·教育研究·

# 课程论文评分中的主题偏倚分析方法研究

张孟科[1,2]　宋伟[1,2]　李丹[1,2]　沈秉正[1,2]　陈莹[1,2]　王宗春[1,2]　郭咸希[1,2]
[1]武汉大学人民医院药学部，武汉　430060；[2]武汉大学第一临床学院，武汉　430060
通信作者：张孟科，Email：zhmke@126.com

【摘要】 课程论文评分往往主观性较强，教师在对课程论文进行评分时可能因对特定主题的论文存在偏倚而在评分中引入偏差，导致论文评分不能准确反映学生的真实水平。本文以《临床药理学》课程论文评分结果为实例，对检验课程论文评分结果中是否包含主题偏倚的分析方法进行了探讨。通过计算纵向偏倚因子和横向偏倚因子，比较了同一教师对不同主题的论文和不同教师对同一主题的论文评分上的差异，构建了一种科学、易行的分析方法。该方法可帮助教师快速发现课程论文评分中是否隐含相关偏差，从而使课程论文评分更加客观准确。

【关键词】 课程论文；　评分；　主题偏倚；　偏倚因子
【中图分类号】 R-05
基金项目：武汉大学医学部教学改革研究项目（2019023、2022ZG187）
DOI: 10.3760/cma.j.cn116021-20220107-01443

**An analytical method for topic bias in course paper grading**

*Zhang Mengke[1,2], Song Wei[1,2], Li Dan[1,2], Shen Bingzheng[1,2], Chen Ying[1,2], Wang Zongchun[1,2], Guo Xianxi[1,2]*
*[1] Department of Pharmacy, Renmin Hospital of Wuhan University, Wuhan 430060, China; [2] The First Clinical School of Wuhan University, Wuhan 430060, China*
*Corresponding author: Zhang Mengke, Email: zhmke@126.com*

【Abstract】 Course paper grading often involves subjective factors. Teachers may introduce biases into the grading due to their preferences for certain topics, resulting in inaccurate grading results that fail to reflect the true abilities of students. Taking the grading results of the "Clinical Pharmacology" course as an example, this article investigates the analytical method for detecting topic bias in the course paper grading. A comparative analysis was performed on the differences in the scores between different topics graded by the same teacher and between the same topic graded by different teachers by calculating the vertical bias factor and the horizontal bias factor, and a scientific and feasible analytical method was established. This method can help teachers quickly discover biases in their course paper grading, thereby making the grading more objective and accurate.

【Key words】 Course paper;　Grading;　Topic bias;　Bias factor
**Fund program:** Teaching Reform Foundation of Wuhan University School of Medicine (2019023, 2022ZG187)
DOI: 10.3760/cma.j.cn116021-20220107-01443

课程考核作为教学的重要环节，不仅是检验学生学习效果、促进学习的重要手段，同时也是教师自我检查教学效果、分析和总结教学经验与不足、不断改进并完善教学的重要途径[1]。课程论文是学生在学习了一门课程之后，根据学术论文写作要求提交的探讨课程相关学术问题的文章。课程论文作为一种重要的考核形式，在不同层次高等教育的教学中均有广泛应用。课程论文的考核有利于激发学生的学习热情，促进学生积极主动地对学科专业的某一问题进行深入的探索，从而对相关领域的发展有一系统、全面的了解；同时，通过完整的课程论文写作训练，有助于提高学生的文献检索和论文写作能力，为后续的学术论文和学位论文撰写打下基础。尽管这一考核形式具有诸多优点，但课程论

文评价主观性太强，评分标准难以量化，评分过程难以控制，导致评分结果不能准确反映学生对课程的掌握情况。

研究者查阅文献发现，已有学者对课程论文这种考核形式的优缺点，以及如何利用课程论文考核形式的优点来提高教学质量进行了研究和论述[2-5]，但对于如何使课程论文评分过程更加客观准确的研究较少，且尚无文献报道可用于检验因评分教师对不同主题的论文有偏倚而导致评分出现偏差的分析方法。本文将以武汉大学 2020 年《临床药理学》课程论文评分结果为实例，对课程论文评分结果分析方法进行探讨，以期建立一种科学、易行的分析方法，帮助教师快速发现课程论文评分中是否隐含此类偏差，从而使课程论文评分更加客观、准确。

## 1　研究背景

2020 年武汉大学共有 42 位临床医学本科生选修《临床药理学》。课程结束后以自主选题撰写课程论文的形式进行了考核，五位授课教师分别对 42 篇课程论文进行了独立评分。本文将以 5 位教师对课程论文的评分为研究对象，对检验课程论文评分中是否存在主题偏倚的分析方法进行探讨。

## 2　研究方法

### 2.1　课程论文评分汇总与分组

为了进行后面的主题偏倚分析，首先将 5 位评分教师对 42 名学生的课程论文评分结果进行汇总，然后根据课程论文的主题对论文评分进行分组，将主题相同或相近的论文分作同一组。为了便于考察评分偏差与教师研究领域的关系，同时标注教师的研究领域（表 1）。

### 2.2　纵向偏倚因子的计算

为了比较同一教师对不同主题的论文评分的差异，研究者引入了纵向偏倚因子这一概念（计算公式见表 2 注释）。该计算方法的核心假设是：尽管不同学生的课程论文质量可能存在较大差异，但在自由选题的课程论文考核中，学生会自主选择最擅长的主题进行作文。因此，质量较高的论文和质量较低的论文会随机分布在不同主题组，导致不同主题下的论文质量在整体上趋于相近。反映在论文评分上便是在没有其他偏倚因素的情况下，不同主题组论文的平均分应当相等或接近，即纵向偏倚因子接近零。若某一主题组的论文评分显著高于（或低于）其他论文的评分，则该主题组论文对应的纵向偏倚因子将明显大于（或小于）零，评分高出（或低于）的越多，纵向偏倚因子偏离零的趋势越明显。五位评分教师在不同主题组的纵向偏倚因子计算结果见表 2。

**表 1**　五位教师对课程论文的评分按主题分组结果

| 分组 | 主题 | 评分教师（教师研究领域） | | | | |
|---|---|---|---|---|---|---|
| | | （领域 A）教师 1 | （领域 B）教师 2 | （领域 C）教师 3 | （领域 D）教师 4 | （领域 E）教师 5 |
| 1 | 主题 A | 82 | 75 | 91 | 82 | 78 |
| | | 78 | 82 | 88 | 84 | 75 |
| | | 84 | 77 | 88 | 92 | 80 |
| | | 70 | 60 | 80 | 70 | 70 |
| | | 73 | 80 | 90 | 80 | 72 |
| | | 89 | 90 | 92 | 90 | 85 |
| | | 88 | 92 | 91 | 86 | 90 |
| | | 81 | 80 | 89 | 95 | 84 |
| | | 86 | 83 | 93 | 86 | 82 |
| 2 | 主题 B | 87 | 82 | 88 | 88 | 80 |
| | | 85 | 86 | 93 | 85 | 82 |
| | | 85 | 78 | 91 | 92 | 81 |
| | | 74 | 75 | 88 | 84 | 75 |
| | | 88 | 82 | 90 | 86 | 88 |
| | | 75 | 80 | 91 | 87 | 71 |
| 3 | 主题 C | 81 | 80 | 88 | 88 | 78 |
| | | 85 | 85 | 91 | 88 | 80 |
| | | 86 | 84 | 91 | 82 | 80 |
| | | 84 | 82 | 91 | 83 | 84 |
| | | 83 | 78 | 90 | 83 | 70 |
| | | 90 | 86 | 94 | 90 | 90 |
| | | 80 | 82 | 95 | 85 | 80 |
| | | 72 | 60 | 89 | 81 | 72 |
| 4 | 主题 D | 80 | 85 | 90 | 98 | 80 |
| | | 81 | 80 | 85 | 91 | 78 |
| | | 88 | 72 | 94 | 88 | 80 |
| | | 82 | 74 | 90 | 90 | 78 |
| | | 79 | 82 | 93 | 93 | 79 |
| | | 78 | 80 | 90 | 96 | 74 |
| | | 81 | 78 | 89 | 89 | 73 |
| 5 | 主题 E | 86 | 77 | 89 | 90 | 75 |
| | | 75 | 85 | 89 | 80 | 83 |
| | | 76 | 65 | 91 | 88 | 78 |
| | | 76 | 85 | 90 | 88 | 82 |
| | | 85 | 80 | 92 | 82 | 76 |
| 6 | 其他主题 | 90 | 90 | 95 | 90 | 93 |
| | | 86 | 85 | 93 | 89 | 82 |
| | | 86 | 80 | 88 | 90 | 84 |
| | | 78 | 78 | 88 | 85 | 75 |
| | | 84 | 75 | 92 | 86 | 87 |
| | | 81 | 80 | 93 | 84 | 80 |
| | | 78 | 82 | 92 | 88 | 72 |

### 2.3　横向偏倚因子的计算

从上文对纵向偏倚因子的介绍可知，该方法的一个重要前提是不同层次的论文在各组随机分布。但在实际教学过程中，一个班级的学生数量往往十分有限，若课程论文同时存在多个主题分组，分布到各组的论文数量必然更加有限，可能出现几篇高质量（或低质量）的论文正好都选择了同一主题的情形。这种情况必然导致该主题下的论文整体质量高于（或低于）平均水平，此时按照前面的方法计算得到的纵向偏倚因子必然偏大（或偏小）。这种情况下，该组论文得分高于（或低于）整体论文平均分是论文质量高于（或低于）平均水平的正确反映，并非由评分教师在评分过程中引入偏差导致。由于这种情况的存在，使得单纯依赖纵向偏倚因子无法有效区分具体是论文本身质量导致了评分的差异还是评分过程导致了这种差异。因此，研究者引入了横向偏倚因子的概念，即通过横向对比不同教师对同一组论文评分上的差异，来对上述情况进行区分（计算公式见表3注释）。

该方法的基本假设是：若一组学生的论文质量高于（或低于）整体平均水平，那么该组学生的平均分就会高于（或低于）平均水平，对应的纵向偏倚因子就会偏大（或偏小）。但面对质量较高（或较低）的论文，不同的评分教师都会给出一个相对高于（或低于）其他论文的分数，即评分教师都会给质量好（或差）的论文打高（或低）分。因此，代表横向对比结果的横向偏倚因子应该仍趋近零。相反，若一位教师因为对某个主题的论文有偏倚而给出较高（或较低）的评分，导致纵向偏倚因子偏大（或偏小），而其他教师因为不存在该偏倚因素，他们对该主题的论文评分和纵向偏倚因子不会出现相应的变化。因此，在计算横向偏倚因子时，出现偏倚的评分教师对应的横向偏倚因子就会显著偏离零。五位评分教师在不同主题的横向偏倚因子计算结果如表3所示。

## 3　研究结论

### 3.1　课程论文评分分组结果

将5位授课教师的评分汇总后，根据课程论文主题分成了6组，相同或相近主题的论文评分放在同一组，个别其他主题的论文全部纳入“其他主题”组。分组后分别以字母（A、B、C、D、E）代替原主题。相应地，教师研究领域也以字母（A、B、C、D、E）代替，且同一字母代表的课程论文主题与研究领域具有相关性，例如：主题为A的课程论文所涉及的内容属于研究领域A的范畴（表1）。

### 3.2　纵向偏倚因子计算结果

纵向偏倚因子为评分教师对某一主题的论文

**表2**　各教师不同主题组纵向偏倚因子计算结果（%）

| 分组 | 主题 | 纵向偏倚因子 | | | | |
|---|---|---|---|---|---|---|
| | | （领域A）教师1 | （领域B）教师2 | （领域C）教师3 | （领域D）教师4 | （领域E）教师5 |
| 1 | 主题A | −0.91 | 0.13 | −1.75 | −2.84 | 0.20 |
| 2 | 主题B | 0.75 | 1.01 | −0.25 | 0.06 | 0.10 |
| 3 | 主题C | 1.23 | −0.29 | 1.05 | −2.76 | −0.28 |
| 4 | 主题D | −0.77 | −1.64 | −0.28 | 7.25 | −3.01 |
| 5 | 主题E | −3.05 | −2.00 | −0.20 | −1.76 | −0.90 |
| 6 | 其他主题 | 2.17 | 2.44 | 1.62 | 0.66 | 3.69 |

注：纵向偏倚因子计算公式如下：令$\Delta_X$=主题$X$组的纵向偏倚因子，a=主题$X$组论文评分的均值，b=同一教师对主题$X$之外的其他论文评分的均值；则$\Delta_X=(a-b)/b\times100\%$。

**表3**　各教师不同主题组横向偏倚因子计算结果（%）

| 分组 | 主题 | 横向偏倚因子 | | | | |
|---|---|---|---|---|---|---|
| | | （领域A）教师1 | （领域B）教师2 | （领域C）教师3 | （领域D）教师4 | （领域E）教师5 |
| 1 | 主题A | 0.15 | 1.45 | −0.89 | −2.26 | 1.55 |
| 2 | 主题B | 0.51 | 0.84 | −0.73 | −0.34 | −0.29 |
| 3 | 主题C | 1.80 | −0.10 | 1.57 | −3.19 | −0.09 |
| 4 | 主题D | −1.35 | −2.44 | −0.74 | 8.67 | −4.15 |
| 5 | 主题E | −1.84 | −0.52 | 1.73 | −0.22 | 0.86 |
| 6 | 其他主题 | 0.07 | 0.41 | −0.62 | −1.82 | 1.97 |

注：横向偏倚因子的计算公式为：令$\Delta_i$=教师$i$的横向偏倚因子，$m$=教师$i$在某一主题的纵向偏倚因子，$n$=除教师$i$以外其他教师在该主题的纵向偏倚因子的平均值；则$\Delta_i=m-n$。

评分均值高出(或低于)其他主题的论文评分均值的百分比,考察的是同一教师对不同主题的论文评分的差异。大部分教师在不同主题组的纵向偏倚因子都小于3%,而教师4在主题D这组课程论文的纵向偏倚因子达到了7.25%,即教师4给这一组论文的平均分比其他论文的平均分高出了7.25%(表2)。这一结果提示该教师对该组论文的评分与其他论文评分存在明显差异,需要对该组论文的评分作进一步检验,以判断该差异是论文本身质量所致还是由于评分教师的偏倚所致。

### 3.3　横向偏倚因子计算结果

横向偏倚因子考察的是不同教师对同一组论文评分的差异情况,当不同教师对一组论文评分趋同时,代表不同教师评分差异的横向偏倚因子将趋于零,只有某一教师对一组论文的评分与其他教师出现分歧时,该教师在该组论文的横向偏倚因子才会明显偏离零。例如在表3中,教师5在第六组论文的纵向偏倚因子略大,但是其他教师对该组论文的评分也略高于平均分,相应的纵向偏倚因子也略大,使得教师5在第六组的横向偏倚因子并没有明显变大。相反,教师4对第四组论文的评分明显高于整体水平,即该教师对这组论文的评价高于整体水平,对应的纵向偏倚因子为正且数值较大,而其他几位教师对该组论文的评分则略低于整体水平,即其他几位教师对这组论文的评价低于整体水平,对应的纵向偏倚因子为负,上述差异使得教师4在第四组的横向偏倚因子明显偏大。这一结果提示该教师在对主题为D的论文评分时可能存在一定偏倚,导致该组论文评分出现了偏差,使得其平均得分高于其他主题的论文,而这一得分上的差异并非因为论文本身质量较高所致,而是由评分教师在评分过程中引入。因此,应要求相应教师对该组论文重新评分或在后续计算论文平均分时剔除这部分包含偏差的评分,从而消除这部分由评分教师引入的偏差,使课程论文得分更好地反映学生的论文质量。

## 4　结语

本研究对本校2020年《临床药理学》课程论文评分的分析结果可知,通过计算纵向偏倚因子和横向偏倚因子可以使评分结果的整体情况有一直观展示,从而有效发现论文评分中可能隐含的偏差。本文介绍的分析方法原理简单,涉及的计算都是基本运算,具有简单易行的特点,便于在实际教学过程中推广使用。相关方法具有良好的通用性,可以在不同学科的开放任务评分中应用。虽然这一方法具有上述优点,但要得到直观、可靠的结论往往需要多人对课程论文进行评分。在单人评分的课程论文考核中,尽管仍然可以利用相关公式获得纵向偏倚因子,但在对明显偏离零点的纵向偏倚因子进行判定时,因为没有横向偏倚因子作参考,教师必须对课程论文本身进行系统深入的分析、比较,才能判断评分中是否存在主题偏倚。

客观公正的评分是真实反映学生学习状况的前提,也是实现考核目的的先决条件。模糊的评分标准可能造成评分结果不能真实反映学生学习情况,不仅无法达到充分激发学生主动性和创造力来促进学习的目的,反而助长了学生应付了事,甚至抄袭剽窃的不良作风,最终对学生的学习带来严重的消极影响。在教学过程中,除了在完成论文评分后对评分结果进行分析,及时发现评分中存在的偏差之外。在论文评阅之前,评分教师可以利用一些现有的针对开放任务评价的方法,例如PTA量表法[6]、沃尔比重评分法等[7],根据不同学科的论文特点制定细化的评分标准,使论文评分标准在一定程度上得以量化,尽量减少评分过程中引入的偏差,从而使论文评分能更加客观准确地反映学生的学习情况。

**利益冲突**　所有作者声明无利益冲突

**作者贡献声明**　张孟科:提出论文构思、撰写论文;宋伟、李丹、沈秉正、陈莹:整理数据、制作表格;王宗春、郭咸希:总体把关、审订论文

**参考文献**

[1] 王艳霞, 胡优敏, 冯雪梅, 等. 实验整合课程考核评估体系建设实践[J]. 中华医学教育探索杂志, 2019, 18(11): 1103-1107. DOI: 10.3760/cma.j.issn.2095-1485.2019.11.006.

[2] Lynch GJ. The effect of theacher course work on student learning-evidence from the TEL [J]. J Econ Educ, 2014, 21(3): 287-296. DOI: 10.2307/1182248.

[3] Ma X. A longitudinal assessment of antecedent course work in mathematics and subsequent mathematical attainment [J]. J Educ Res, 2000, 94(1): 16-28. DOI: 10.1080/00220670009598739.

[4] 张松艳. 基于课程论文的研究生计量经济学教学改革:以浙江科技学院为例[J]. 浙江科技学院学报, 2019, 31(4): 335-338. DOI: 10.3969/j.issn.1671-8798.2019.04.012.

[5] 赵彦芳. 从课程论文到学位论文: 提升论文写作素养,深化课程教学改革[J]. 扬州大学学报(高教研究版), 2017, 21(3): 78-81. DOI: 10.19411/j.cnki.1007-8606.2017.03.016.

[6] Sheehe JB, Schoener L. Primary trait analysis student, course, and program assessment [J]. Nurse Educ, 2009, 34(3): 114-117. DOI: 10.1097/NNE.0b013e31819fca9d.

[7] 柴斌锋, 竺素娥. 沃尔比重评分法在学生平时成绩考核中的应用[J]. 亚太教育, 2016(17): 228-229. DOI: 10.16550/j.cnki.2095-9214.2016.17.186.

(收稿日期:2022-01-07)
(本文编辑:唐宗顺)