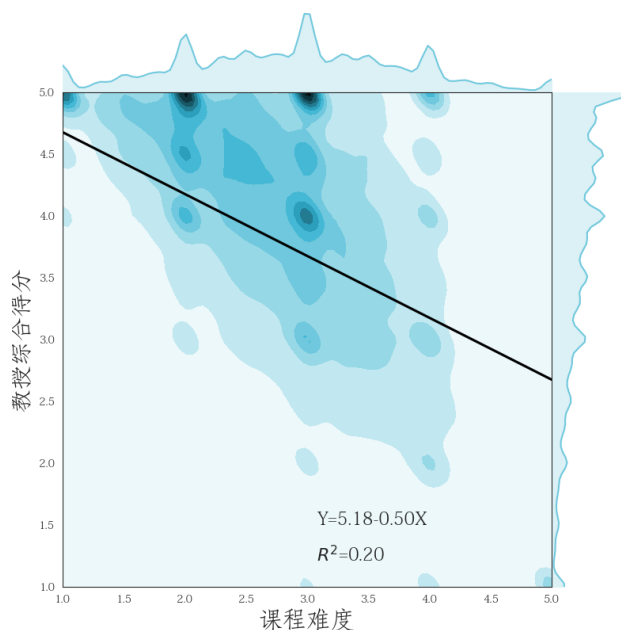


图 5.5 课程难度与教授综合得分的相关关系

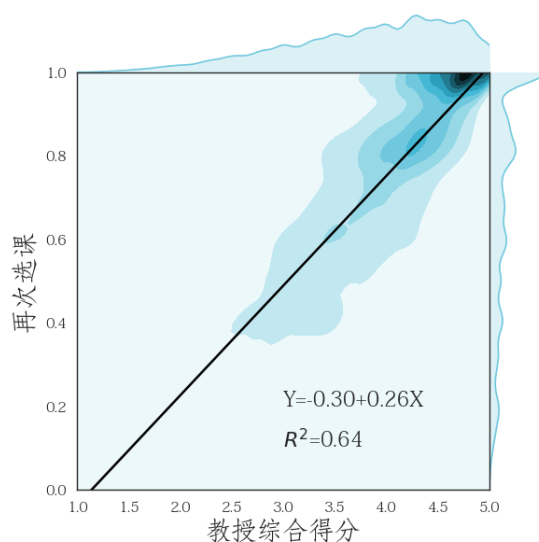


图 5.6 再次选课比例与教授综合得分的相关关系

5.3.6 学生成绩、出勤与教授综合得分的关系分析

学生是否再次选择教授课程来也是本研究关注的问题。如图 5.7 所示，每个学生除了给予教授综合评分外，还会说明是否为了学分选择本课程、本课程是否强制出勤及自己的期末成绩。



图 5.7 RMP 上学生给予教授评价的示例图

表 5.4 表明，教授给予学生的成绩都比较高，A+(15.03%)和 A(35.65%)两类成绩占比已经达到 50%。图 5.8 是学生的成绩与教授评分之间的关系。结果表明，学生是否为了学分而选择本课程，在给予教授的评分上存在显著差异(Mann-Whitney $U = 20842019452$, $p < .001$, Cohen's $d = 0.29$)。为了学分而选择本课程的学生($N = 2487432$, $M = 3.69$, $SD = 1.45$)给予教授的评分显著高于不是为了学分而选择本课程学生给予教授的评分($N = 14810$, $M = 3.23$, $SD = 1.70$)。

通过分析课程是否强制出勤与学生给予教授评分的关系发现，尽管强制出勤及非强制出勤的学生给予教授的评分存在显著差异，但是效应量较小(Mann-Whitney $U = 601251982391.5$, $p < .001$, Cohen's $d = 0.10$)。选择强制出勤课程的学生给予教授的评分($N = 1721204$, $M = 3.73$, $SD = 1.46$)显著高于不强制出勤的学生给予教授的评分($N = 655633$, $M = 3.58$, $SD = 1.47$)。

得到不同成绩的学生给予的教授有显著差异(Kruskal-Wallis $H = 384846.43$, $p < .001$, $\eta^2 = 0.20$)。由图 5.8 可知，学生的分数越高，教授的评分就越高。图 5.10 显示了是否再次选择本课程和学生的成绩分布。从该图中可以看到，期末成绩是 A+到 B 的学生更有可能再次选择这门课程。

表 5.4 不同期末成绩的学生给予教授综合得分的描述统计

学生成绩	<i>M</i>	<i>SD</i>	<i>N</i>	成绩占比
A+	4.45	0.99	294055	15.07%
A	4.26	1.08	695633	35.65%
A-	3.94	1.25	172567	8.84%
B+	3.79	1.33	191888	9.83%
B	3.49	1.41	213334	10.93%
B-	2.97	1.44	52116	2.67%
C+	2.76	1.46	47142	2.42%
C	2.64	1.44	62657	3.21%
C-	2.12	1.28	18366	0.94%
D+	1.98	1.24	9483	0.49%
D	1.94	1.24	16025	0.82%
D-	1.72	1.14	4278	0.22%
F(不及格)	1.81	1.28	17189	0.88%
Drop/Withdrawal(退课)	1.80	1.19	30402	1.56%
Incomplete(未完成)	2.91	1.64	9414	0.48%
Not sure yet(未报告)	3.18	1.57	113687	5.83%
Audit/No Grade(旁听生)	3.15	1.67	3197	0.16%
总计			1951433	100.00%

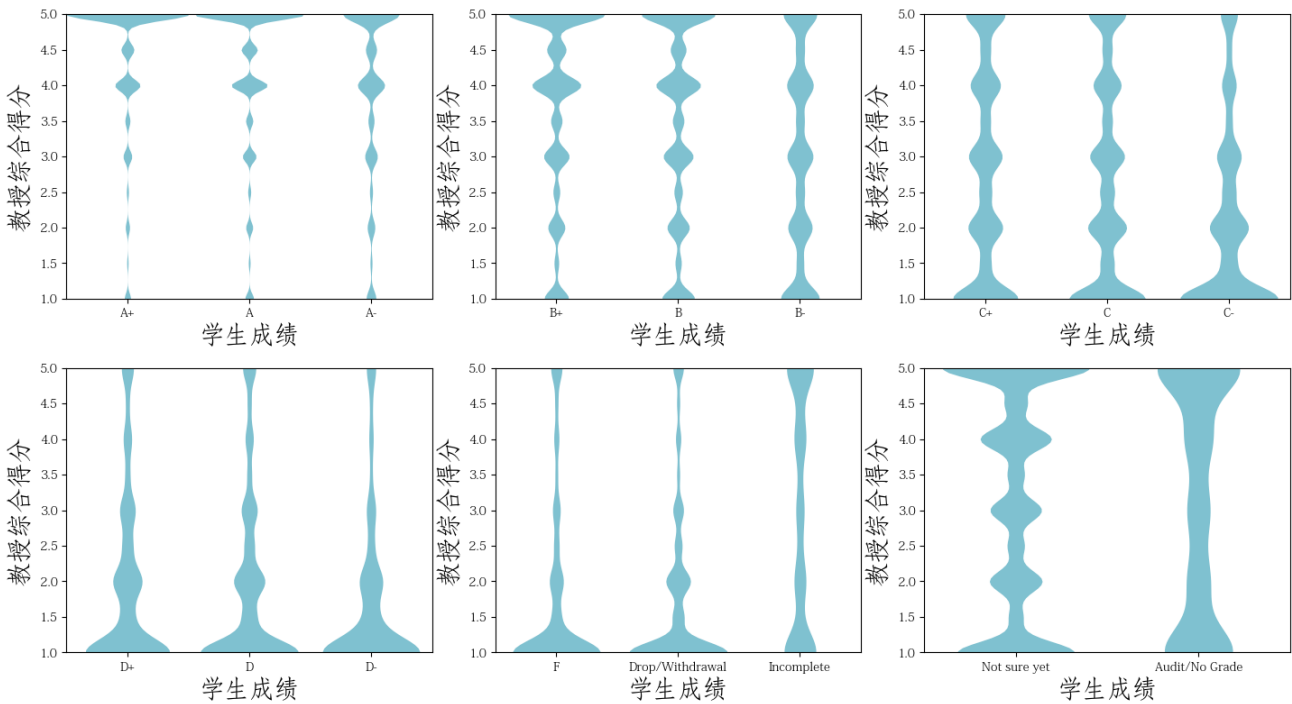


图 5.8 不同学生的成绩与学生给予教授综合得分的小提琴图

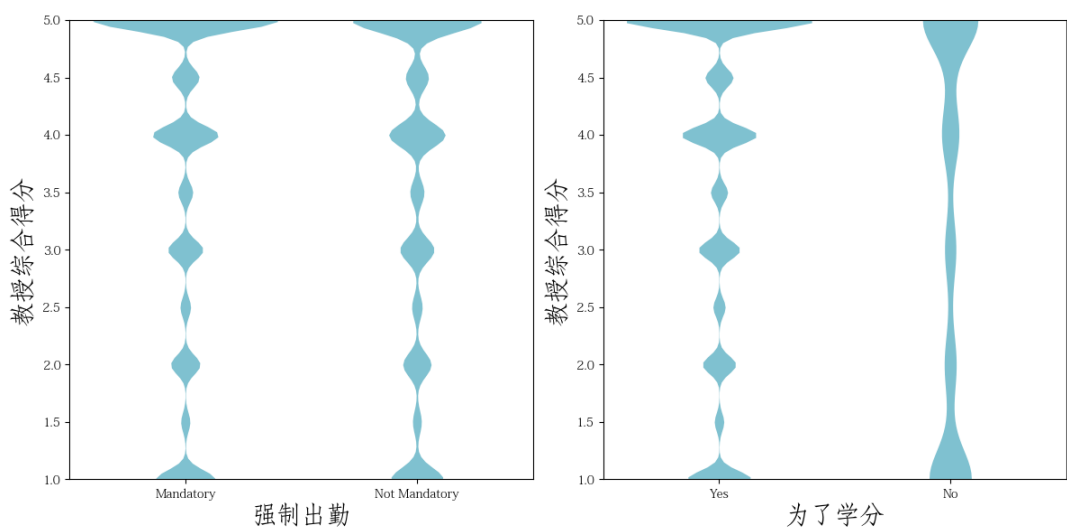


图 5.9 课程是否出勤强制与教授综合得分的小提琴图及是否为了学分与教授综合得分的小提琴图

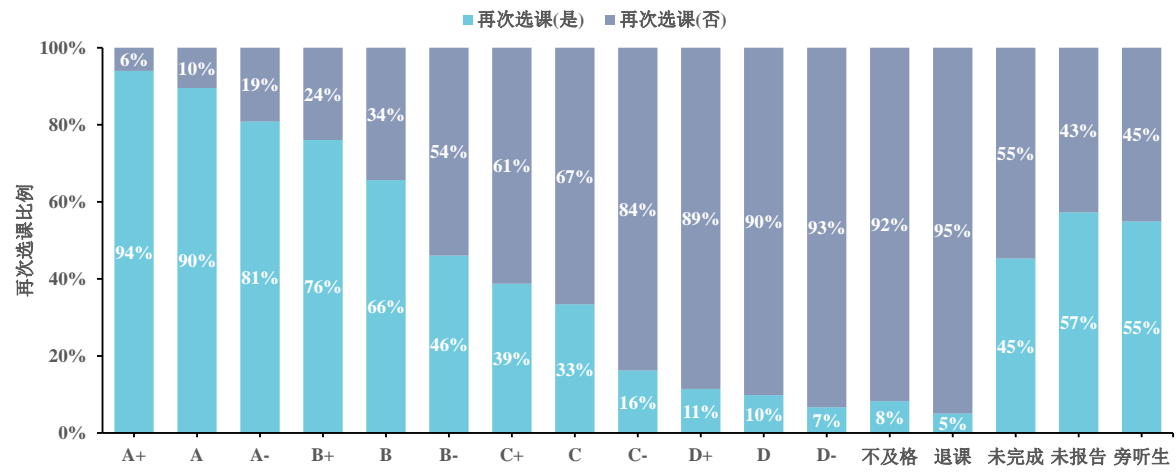


图 5.10 学生的成绩与是否再次选课的堆积图

5.3.7 教授的线下课程及在线课程的综合得分及课程难度分析

在RMP上，教授的线下课程($N = 9294420$, $M = 3.70$, $SD = 1.42$)的得分显著高于在线课程($N = 247365$, $M = 3.45$, $SD = 1.56$)的得分(Mann–Whitney $U = 1063841030165$, $p < .001$, Cohen's $d = 0.17$)。而且在线课程 ($N = 247365$, $M = 3.03$, $SD = 1.27$)的课程难度显著高于线下课程($N = 9294420$, $M = 2.90$, $SD = 1.29$)的课程难度(Mann–Whitney $U = 1212813286011$, $p < .001$, Cohen's $d = 0.10$)。

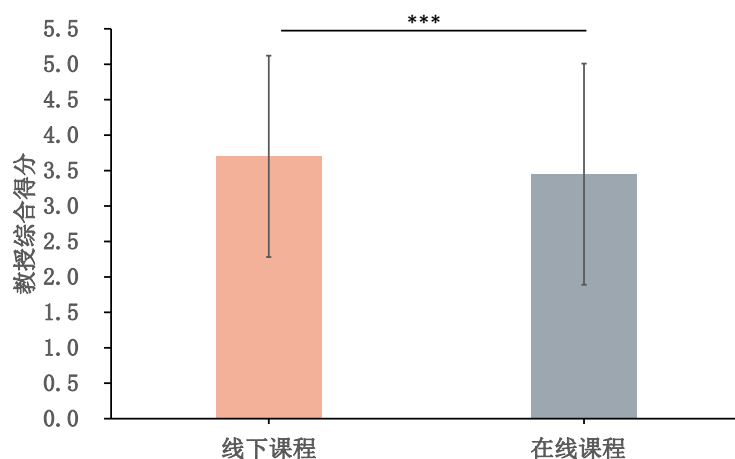


图5.11 RMP线下课程及在线课程的综合得分的平均分及标准差

对教授的职称与课程类型进行非参数多因素方差分析，采用R语言的ARTool(Wobbrock, Findlater, Gergle, & Higgins, 2011; Kay & Wobbrock, 2019)。经分析发现，RMP上的不同职称教授(助理教授、副教授及教授)与课程类型(在线课程与线下课程)交互作用显著($F_{(2,78443338)} = 1237.12$, $p < .001$, $\eta_p^2 = 0.00$)，但是效应量非常小，描述统计见表6.8。其次，RMP课程水平可以根据课程号判断，因此本研究根据课程号提取难度为1-4的课程，难度从低到高，通过多因素非参数方差分析可知，不同职称教授(助理教授、副教授及教授)与课程难度水平(课程号是1,2,3,4)交互作用显著($F_{(2,78443318)} = 2483.2$, $p < .001$, $\eta_p^2 = 0.00$)，但是其效应量依然非常小。

表5.5 不同职称教师的线下课程及在线课程得分的描述统计

教师职称	线下课程			在线课程		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
助理教授	3887463	3.68	1.44	125946	3.49	1.56
副教授	3129305	3.71	1.39	85083	3.41	1.54
教授	604458	3.76	1.37	12085	3.41	1.55

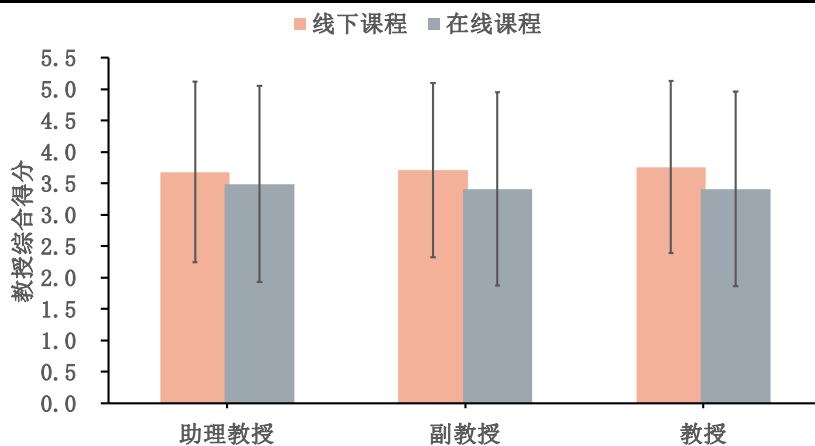


图5.12 不同职称教授线下课程、在线课程的综合得分平均分及标准差

表5.6 教师的课程难度水平得分的描述统计

课程难度水平	助理教授			副教授			教授		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
1	2205415	3.68	1.45	1612060	3.72	1.39	276917	3.77	1.37
2	1019658	3.66	1.44	816693	3.69	1.39	163376	3.73	1.37
3	576306	3.64	1.45	558152	3.68	1.39	123303	3.72	1.38
4	212030	3.71	1.45	227483	3.74	1.40	52947	3.76	1.41

5.3.8 不同学科的线下课程及在线课程的综合得分及课程难度分析

根据RMP课程的缩写识别12门课程，课程分别如下：会计学、艺术学、商科、化学、计算机科学、经济学、英语、法律、数学、物理学、心理学、社会科学。

通过非参数检验发现，会计学的线下课程得分($M = 3.55$, $SD = 1.74$)与在线课程得分($M = 3.46$, $SD = 1.53$)存在显著差异(Mann-Whitney $U = 97143706$, $p < .05$, Cohen's $d = 0.06$)；艺术学线下课程得分($M = 3.76$, $SD = 1.49$)与在线课程($M = 3.52$, $SD = 1.52$)存在显著差异(Mann-Whitney $U = 273249809$, $p < .05$, Cohen's $d = 0.16$)；商科线下课程得分($M = 3.72$, $SD = 1.44$)与在线课程($M = 3.53$, $SD = 1.56$)存在显著差异(Mann-Whitney $U = 295743515$, $p < .001$, Cohen's $d = 0.12$)；计算机科学线下课程得分($M = 3.50$, $SD = 1.50$)与在线课程($M = 3.38$, $SD = 1.59$)存在显著差异(Mann-Whitney $U = 383977283$, $p < .05$, Cohen's $d = 0.16$)；英语线下课程得分($M = 3.78$, $SD = 1.40$)与在线课程($M = 3.41$, $SD = 1.58$)存在显著差异(Mann-Whitney $U = 7527711768$, $p < .001$, Cohen's $d = 0.25$)；法律线下课程得分($M = 3.79$, $SD = 1.49$)与在线课程($M = 3.51$, $SD = 1.54$)存在显著差异(Mann-Whitney $U = 8443821$, $p < .001$, Cohen's $d = 0.19$)；物理学线下课程得分($M = 3.44$, $SD = 1.48$)与在线课程($M = 3.22$, $SD = 1.60$)存在显著差异(Mann-Whitney $U = 86434146$, $p < .001$, Cohen's $d = 0.14$)；心理学线下课程得分($M = 3.86$, $SD = 1.34$)与在线课程($M = 3.66$, $SD = 1.50$)存在显著差异(Mann-Whitney $U = 2612794805$, $p < .001$, Cohen's $d = 0.14$)；社会科学线下课程得分($M = 3.72$, $SD = 1.40$)与在线课程($M = 3.44$, $SD = 1.54$)存在显著差异(Mann-Whitney $U = 775892896$, $p < .001$, Cohen's $d = 0.19$)，余下2门课程化学、数学的在线课程及线下课程的综合得分不存在显著差异。尽管10门课程存在显著差异，但是会计学、艺术学、商科、计算机科学、经济学、法律、物理学、心理学、社会科学学科的效应量较小，英语学科的为中等。

表5.7 不同学科的线下课程及在线课程综合得分分析

学科	线下课程			在线课程			Mann-Whitney <i>U</i>	<i>p</i>	Cohen's <i>d</i>
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>			
会计学	84387	3.55	1.47	2359	3.46	1.53	97143706	.04*	0.06
艺术学	160460	3.76	1.39	3667	3.52	1.52	273249809	< .001	0.16
商科	113760	3.72	1.44	5476	3.53	1.56	295743515	< .001	0.12
化学	215963	3.44	1.47	1474	3.40	1.58	160100815	.69	0.03
计算机科学	181966	3.50	1.50	4363	3.38	1.59	383977283	< .001	0.08
经济学	147635	3.41	1.45	3894	3.34	1.53	281625020	.03*	0.05
英语	792289	3.78	1.40	21548	3.41	1.58	7527711768	< .001	0.25
法律	36086	3.79	1.40	513	3.51	1.54	8443821	< .001	0.19
数学	476787	3.53	1.51	8354	3.50	1.58	2005801141	.25	0.01
物理学	148462	3.44	1.48	1250	3.22	1.60	86434146	< .001	0.14
心理学	392230	3.86	1.34	13992	3.66	1.50	2612794805	< .001	0.14
社会科学	218067	3.72	1.40	7784	3.44	1.54	775892896	< .001	0.19

注：* $p < .05$ 。

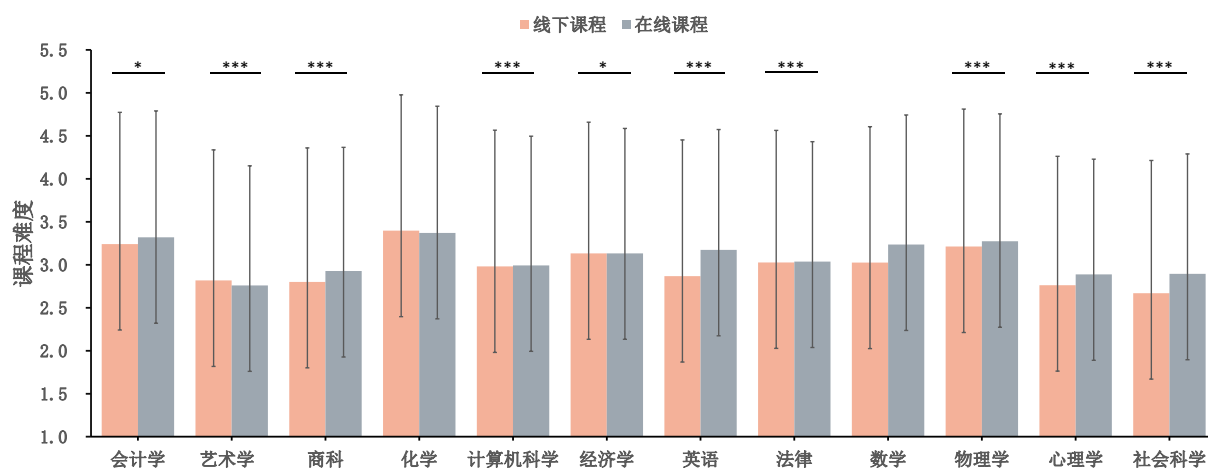


图5.12 不同学科的在线课程、线下课程的综合得分平均分及标准差

通过分析不同学科的难度发现，会计学的线下课程难度($M = 3.24$, $SD = 1.28$)与在线课程难度($M = 3.32$, $SD = 1.19$)存在显著差异(Mann-Whitney $U = 102764668$, $p < .05$, Cohen's $d = 0.06$)；艺术学的线下课程难度($M = 2.82$, $SD = 1.29$)与在线课程难度($M = 2.76$, $SD = 1.33$)存在显著差异(Mann-Whitney $U = 285817932$, $p < .05$, Cohen's $d = 0.04$)；商科的线下课程难度($M = 2.80$, $SD = 1.30$)

与在线课程难度($M = 2.93$, $SD = 1.28$)存在显著差异(Mann-Whitney $U = 328941639$, $p < .001$, Cohen's $d = 0.10$); 英语的线下课程难度($M = 2.87$, $SD = 1.27$)与在线课程难度($M = 3.17$, $SD = 1.24$)存在显著差异(Mann-Whitney $U = 9703575329$, $p < .001$, Cohen's $d = 0.24$); 数学的线下课程难度($M = 3.03$, $SD = 1.29$)与在线课程难度($M = 2.76$, $SD = 1.33$)存在显著差异(Mann-Whitney $U = 2172945322$, $p < .05$, Cohen's $d = 0.04$); 心理学的线下课程难度($M = 2.76$, $SD = 1.25$)与在线课程难度($M = 2.89$, $SD = 1.26$)存在显著差异(Mann-Whitney $U = 2900003570$, $p < .001$, Cohen's $d = 0.10$); 社会科学的线下课程难度($M = 2.67$, $SD = 1.25$)与在线课程难度($M = 2.89$, $SD = 1.27$)存在显著差异(Mann-Whitney $U = 934003821$, $p < .001$, Cohen's $d = 0.18$)。

表5.8 不同学科的在线课程、线下课程课程难度分析

学科	线下课程			在线课程			Mann-Whitney U	p	Cohen's d
	N	M	SD	N	M	SD			
会计学	84387	3.24	1.28	2359	3.32	1.19	102764668	$< .05$	0.06
艺术学	160460	2.82	1.29	3667	2.76	1.33	285817932	$< .05$	0.04
商科	113760	2.80	1.30	5476	2.93	1.28	328941639	$< .001$	0.10
化学	215963	3.40	1.23	1474	3.37	1.22	157279736	.42	0.02
计算机科学	181966	2.98	1.32	4363	2.99	1.30	399477597	.47	0.01
经济学	147635	3.13	1.27	3894	3.13	1.29	287964359	.85	0.00
英语	792289	2.87	1.27	21548	3.17	1.24	9703575329	$< .001$	0.24
法律	36086	3.03	1.25	513	3.04	1.35	9336559	.73	0.01
数学	476787	3.03	1.31	8354	3.24	1.23	2172945322	$< .001$	0.17
物理学	148462	3.21	1.29	1250	3.27	1.28	95319671	.09	0.05
心理学	392230	2.76	1.25	13992	2.89	1.26	2900003570	$< .001$	0.10
社会科学	218067	2.67	1.25	7784	2.89	1.27	934003821	$< .001$	0.18

5.4 讨论

通过标签分析发现, 高分的教授标签中出现频率比较高的是“给予良好反馈(Gives good feedback)”(9.54%), “关怀(Caring)”(8.57%)和“受人尊敬(Respected)”(8.56%)等比较积极的标签, 低分的教授通常出现“给分低(Tough grader)”(14.67%), “逃课? 你会挂科(Skip class? You won't pass.)”(9.66%)等比较消极的标签, 从这里看出, 高分教授和低分教授在教学上采取了不同的方式, 学生给予了不同的评分, 教授可以通过增加反馈、给予学生更多的关怀和鼓励来提高学生对其教学效能的评价。助理教授、副教授、教授在标签上出现比较多都有“给予良好反馈

(Gives good feedback)”,说明教授在学习上给予学生很多的反馈,帮助学生及时了解学习进展,而副教授、教授相比于助理教授在标签上出现比较多的还有“给分低(Tough grader)”,说明副教授、教授在教学上要比助理教授严苛。

随着教授教龄的增加,教授的综合得分基本是不变的。助理教授、副教授和教授的综合得分尽管存在显著差异,但是其效应量较小,可以忽略。教授的综合得分($N=50261$, $M=3.77$, $SD=0.77$)高于助理教授的综合得分($N=609193$, $M=3.75$, $SD=1.10$)和副教授的综合得分($N=260284$, $M=3.75$, $SD=0.83$),获得终身职位的教授,其综合得分并没有下降,反而略有上升了。有研究表明,终身制对教师的教学有很大的激励作用,可以帮助教授在教学和研究方面取得成功。终身制给予职业保障,教授可以将更多时间用于教学和提高绩效。与非终身教授相比,终身教授对学生提高毕业率更有利(Sav, 2016)。在美国,约有70%的教职员工不在终身任职之列,他们因工作是临时职位而感到压力(Reevy & Deason, 2014)。非终身制教授的教学经验不足会对本科生的表现产生负面影响(Banasik & Dean, 2016)。就目前而言,终身制度对教授个体和大学等机构依然起着至关重要的作用。教授的课程难度和教授的教学效能之间的呈显著负相关($r=-0.45$, $p<.001$),教授的课程难度越高,学生给予教授的评分越低。为了学分而选择本课程的学生往往会给予教授的更好评分。从这里可以看出,教授通过降低课程难度、将课程设置成学分课能帮助教师提高教学效能,而且学生期末成绩越高,学生也将给予教授更高的评分,教师的教学效能与学生的成绩呈正相关(Good, 1979)。期末成绩越高,再次选择这门课程的比例也会更高。

在线评估的主要优势之一是每个人都可以使用它们,在学期初退课的学生也是非常重要的一群人,这些学生的反馈意见非常重要,在RMP中,退课(WD)占1.55%。尽管已退课,学生也对教授进行评分,这些学生给予教授的评分的平均值仅为5分制的1.80。这意味着即使学生退课,他们也可以给出负面反馈或较低的评分。这类似于电子商务网站上的评论,匿名评论意味着即使客户没有购买,他们仍然给出负面反馈(Anderson & Simester, 2014)一样,对于这种偏见反馈是目前公开评论中无法避免的,而且学生的反馈对教师的教学评估也存在一定偏见(Chitre & Srinivasan, 2018),因此也要谨慎对待学生的单方面反馈。

基于RMP探索国外教授的在线课程及线下课程教学效能差异,发现教授的线下课程的综合得分高于在线课程,而在线课程的课程难度高于线下课程。通过对RMP上的12门学科分析发现,会计学、艺术学、商科、计算机科学、英语、法律、物理学、心理学、社会科学10门课程的在线课程及线下课程的综合得分存在显著差异,而且线下课程的综合得分均高于在线课程的得分。

综上所述,本研究发现随着教学年龄的增长,教师的教学效能的综合得分保持不变。助理教授、副教授、教授之间的教学效能综合得分尽管存在显著差异,但效应量较少,建议忽略,不过教授的教学效能综合得分要好于副教授及助理教授。对于教师而言,通过在课堂内外关注

学生成长、积极给予学生良好的反馈，同时在课程设置上强制要求学生出勤及将课程设置为学分课，教师可以获得更好教学效能综合得分。**RMP** 上的教授评价结果会影响学生对教授的选择(Swiggard & Seidel, 2012)，因此，教授可以时刻关注 **RMP** 上的评价结果，以改进自己的教学方式。对于学生而言，通过 **RMP** 选择较高教学效能的教师也非常重要。教学经验丰富的教师，教学效能一般也比较高，而且学生想要往往也会获得好的成绩(Kini & Podolsky, 2016)。基于 **RMP** 上的 12 门学科的教学效能综合得分和课程难度的分析结果发现，教授在线下课程上的教学效能得分均高于在线课程，在线课程的难度普遍高于在线课程。所以对教师而言，可以适当降低在线课程的难度以满足学生的学习需要，并获得较高的教学效能评价。

第6章 研究二：线下课程及在线课程教学效能的差异研究

6.1 目的

当下在线课程兴起，对于教师的教学策略、教学方法等提出了新的要求，针对教师在线课程及线下课程是否存在不同的教学效能是本研究关注的问题。

6.2 方法

6.2.1 研究对象

本研究调查了北京联合大学及济南大学的大学生，采用随机抽样方法，通过网络调研最终收集到375份有效问卷，其中调查大学生，男156人($M = 21.98$ 岁, $SD = 2.39$ 岁)，女219人($M = 21.59$ 岁, $SD = 2.70$ 岁)。从年级上看，大一学生98人，大二学生56人，大三学生87人，大四学生69人，研究生65人。被评价的讲师有182人，副教授有107人，教授有86人。

6.2.2 研究工具

本研究使用的问卷是由北京师范大学孟庆茂、刘红云(2003)翻译及修订的《大学教师教学效果评价问卷》，本问卷应用广泛，信、效度良好。本问卷有32个题目，分为7个维度，分别是学习价值感，教学热情与组织，群体互动，人际和谐，知识宽度，教学管理，功课量及难度。问卷采取Likert 5点评分，要求学生在项目描述中选择与自己的最符合情况，从完全反对、不同意、一般、比较同意、非常同意做出选择。其中，非常同意计5分、完全反对计1分。本问卷最终以所有项目的平均分作为教师教学效果综合得分，即教师的教学效能得分。

由于本次不同职称教师的样本量差异较大，因此首先对数据进行正态性检验，结果显示各个维度的P-P图中数据点与理论直线基本重合，因此各维度的数据符合正态分布，可以进行正态检验。

6.3 结果

6.3.1 国内学生给予教师线下课程及在线课程教学效果得分的描述统计

根据表6.1可知，男生与女生给教师在线课程的得分均高于线下课程得分。针对不同年级学生对教师不同课程类型的评分统计发现，大一、大二、大三的学生给教师在线课程的得分低于线下课程的得分，大四、研一、研二、研三的学生给予教师在线课程的得分高于线下课程的得分，从这里可以看出，低年级的学生倾向于给教师线下课程更高的得分，而高年级学生倾向于给教师在线课程更高的得分。

表 6.1 不同性别学生给国内教师教学效果得分的描述统计

性别	线下课程			线上课程		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
男生	81	3.96	0.56	75	4.03	0.50
女生	98	4.12	0.43	121	4.14	0.49

表 6.2 不同年级学生给国内教师教学效果得分的描述统计

年级	线下课程			在线课程		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
大一	23	4.37	0.42	75	4.23	0.43
大二	35	4.15	0.43	21	4.11	0.49
大三	39	4.09	0.31	48	3.87	0.59
大四	49	3.82	0.59	20	3.94	0.41
研一	9	3.65	0.48	12	4.04	0.39
研二	11	4.29	0.46	14	4.36	0.34
研三	13	3.98	0.40	6	4.28	0.48

6.3.2 国内教师线下课程及在线课程的教学效果得分分析

国内教师线下课程及在线课程及的教学效果得分方差齐性($F_{(1,373)} = 0.10, p > .05$), 在线课程($M = 4.09, SD = 0.49$)与线下课程($M = 4.04, SD = 0.50$)的得分不存在显著差异($t = -1.02, p > .05$)。对学生性别及课程类型进行方差齐性检验, 结果显示方差齐性($F_{(1,373)} = 2.62, p > .05$), 学生的性别与国内教师的课程类型(线下课程、在线课程)之间不存在交互作用($F_{(1,371)} = 0.12, p > .05$)。

不同性别学生对国内教师的线下课程及在线课程的教学效果得分的主效应显著($F_{(1,373)} = 6.686, p < .05, \eta_p^2 = 0.02$), 男生给教师在线课程的得分($M = 4.03, SD = 0.50$)高于线下课程($M = 3.96, SD = 0.56$), 女生给教师在线课程的得分($M = 4.14, SD = 0.49$)高于线下课程($M = 4.12, SD = 0.43$)。不同学科性质(理科、文科)的学生对国内教师的线下课程及在线课程的综合得分不存在交互作用($F_{(1,371)} = 0.70, p > .05$), 学生的年级与国内教师的课程类型(线下课程及在线课程)交互作用不显著($F_{(6,361)} = 2.16, p > .05, \eta_p^2 = 0.04$)。

表 6.3 学生性别与课程类型的方差分析

变异来源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
学生性别	1.55	1	1.55	6.38	.01*
课程类型	0.20	1	0.20	0.82	.37
学生性别*课程类型	0.05	1	0.05	0.21	.64

注：* $p < .05$ 。

6.3.3 不同职称国内教师的线下课程及在线课程的教学效果得分分析

首先对国内教师综合得分进行方差齐性检验,结果显示,不同职称教师的教学效果得分方差齐性($F_{(2,372)} = 1.30, p > .05$),通过单因素方差分析发现,讲师($M = 4.10, SD = 0.52$)、副教授($M = 4.05, SD = 0.44$)、教授($M = 4.05, SD = 0.51$)的综合得分不存在显著差异($F_{(2,372)} = 0.40, p > .05$)。通过两因素方差分析发现,教师职称及课程类型方差齐性($F_{(2,372)} = 2.10, p > .05$),国内教师的职称与课程类型交互作用显著($F_{(1,369)} = 3.19, p < .05, \eta_p^2 = 0.02$)。经简单主效应分析发现,讲师线下课程($M = 3.99, SD = 0.59$)及线上课程($M = 4.17, SD = 0.46$)存在显著差异($F_{(1,369)} = 5.73, p < .05$),但副教授($F_{(1,369)} = 5.73, p > .05$)及教授($F_{(1,369)} = 5.73, p > .05$)的线下、线上课程不存在显著差异。

表 6.4 不同职称国内教师线下课程及在线课程的教学效果得分的描述统计

教师职称	线下课程			在线课程		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
讲师	76	3.99	0.59	106	4.17	0.46
副教授	56	4.07	0.39	51	4.04	0.50
教授	47	4.10	0.45	39	3.98	0.56

表 6.5 国内教师线下课程及在线课程教学效果得分分析

变异来源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
教师职称	0.11	2	0.06	0.23	.79	0.00
课程类型	0.00	1	0.00	0.02	.90	0.00
教师职称*课程类型	1.56	2	0.78	3.19	.04*	0.02

注：* $p < .05$ 。

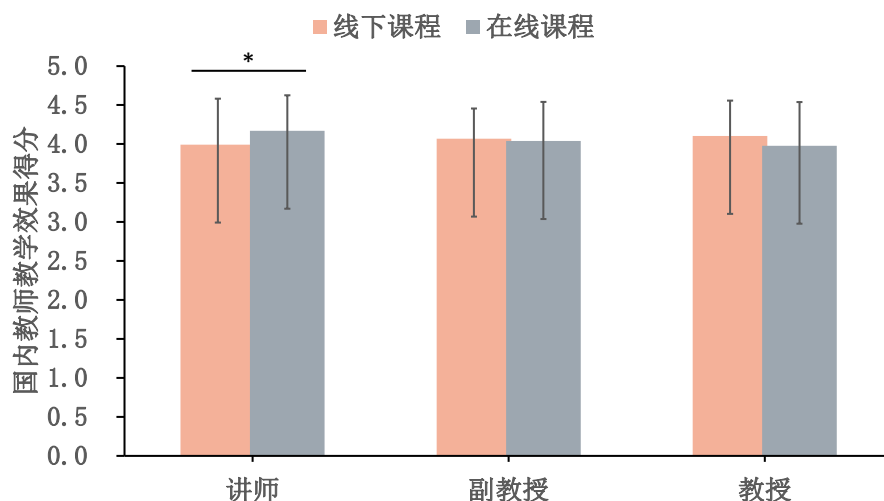


图 6.1 不同职称国内教师线下课程及在线课程的教学效果得分平均分及标准差

6.3.4 国内教师线下课程及在线课程的教学效果各维度分析

国内教师线下课程的功课量得分($M = 3.09$, $SD = 0.57$)及在线课程功课量得分($M = 3.27$, $SD = 0.61$)存在显著差异($t_{(373)} = -2.918$, $p < .05$, $Cohen's d = 0.16$), 即在线课程的课程难度高于线下课程、课程进度快于线下课程、作业量多于线下课程, 描述统计见表 6.7。除了功课量得分存在显著差异外, 其他维度, 如学习价值感($t_{(373)} = -0.28$, $p > .05$), 教学热情与组织($t_{(373)} = -0.48$, $p > .05$), 教学管理($t_{(373)} = -0.59$, $p > .05$), 群体互动($t_{(373)} = -1.018$, $p > .05$), 人际和谐($t_{(373)} = -0.09$, $p > .05$)和知识宽度($t_{(373)} = -1.67$, $p > .05$)6 个维度上均不存在显著差异。

表 6.6 国内教师线下课程及在线课程各维度的得分分析

变量	线下课程			线上课程			t	df	p	$Cohen's d$
	N	M	SD	N	M	SD				
学习价值感	179	4.23	0.60	196	4.24	0.62	-0.28	373	.78	0.02
教学热情与组织	179	4.20	0.65	196	4.23	0.62	-0.48	373	.63	0.03
教学管理	179	4.13	0.52	196	4.17	0.55	-0.59	373	.55	0.03
群体互动	179	4.19	0.64	196	4.25	0.58	-1.02	373	.31	0.06
人际和谐	179	4.04	0.65	196	4.05	0.63	-0.09	373	.93	0.01
知识宽度	179	4.12	0.58	196	4.21	0.55	-1.67	373	.10	0.09
功课量	179	3.09	0.57	196	3.27	0.61	-2.92	373	<.05	0.16

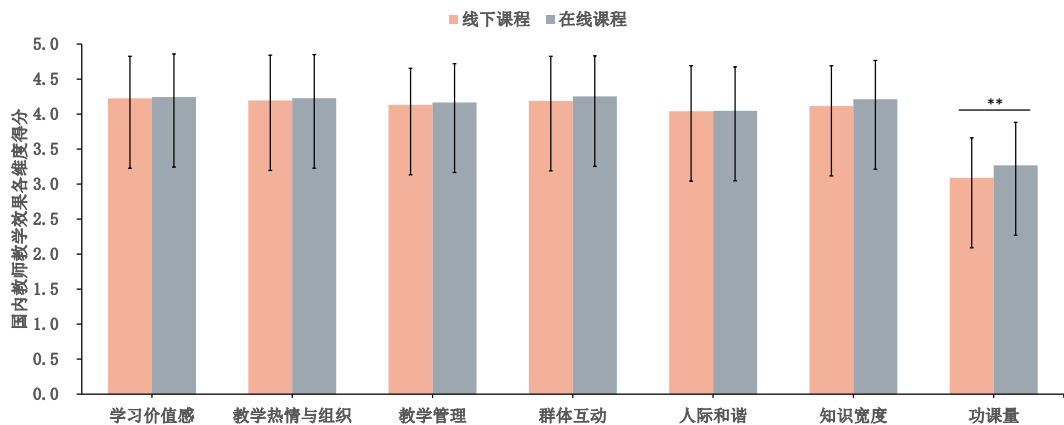


图 6.2 国内教师线下课程及在线课程的各维度平均分及标准差

表 6.7 功课量维度各项目平均分及标准差

功课量	线下课程		在线课程	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
课程难度	2.99	0.84	3.34	0.82
课程进度	3.17	0.76	3.28	0.78
作业量	3.11	0.81	3.19	0.83

6.4 讨论

本研究基于《大学教师教学效果评价问卷》调查结果发现，国内教师的在线课程及线下课程教学效果得分不存在显著差异，这与多层次增长模型中的教师的教学效能在线和线下不存在显著差异是一致的(Carle, 2009)。不同性别学生对国内教师线下课程及在线课程的主效应显著，男生与女生给教师在线课程的得分高于线下课程得分，相比于线下课程，国内教师的在线课程教学效能要好于线下课程，而且针对不同年级学生对教师不同课程的统计发现，大一、大二、大三的学生给教师在线课程的得分低于线下课程的得分，大四、研一、研二、研三的学生给予教师在线课程的得分高于线下课程的得分，从这里可以看出，低年级的学生倾向于线下课程，给予教师线下课程更高的得分，而高年级学生偏爱在线课程，给予教师在线课程更高的评分。

通过对国内教师教学效能各维度的分析发现，国内教师在功课量及难度这一维度上存在显著差异。通过功课量各项目分析发现，在线课程的课程难度高于线下课程、课程进度快于线下课程、作业量多于线下课程，因此，教师在两种不同课程的教学上存在一定差异的，以往的研究发现，教师在线下课程的团体互动，软件使用经验、教师指导和模拟工具易用性方面会影响的学生的学习，而在线上课程，只有教师指导和工具易用性会影响到学生的学习(Riley, Ellegood, Solomon, & Baker, 2017)。因此未来的研究可以关注两种不同课程的在具体教学行为上的差异。

教师的职称与课程类型不存在显著差异，只有讲师在线下课程及在线课程上的得分存在差异，说明教学经验丰富的副教授及教授在两种不同的上课形式上依然能保持一致的教学水平，相比教学经验较少的讲师来说，通过增加在线课程互动方式，例如弹幕、随堂提问等，可以增强与学生的交流，提高学生学习的满意度，有研究表明，教师的互动与学生的在线课程的满意度呈显著正相关($r = 0.20, p < .05$) (Horzum, 2015)，教师想要提高学生的在线课程的参与，清晰的授课、充满激情，有同伴互动，充分利用课程资源等是关键因素(Hew, 2016)。

本研究仅仅使用了学生评价教师的数据，未加入学生的期末成绩，因此对于在线课程及线下的哪种课程形式更好存在不足，未来的研究也可以考虑增加学生的成绩作为研究内容。不过，学生在线下课程及线下课程的学习结果是一样的(Lou, Bernard, & Abrami, 2006; Means et al., 2009)，在线学生的学习表现和线下学生一样更好(Tucker, 2001)，所以两种课程下学生的学习哪种更好是未来探索的方向。

综上所述，国内教师的在线课程及线下课程的教学效能整体上基本一致的，但是在线课程的教学效能得分要高于线下课程。从教学效能的各维度的分析发现，国内教师的在线课程的功课量显著高于线下课程，说明在线课程的课程难度高于线下课程、课程进度快于线下课程、作业量多于线下课程，因此，教师在线授课时需要了解学生的学习进度，适当降低一定的学习难度和进度以帮助学生。

第7章 总讨论

7.1 主要研究发现

本文基于教学效果内隐理论,利用大数据技术来探讨教师教学效能与教龄、终身制度等因素的作用机制,为多层次增长模型提供大样本的支持,同时基于国内调查研究,探讨了教师线下课程及在线课程的教学效能差异,本研究主要发现有:

第一,教师的教学效能综合得分与教龄呈正相关,但是相关关系非常小,建议忽略,所以随着教学年龄的增长,教学效能保持相对稳定,因此假设一不成立,该结论支持了多层次增长模型,即教授的教学效能评分相对比较稳定(Marsh, 2007; Carle, 2009);

第二,助理教授、副教授、教授之间的教学效能综合得分尽管存在显著差异,但效应量较少,建议忽略,因此获得终身职位的教授的教学效能与非终身教授没有差异,但是教授的教学效能高于副教授及助理教授。本结论支持了多层次增长模型中终身制没有影响教师的教学效能这一观点(Carle, 2009),本研究也发现,教授的教学效能得分要高于助理教授和副教授,因此,假设二是部分成立的;

第三,基于 RMP 探索发现,线下课程的教学效能综合得分要高于在线课程,而在线课程的课程难度普遍高于线下课程。为了学分而选择课程、获得比较好的期末成绩及课程要求强制出勤,学生往往给予这些课程更高的评价;

第四,国内教师的在线课程及线下课程的教学效果得分整体上基本一致的,并无显著差异,所以在线课程及线下课程的教学效能不存在差异,假设三不成立,本结论与多层次增长模型的中,教师的在线、线下的课程的教学效能的评价是一致的(Carle, 2009),但从教学效果各维度的分析结果可知,国内教师的在线课程的功课量得分显著高于线下课程。

7.2 研究贡献及启示

7.2.1 理论意义

基于以上发现,本研究支持了多层次增长模型,即教授的教学效能评分相对比较稳定,教学效能随着教龄的增加,教师的教学效能并没有太大变化(Marsh, 2007; Carle, 2009),而且在线、线下的课程的教学效能的评价在总体上一致的,终身教授的教学效能得分显著提高了(Carle, 2009),这为多层次增长提供了大样本的数据支持,弥补了该模型调查研究中的小样本的不足;其次,也会一定程度上支持了教学效果内隐理论,即教师的教学行为会影响学生对教学效能的总体评价,例如教授不同难度的课程、给予学生的不同的成绩、强制要求出勤会影响到学生对教师的评价;最后也为增值性评价模型提供学生评价内容提供一定的启发,因为该评估模型以学生的学生成绩作为教学效能的衡量指标,尽管已经加入了学生的历史学习成绩、学校因素、同事因素、家庭等因素(Hanushek & Rivkin, 2010),但并不是对每个教师的评价都能使用学生的

成绩,有三分之二的教师是无法进行增值性评价(Papay, 2012),多数教师还是需要通过传统的观察课堂表现和学生问卷调研来进行评估(Bacher-Hicks, Chin, Kane, & Staiger, 2017),在未来可以尝试增加学生评教内容作为教学效能的评价指标,不断拓展增值性评价模型应用场景。

7.2.2 现实意义

RMP 作为一个公开在线评教平台,数据丰富且应用广泛,学生普遍认为 RMP 上的内容是可信且有效的,因此通过 RMP 查看教师的授课信息,并根据自己的需求来选择课程(Hayes & Prus, 2014)。本研究具有一定的应用价值:

第一,本文为了解教师在职业生涯中的教学效能的变化趋势提供一定的参考依据。本文基于 RMP,利用大数据技术,探讨了教师教学年龄与教学效能之间的关系,为教学效能的多层次增长模型提供大样本的数据支持,有利于帮助学生、教师、学校了解教师的教学效能的发展趋势。教师教学效能随着年龄增长是保持不变(Marsh, 2007; Carle, 2009),因此单从教龄上区分教师是否具有较高的教学效能是不太实际的,这启示学生在选择教师时应减少对教师的教龄的关注。对于教师而言,教师如果想要促进专业发展、提高教学效能,应在学校环境下通过教学实践或者学生的反馈来不断改善教学策略(Postholm, 2012),达到提高教学效能的目的。

第二,本文为提高教学效能提供一定的指导。本文基于 RMP 进行探索,发现了高教学效能的教师的一些特点,例如教师要关注学生的成长,及时给予学生反馈,同时要尊重学生;其次,在课程设置上,课程安排要符合学生的需求,如果课程难度较大建议适当降低一定的难度;还有,根据学生的需要多设置学分课程,增加学生选课的概率;最后,一般给予学生的期末成绩可以大方些。以往的研究也发现,学生喜欢选择在 RMP 上得到积极评价的教授,因为学生在课程结束后容易获得高分(Reber, Ridge, & Downs, 2017)。通过这种教学评价,可以改善教师教学和学生的学习(Umar et al., 2016)。

第三,本文为终身制度的实施标准提供一定的参考依据。本文基于 RMP,探讨教师教学效能及终身职位之间的关系,发现获得终身职位与未获得终身职位的教授之间教学效能是不存在差异的,因此,高校将学生评价纳入授予终身职位的指标体系中时需要慎重考虑,目前在美国已经不少州将教授的绩效作为授予终身的依据,教师的教学效能是教师绩效的其中一环,尽管终身制度对于教学效能无影响,教授们也应该努力提高自己的教学效能。需要主要的是, RMP 是一个公开评价平台,学生的评价可能存在偏见(Clayson, 2014; Legg & Wilson, 2012),因此可在学校层面建立教学评价共享平台,教师可以及时了解学生对学习的反馈,从而督促教师提高教学效能。

第四,本文为学生选择在线课程及线下课程提供一定的参考。目前基于 RMP 及国内调查研究发现,教师的在线课程难度普遍高于在线课程的难度,因此从课程难度来说,学生对于难度较高的课程往往再次选择的可能性是比较低的。学生在选课时会考虑课程的难度,因此,一般较少选择难度大的课程(Babad, Darley, & Kaplowitz, 1999),但如果课程其他方面评价较好或

者课程很有价值时, 学生也会选择这些难度大的课程(Coleman & McKeachie, 1981)。所以如果教师想要提高学生再次选课的比例, 建议适度降低一定的课程难度。学生可以根据公开平台获取教师的教学评价, 但如果学生在线下课程及线下课程的学习结果是一样的(Lou, Bernard, & Abrami, 2006; Means et al., 2009), 那么选择哪一种都可以, 不过也可以根据其他方面的需要选择适合自己的教师及课程。

第五, 为国内建设公开评价教师教学的平台提供一些启发。国内目前还缺少公开评价教师教学的平台, 尽管国内已经存在在线评教网站有学邦网(www.xuebang.com.cn)、导师评价网(www.mysupervisor.org), 但是学邦网已经停止运营、导师评价网是导师的学术评价而不是教学评价, 因此对于学校来说, 可以考虑建立学校内部的公开评教平台。这也对目前的教育机构建设公开教师评价平台提供也一些启发, 因为学生倾向于给老师反馈, 期望教师不断提高教学效能(Chen & Hoshower, 2003), 因此建立一个公开的评教系统对于帮助教师提高教学效能很有必要。

7.3 研究不足

本研究的局限性是, 第一, 本研究从学生的角度对教师的教学效能进行研究, 探讨在终身制度下, 教授的教学效能的变化, 未考虑教师的其他教学服务及学术研究和贡献, 也未考虑教师的自我评价、同行评价等重要内容; 第二, 学生评价存在一定偏见, 教师可能会用给学生更高的期末成绩来换取学生给自己的打高分。有研究表明, RMP 的问卷本身存在偏见, RMP 的评价问卷是一种态度量表, 而不是教学绩效量表(Clayson, 2014; Legg & Wilson, 2012)。虽然 RMP 涵盖了许多学生对教授的评价, 但他们仍然遗漏了很大一部分学生。这种情况发生的原因是, 与大学的内部调查相比, 在线评级并不是一种强制要求。教师只有在课程上传达教学评价的重要性时, 学生们才会提高评教的概率(Young, Joines, Standish, & Gallagher, 2019)。另外, 由于评分网站存在很多, 学生们不会每个都使用。在线评级也受到光环效应等偏见的影响(Otto, Sanford, & Ross, 2008), 例如光环效应导致人们根据教授的个性和吸引力而不是教学质量来给教授打分, 这种偏见会降低对教授评分的准确性, 而不同种族及性别的教师的得分也会受到偏见影响(Chisadza, Nicholls, & Yitbarek, 2019), 甚至有研究者指出, 学生评教与学生的学习无相关关系(Uttl, White, & Gonzalez, 2017), 建议谨慎使用这类学生评价综合评价分数(Feistauer & Richter, 2017)。第三, 本文基于 RMP 进行大数据分析, 大数据的优势之一是数量庞大, 但是劣势是数据质量及准确性存在问题。

7.4 未来展望

首先，未来的研究可从学术成就、师生关系等方面多维度评估教师，从领英获取教授的职业数据，从 [researchgate.com](https://www.researchgate.com)、谷歌学术等网站获取教授的学术发表情况，丰富教师评价维度；其次，对于教师的教学评价还可以增加自我评价、同行评价等内容。

本研究中已采用了 Word Cloud 来可视化数据，未来通过使用更高级的文本挖掘工具从学生的评论中提取学生关于教师的情感表达，以了解教师的教学表现。涉及到大数据数据管理可以建立数据质量评价体系，从完整性(数据完整，无缺失)；一致性(数据规范统一)；准确性(数据准确，无异常或者错误信息)等方面确保数据质量。

参考文献

- 边玉芳 & 孙丽萍. (2015). 教师增值性评价的进展及在我国应用的建议. *教师教育研究*, 27(1), 88-112.
- 陈琦 & 刘儒德. (2011). *教育心理学(第2版)*. 高等教育出版社.
- 孟庆茂 & 刘红云. (2003). 大学教师教学效果评价维度结构及影响因素研究. *心理科学*, 26(4), 617-619.
- 喻丰, 彭凯平, & 郑先隽. (2015). 大数据背景下的心理学: 中国心理学的学科体系重构及特征. *科学通报*, 60(5-6), 520-533.
- 朱廷劭. (2016). *大数据时代的心理学研究及应用*. 科学出版社.
- Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51(3), 249-269.
- Babad, E., Darley, J. M., & Kaplowitz, H. (1999). Developmental aspects in students' course selection. *Journal of Educational Psychology*, 91(1), 157-168.
- Bacher-Hicks, A., Chin, M., Kane, T., & Staiger, D. (2017). An evaluation of bias in three measures of teacher quality: Value-added, classroom observations, and student surveys. *National Bureau of Economic Research*, 121, 1-47.
- Banasik, M. J. D., & Dean, J. L. (2016). Non-Tenure track faculty and learning communities: Bridging the divide to enhance teaching quality. *Innovative Higher Education*, 41(4), 333-342.
- Bleske-rechek, A., & Michels, K. (2010). RateMyProfessors.com : Testing assumptions. *Practical Assessment Research & Evaluation*, 15(5), 1-12.
- Bonds-Raacke, J., & Raacke, J. (2007). The relationship between physical attractiveness of professors and students' ratings of professor quality. *Journal of Psychiatry, Psychology and Mental Health*, 1(2), 1-7.
- Boyer, E. L. (1996). The scholarship of engagement. *Bulletin of the American Academy of Arts and Sciences*, 49(7), 18-33.
- Brown, M. J., Baillie, M., & Fraser, S. (2009). Rating ratemyprofessors.com: A comparison of online and official student evaluations of teaching. *College Teaching*, 57(2), 89-92.
- Burns, S. M., & Ludlow, L. H. (2005). Understanding student evaluations of teaching quality: The contributions of class attendance. *Journal of Personnel Evaluation in Education*, 18(2), 127-138.
- Carle, A. C. (2009). Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach. *Computers and Education*, 53(2), 429-435.

- Carini, Hayek, Kuh, Kennedy, & Ouimet. (2003). College student responses to web and paper surveys: does mode matter? *Research in Higher Education*, 44(1), 1–19.
- Chen, E. E., & Wojcik, S. P. (2016). A practical guide to Big Data research in psychology. *Psychological Methods*, 21(4), 458–474.
- Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28(1), 71–88.
- Cheng, D. A. (2015). Effects of professorial tenure on undergraduate ratings of teaching performance. *Education Economics*, 23(3), 338–357.
- Chisadza, C., Nicholls, N., & Yitbarek, E. (2019). Race and gender biases in student evaluations of teachers. *Economics Letters*, 179, 66–71.
- Chitre, M., & Srinivasan, D. (2018). Evaluating teaching effectiveness using quantitative student feedback. In *Proceedings of 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2018, December*, 155–160.
- Clayson, D. E. (2014). What does ratemyprofessors.com actually rate? *Assessment & Evaluation in Higher Education*, 39(6), 678–698.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*(2nd ed.). Lawrence Erlbaum Associates. New York, the United States.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–149.
- Coladarci, T., & Kornfield, I. (2007). Ratemyprofessors.com versus formal in-class student evaluations of teaching. In *Practical Assessment, Research and Evaluation*, 12(6), 1–15.
- Coleman, J., & McKeachie, W. J. (1981). Effects of instructor/course evaluations on student course selection. *Journal of Educational Psychology*, 73(2), 224–226.
- Constand, R. L., Clarke, N., & Morgan, M. (2018). An analysis of the relationships between management faculty teaching ratings and characteristics of the classes they teach. *International Journal of Management Education*, 16(2), 166–179.
- Cornell University. (2020, January 27). Associate professor. Retrieved January 27, 2020, from <http://theuniversityfaculty.cornell.edu/the-newp-faculty-handbook/3-titles-and-appointments-leaves/3-1-titles/associate-professor/>.
- Davison, Elizabeth, & Price. (2009). How do we rate an evaluation of online student evaluations. *Assessment & Evaluation in Higher Education*, 34(1), 51–65.
- Driscoll A., Jicha K., Hunt A. N., Tichavsky L., & Thompson G.. (2012). Can online courses deliver In-class results?. *Teaching Sociology*, 40, 312–331.

- Feistauer, D., & Richter, T. (2017). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education*, 42(8), 1263–1279.
- Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view : matched or mismatched priorities ? *Research in Higher Education*, 28(4), 291–344.
- Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness and other issues: Student evaluations of professors on Ratemyprofessors.com. *Assessment & Evaluation in Higher Education*, 33(1), 45–61.
- Figlio, D. N., Schapiro, M. O., & Soter, K. B.. (2015). Are tenure track professors better teachers? *Social Science Electronic Publishing*, 97(3), 715-724.
- Gentry, R., & Stokes, D. (2015). Strategies for professors who service the university to earn tenure and promotion. *Research in Higher Education Journal*, 29, 1–13.
- Gonzalez, C., Liu, Y., & Shu, X. (2012). The faculty promotion and merit system in China and the United States: the cases of Wuhan university and the University of California, Davis. *Research & Occasional Paper Series:CSHE.13.12*, 1–11.
- Good, T. L. (1979). Teacher effectiveness in the elementary school. *Journal of Teacher Education*, 30(2), 52–64.
- Gorsky, P., & Blau, I. (2009). Online teaching effectiveness: A tale of two instructors. *International Review of Research in Open and Distance Learning*, 10(3), 1–27.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2), 267–271.
- Harrison, P. D., Ryan, J. M., & Moore, P. S. (1996). College students' self-Insight and common implicit theories in ratings of teaching effectiveness. *Journal of Educational Psychology*, 88(4), 775–782.
- Hayes, M. W., & Prus, J. (2014). Student use of quantitative and qualitative information on ratemyprofessors.com for course selection. *College Student Journal*, 48(4), 675–688.
- Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, 47(2), 320–341.
- Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations: What all faculty should know. *College Teaching*, 49(1), 26–31.
- Horzum, M. B. (2015). Interaction, structure, social presence, and satisfaction in online learning. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(3), 505–512.

- Ismail, Rahida. Aini. M., Arshad, R., & Abas, Z.(2018). Can teachers' age and experience influence teacher effectiveness in HOTS? *International Journal of Advanced Studies in Social Science & Innovation*,2(1),144-158.
- Jahangiri, L., Mucciolo, T. W., Choi, M., & Spielman, A. I. (2008). Assessment of teaching effectiveness in U.S. dental schools and the value of triangulation. *Journal of Dental Education*, 72(6), 707–718.
- Kasten, K. L. (1984). Tenure and merit pay as rewards for research, teaching, and service at a research university. *The Journal of Higher Education*, 55(4), 500–514.
- Kay M. & Wobbrock J. (2019). ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs. Retrieved January 27, 2020, from <https://github.com/mjskay/ARTool>.
- Kini, T., & Podolsky, A. (2016). Does teaching experience increase teacher effectiveness? A review of Research. *Learning Policy Institute*, 1–64.
- Kim, N., Yu, X., & Schwartz, Z. (2013). Can online surveys substitute traditional modes? An error-based comparison of online and on-site tourism destination surveys. *Tourism Review International*, 17(1), 31–45.
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the tennessee value added assessment system. *Educational Evaluation and Policy Analysis*, 25(3), 287–298.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group*, 70–73.
- Larson, J. R. (1979). The limited utility of factor analytic techniques for the study of implicit theories in student ratings of teacher behavior. *American Educational Research Journal*, 16(2), 201–211.
- Le, T. T. K., Tran, T. T. B., Ho, H. T. M., Vu, A. T. L., & Lopata, A. L. (2018). Prevalence of food allergy in Vietnam: Comparison of web-based with traditional paper-based survey. *World Allergy Organization Journal*, 11(1), 1–10.
- Legg, A. M., & Wilson, J. H. (2012). RateMyProfessors.com offers biased evaluations. *Assessment & Evaluation in Higher Education*, 37(1), 89–97.
- Lou, Y., Bernard, R. M., & Abrami, P. C. (2006). Media and pedagogy in undergraduate distance education: A theory-based meta-analysis of empirical literature. *Educational Technology Research and Development*, 54(2), 141–176.
- Lutter, M., & Schröder, M. (2016). Who becomes a tenured professor, and why? Panel data evidence from German sociology, 1980-2013. *Research Policy*, 45(5), 999–1013.

- Mariano L. T. , D. F. McCaffrey, and J. Lockwood.(2010).A model for teacher effects from longitudinal data without assuming vertical scaling. *Journal of Educational and Behavioral Statistics*. 35(3),253-279.
- Marsh, H. W. (1982). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, 74(2), 264–279.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707–754.
- Marsh, H. W., & Hocevar, D. (1991). Students' evaluations of teaching effectiveness : The stability of mean ratings of the same teachers over a 13-year period. *Teaching & Teacher Education*, 7(4), 303–314.
- Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, 99(4), 775–790.
- Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured Intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30(1), 217–251.
- Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92(1), 202–228.
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC. US Department of Education.
- Mart, C. T. (2017). Student evaluations of teaching effectiveness in higher education. *International Journal of Academic Research in Business and Social Sciences*, 7(10), 57-61.
- Matz, S. C., & Netzer, O. (2017). Using Big Data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences*, 18, 7–12.
- Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, 12(1), 53–74.
- O'Neill, D. K., & Sai, T. H. (2014). Why not? Examining college students' reasons for avoiding an online course. *Higher Education*, 68(1), 1-14.
- Otto, J., Sanford, D. A., & Ross, D. N. (2008). Does ratemyprofessor.com really rate my professor? *Assessment & Evaluation in Higher Education*, 33(4). 355–368.

- Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123–141.
- Postholm, M. B. (2012). Teachers' professional development: A theoretical review. *Educational Research*, 54(4), 405–429.
- Radmacher, S. A., & Martin, D. J. (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *Journal of Psychology: Interdisciplinary and Applied*, 135(3), 259–268.
- RateMyProfessors.com. (2020, January 27). About RateMyProfessors.com. Retrieved January 27, 2020, from <https://www.ratemyprofessors.com/About.jsp>.
- Reber, J. S., Ridge, R. D., & Downs, S. D. (2017). Perceptual and behavioral effects of expectations formed by exposure to positive or negative Ratemyprofessors.com evaluations. *Cogent Psychology*, 4(1), 1–16.
- Reevy, G. M., & Deason, G. (2014). Predictors of depression, stress, and anxiety among non-tenure track faculty. *Frontiers in Psychology*, 5, 1–17.
- Richardson, John T. E. (2016). Face-to-face versus online tutorial support in distance education: preference, performance, and pass rates in students with disabilities. *Journal of Postsecondary Education & Disability*, 29, 83-90.
- Riley, J. M., Ellegood, W. A., Solomon, S., & Baker, J. (2017). How mode of delivery affects comprehension of an operations management simulation online vs face-to-face classrooms. *Journal of International Education in Business*, 10(2), 183–200.
- Robert D. R., & Harry G. M.. (1996). Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education*, 37(3), 223-240.
- Rothgeb, J. M. (2014). When tenure protects the incompetent: Results from a survey of department chairs. *Political Science and Politics*, 47(1), 182–187.
- Sanders, W. L., & Horn, S. P. (1994). The tennessee value-added assessment system (TVAAS): mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(3), 299–311.
- Sav, G. T. (2016). Are american universities mismanaged?: Tenure vs non-tenure faculty employment decisions. *Information Management and Business Review*, 8(6), 19-31.
- Seiler, M. J., Seiler, V. L., & Chiang, D. (1999). Professor , student , and course attributes that contribute to successfui teaching evaluations. *Financial Practice and Education*, 91–99.
- Sass, T. R., Semykina, A., & Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, 38, 9–23.

- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. NY: Oxford University Press.
- Sonntag, M. E., Bassett, J. F., & Snyder, T. (2009). An empirical test of the validity of student evaluations of teaching made on RateMyProfessors.com. *Assessment & Evaluation in Higher Education*, 34(5), 499–504.
- Stella, A., & Gnanam, A. (2004). Quality assurance in distance education: The challenges to be addressed. *Higher Education*, 47 (2), 143–160.
- Swiggard, J. C., & Seidel, S. (2012). An evaluation of student conformity when using professor rating websites. *Journal of Young Investigators*, 24(4), 20–25.
- Tang, T. L. P. (1997). Teaching evaluation at a public institution of higher education: Factors related to the overall teaching effectiveness. *Public Personnel Management*, 26(3), 379–389.
- Theyson, K. C. (2015). Hot or not: The role of instructor quality and gender on the formation of positive illusions among students using RateMyProfessors.com. *Practical Assessment, Research and Evaluation*, 20(4), 1–12.
- Thomsen Jennifer (2014, May 1). 50-States comparison: Teacher tenure continuing contract policies. Education Commision of the States. Retrieved January 27, 2020, from <https://www.ecs.org/teacher-tenure-continuing-contract-policies/>.
- Tucker S..(2001). Distance education: Better, worse, or as good as traditional education? *Online Journal of Distance Learning Administration*, 4(4). 1556-3847.
- Umar, M. A., Ahmad, B. I., Kufena, A. M., Abdulsalami, A. O., Tenuche, S. S., Sahabi, Y. A., & Ahmad, U. M. (2016). Students' perception of online student evaluation of teaching (SET) in Nigeria. *CEUR Workshop Proceedings*, 1830, 23–27.
- University of Houston. (2020, January 27). Ranks of professors and the path to becoming a professor. Retrieved January 27, 2020, from <https://www.math.uh.edu/~tomforde/Ranks-of-Professors.html>.
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42.
- Winters, M. A., & Cowen, J. M. (2013). Would a value-added system of retention improve the distribution of teacher quality? A simulation of alternative policies. *Journal of Policy Analysis and Management*, 32(3), 1–21.
- Whitely, S. E., & Doyle, K. O. (1976). Implicit theories in student ratings. *American Educational Research Journal*, 13(4), 241–253.

- Wobbrock J., Findlater L., Gergle D., & Higgins J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)*, Vancouver, British Columbia (May 7-12, 2011). New York: ACM Press, 143-146.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57–67.
- Young, K., Joines, J., Standish, T., & Gallagher, V. (2019). Student evaluations of teaching: the impact of faculty procedures on response rates. *Assessment & Evaluation in Higher Education*, 44(1), 37–49.
- Yueh, H. P., Chen, T. L., Chiu, L. A., Lee, S. L., & Wang, A. B. (2012). Student evaluation of teaching effectiveness of a nationwide innovative education program on image display technology. *IEEE Transactions on Education*, 55(3), 365–369.
- Zafer Ünal & Aslihan Ünal. (2012). The impact of years of teaching experience on the classroom management approaches of elementary school teachers. *International Journal of Instruction*, 5(2), 1308-1470.

附录 A 大学教师教学效果评价问卷

同学您好，我们正在进行一项关于教师教学效果的问卷调查。本问卷目的是为了评价教师教学效果，请回忆最近已完成的一门课程，根据自己的观察和体验对教师教学做出评定。本问卷不记名，所有答案均无对错之分，所有数据仅用于统计分析和学术研究，请您根据实际情况填写。

1.性别

男 女

2.年龄

3.年级

大一 大二 大三 大四 研一 研二 研三

4.学科性质

文科 理科

请回忆最近已经完成的课程，并填写教师相关信息

5.授课教师性别

男 女

6.授课教师职位

讲师 副教授 教授

7.课程类型

线下课程 线上课程

请回忆最近已完成的一门课程，根据自己的观察和体验对教师教学效果做出评价。

1.学会并理解教师所讲授的学科内容

完全反对 不同意 一般 比较同意 非常同意

2.教师的讲授方式让学生在课堂上保持兴趣

完全反对 不同意 一般 比较同意 非常同意

3.教师的讲授有助于学生作笔记

完全反对 不同意 一般 比较同意 非常同意

4.教师鼓励学生表达自己的观点和对教师提出质疑

完全反对 不同意 一般 比较同意 非常同意

5.在办公时间内或课后，教师与学生有充分的接触

完全反对 不同意 一般 比较同意 非常同意

6.教师充分讨论本学科领域目前的发展趋势

完全反对 不同意 一般 比较同意 非常同意

7.考试或评定成绩的作业,能够检验教师所强调的教学内容

完全反对 不同意 一般 比较同意 非常同意

8.适当时候,除自己的观点外,教师也讲授他人的各种观点

完全反对 不同意 一般 比较同意 非常同意

9.教师真诚地关心每个学生

完全反对 不同意 一般 比较同意 非常同意

10.教师鼓励学生提出问题且给予学生有意义的回答

完全反对 不同意 一般 比较同意 非常同意

11.预定的教学目标应该和实际的讲授结合,学生了解课程的进展情况

完全反对 不同意 一般 比较同意 非常同意

12.教师在讲课中具有幽默感

完全反对 不同意 一般 比较同意 非常同意

13.我觉得上了这门课程,我对这门课的兴趣提高了

完全反对 不同意 一般 比较同意 非常同意

14.在这门课里学到了有价值的东西

完全反对 不同意 一般 比较同意 非常同意

15.教师在教这门课时充满活力与精力

完全反对 不同意 一般 比较同意 非常同意

16.教师对课程内容作精心准备与仔细解释

完全反对 不同意 一般 比较同意 非常同意

17.教师邀请学生发表及共享他们的见解和知识

完全反对 不同意 一般 比较同意 非常同意

18.无论课堂内外,教师欢迎学生向他(她)寻求帮助或建议

完全反对 不同意 一般 比较同意 非常同意

19.教师阐述课程中所用观念(或概念)的背景或来源

完全反对 不同意 一般 比较同意 非常同意

20.教师用来评价学生的方法公平恰当

完全反对 不同意 一般 比较同意 非常同意

21.教师指定的阅读材料,布置的作业等,有助于学生对课程的理解和掌握

完全反对 不同意 一般 比较同意 非常同意

22.教师对每个学生都很友善

完全反对 不同意 一般 比较同意 非常同意

23.教师讲课中比较各种理论或方法的含义

完全反对 不同意 一般 比较同意 非常同意

24.教师在考试或评分作业上所给的反馈有价值

完全反对 不同意 一般 比较同意 非常同意

25.教师所指定的阅读材料或参考资料很有价值

完全反对 不同意 一般 比较同意 非常同意

26.教师讲课在智力上激发学生，有启发和激励性

完全反对 不同意 一般 比较同意 非常同意

27.教师对教学很热情

完全反对 不同意 一般 比较同意 非常同意

28.教师解释很清楚

完全反对 不同意 一般 比较同意 非常同意

29.教师鼓励学生参与课堂讨论

完全反对 不同意 一般 比较同意 非常同意

30.为促进教学，教师所教课程的难度是

非常容易 比较容易 一般 比较困难 非常困难

31.为促进教学，教师所教课程的进度是

非常慢 比较慢 一般 比较快 非常快

32.为促进教学，教师所教课程的作业量是

非常少 比较少 一般 比较多 非常多

致谢

时光如梭，转眼间，两年的研究生生活即将结束，站在毕业的门槛上，回首过往，奋斗和努力回荡在脑海中，甜美与欢笑也都尘埃落定。自强不息、厚德载物，清华大学以其优良的学风、严谨的科研氛围教我求学，以其博大包容的情怀胸襟育我成人。值此毕业论文完成之际，我谨向所有关心、帮助我的人们表示最诚挚的感谢与最美好的祝愿。

本论文是在导师何吉波副教授的悉心指导之下完成的。导师不仅授我以文，而且教我做人，虽历时两载，却赋予我终生受益无穷之道。本论文从选题到完成，几易其稿，每一步都是在导师的指导下完成的，倾注了导师大量的心血，在此我向我的导师何吉波副教授表示深切的谢意与祝福！

本论文的完成也离不开其他各位老师、同学和朋友的关心与帮助。尤其是在开题答辩时，张丹老师、廖江群老师及王非老师提出了宝贵的修改建议，还要感谢王胤丞、梁琮丹等人的帮助，在科研过程中给我以许多鼓励。回想整个论文的写作过程，虽有不易，却让我除却浮躁，经历了思考和启示，也更加深切地体会了心理学的精髓和意义，因此倍感珍惜。

还要感谢父母在我求学生涯中给与我无微不至的关怀和照顾，一如既往地支持我、鼓励我。同时，还要感谢我的室友，张永发同学、朱奥同学、覃康文同学、黄阳同学，愿友谊长存！

声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行的研究工作及取得的研究成果。尽我所知，除了文中已注明引用的内容外，本学位论文中不包含其他人享有著作权的内容。对本论文所涉及的研究工作做出的贡献的其他个人和集体，均已在文中以明确方式标明。

研究生签名：_____日期：_____

个人简历、在学期间发表的学术论文及研究成果

个人简历

1993年1月出生于山东省寿光市。

2016年7月毕业于济南大学教育与心理科学学院心理学系。

2018年9月进入清华大学社会科学学院心理学系攻读硕士至今。

学术成果

1. Jibo He^{1*}, Mingchun Liu¹, Yu Du², Guanlin Liu³, Jingmeng Cui⁴, Yanjie Su^{4*}. (submitted)
How do Professors Teach? An Analysis of Student Comments on almost One Million Professors in RateMyProfessors.com. *Journal of Educational Psychology*.