

学生评教偏差分析与控制模型研究

马秀麟 袁克定 张 倩

(北京师范大学 教育技术学院, 北京 100875)

【摘要】随着学生评教的普及, 学生评教中的偏差问题已经引起教学管理者的警惕。该课题从分析学生评教偏差现象和偏差产生的原因入手, 探索了对学生评教开展控制的控制模型, 并论证了模型的有效性。

【关键词】学生评教; 偏差分析; 控制模型

【中图分类号】G420

【文献标识码】A

【论文编号】1009—8097 (2011) 02—0040—05

随着高等教育大众化的到来, 高等教育质量问题成为影响高等教育能否可持续发展的决定因素。从高等教育质量监控的层次来看, 可以分为微观层次、中观层次和宏观层次, 在各类教学质量监控举措中, 由教学管理部门负责的学生评教、督导评教、院系领导听课制度是中观控制的常见手段。尽管学生评教已经在高校中普遍存在, 但对高校中学生评教有效性的争议一直存在, 对于如何提升学生评教的有效性水平, 也有许多学者进行了有益的探索。

一 针对学生评教数据样本的实证分析

思辨的方法不能解决学生评教数据的有效性和偏差问题。因此, 课题组决定针对真实的评教数据, 开展实证性分析, 从数据分析的视角探索评教数据的有效性和偏差现象。

1 数据样本采集及其信度、效度分析

(1) 评教样本的选取

为保证研究的有效性, 研究的样本量应该达到一定的规模, 并且具有一定的代表性。为此, 课题组选用了三所高等院校的学生评教数据开展研究。其中 A 大学是位于北京的一所综合性重点大学, 共采集了该校自 2004 年上学期至 2009 年上学期 (共 11 学期) 的全部学生评教数据, 大约有 80 万条记录 (7 万 $\times$ 11 条记录); B 大学是位于河南的一所普通院校, 只获得了该校 4 个学期的部分评教数据, 大约有 10 万多条记录。C 大学是位于山西的一所普通院校, 只获得了该校一个院系 2 个学期的评教数据, 仅有 2 万条评教记录。从当前获得的评教样本看, A 校的数据比较全面, 适合开展较为详细的分析, B 校和 C 校的数据规模较小, 但 B 校和 C 校的评教数据能够部分代表普通院校学生评教的情况, 因此可作为对 A 校数据的有效补充。

(2) 评教样本的信度和效度

课题组首先针对数据样本进行了内部一致性检验, 发现全体数据的克隆巴赫系数全部在 0.85 以上, 呈现出较好的可靠性; 然后, 分别以专家评价和教师在其他学期同一课程上

的评教得分作为再测复本, 开展复本检验, 发现数据之间的相关性显著, 其检验概率全部小于 0.05, 呈现出极好的一致性。因此, 从整体上看, 评教样本具有较好的信度和效度。

2 数据样本的离散性与偏差状况

充分肯定评教数据整体有效性, 并不能否认评教数据中偏差的存在。从评教数据的视角看, 学生评教中明显存在着以下问题:

(1) 少量学生的评教给分缺乏离散性。有少量学生对一门课程的所有评教指标项全部给予相同的分值 (比如全部 5 分, 或者全部 3 分), 仿佛没有认真地思考每个指标项的语义和教师在相关方面的表现;

(2) 部分学生的评教给分整体偏低 (或偏高)。部分学生对所有课程都给予了较低 (或较高) 的成绩, 尽管对不同的课程, 这些成绩能够区分出高低的顺序, 但对样本整体的统计分析极为不利;

(3) 通过网上调查和学生座谈得知, 在学生评教活动中有代评现象存在。

3 针对学生评教数据的协同性分析

在已经证明学生评教有效性和评教指标基本合理的情况下, 课题组分别从学生和课程两个维度对学生评教过程中的协同性进行了分析, 论证某一行政班级中多位学生针对某一课程 (学期+教师工号+课程代码) 的评价是否具有协同性、一致性。这类关于多位学生针对某一课程的评价是否保持一致性的研究虽然可以通过学生们评价的离散度 (方差) 来体现, 但在面对众多课程的时候, 利用多配对样本的 Kendall 协同系数检验更是非常有效的方法。在这种思路中, 如果把对每个评判对象的分数看作是来自多个总体的配对样本, 采用变形的 Friedman 检验, 就能通过 Kendall 协同系数检验多个评价者对多门课程开展评价的协同性水平问题^[1]。

(1) 多名学生针对同一课程的肯德尔协同系数分析

针对一个行政班级, 在此班级中有多名学生。在评教活动中, 这些学生充当着评价者角色。由于这个班级的学生同

时参与多门课程（多个教学班）的学习，现在可以检测这些评价者对多门课程的评价数据是否具有协同性和一致性。

首先，选取样本并进行数据准备。从 A 校的评教数据中选择一个人数比较多的行政班级，把该行政班级的所有学生和这些学生选修的所有课程以及评教结论整理到一个数据表

秩		检验统计量	
	秩均值	N	13
授课 05	3.08	Kendall W	.123
授课 34	2.88	卡方	6.400
授课 19	3.27	Df	4
授课 21	2.88	渐近显著性	.171
授课 22	2.88		

图1 执行面向学生的Kendall1协同系数检验

第三，协同性检验结论。由 Kendall 协同系数检验结果(图 1) 可知，几门课程的秩分别为 3.08、2.88、3.27、2.88、2.88，卡方检验统计量的值为 6.400，对应的概率 p 值为 0.174，大于预设的显著性水平 0.05，因此不能拒绝 0 假设，只能认为学生对各门课程的评价没有显著性差异，而协同系数（Kendall W）为 0.123，远远小于 1。因此可以认为各位学生对同一门课程的评价并不一致。

当然，从这个视角获得的结论仅仅可以证明多位学生对同一门课程实施评教的协同性较差，尚不能推出“学生对各门课程的评价是无效的”这种结论。

(2) 同一学生针对不同课程的肯德尔协同系数分析

针对一个行政班级中的多个教学班，课题组检测了该班级修读的各门课程对于多位评价者的评价是否具有协同性和一致性。

首先，选择样本并执行数据清理。从 A 校的评教数据中选择一个人数比较多的行政班级，把该班级的所有学生和这些学生选修的所有课程以及评教结论整理到一个数据表中；

其次，进行转置与清理，剔除无效数据，最终得到一个以课程为行、学生为列的二维数据表。

第三，执行 Kendall 协同系数检验。对获得的二维数据表执行多配对样本的非参数检验，要求实施 Friedman 检验并计算 Kendall W 协同系数。执行分析后获得的结果如图 2 所示。

检验统计量	
N	5
Kendall W	.938
卡方	56.253
df	12
渐近显著性	.000
a. Kendall 协同系数	

图2 执行面向学生的Kendall1协同系数检验

第四，协同性检验结论。由 Kendall 协同系数检验结果可知，几位学生的秩分别为 8.40、1.30、8.40、8.40、3.50，卡方检验统计量的值为 56.253，对应的概率 p 值为 0.00，小于预设的显著性水平 0.05，因此可以拒绝 0 假设，认为不同学

生对同一课程的评价有显著性差异，而协同系数（Kendall W）为 0.938，接近于 1。因此可以认为每位学生对自己修读的 5 门课程的评价具有一致性。

其次，执行 Kendall 协同系数检验。对学生、课程数据表执行多配对样本的非参数检验，要求实施 Friedman 检验并计算 Kendall W 协同系数。执行检验后获得的结果如图 1 所示。

秩		检验统计量	
	秩均值	N	13
授课 05	3.08	Kendall W	.123
授课 34	2.88	卡方	6.400
授课 19	3.27	Df	4
授课 21	2.88	渐近显著性	.171
授课 22	2.88		

图1 执行面向学生的Kendall1协同系数检验

生对同一课程的评价有显著性差异，而协同系数（Kendall W）为 0.938，接近于 1。因此可以认为每位学生对自己修读的 5 门课程的评价具有一致性。

(3) 对学生评教协同性检验的结论

通过对 A 校和 B 校的评教数据分析后得知，A 校和 B 校的多位学生针对同一门课程的评教一致性较差，因为其协同系数较低；而一位学生对不同课程的评教具有较强的协同系数，其一致性较强。

针对这一现象，课题组对部分学生进行了调查和座谈。调研发现，导致上述现象的原因在于：在实际评教过程中，确实存在着学生对评分基准理解的不一致问题（见表 1 的前测数据）。部分学生认为得分 4.5 以上的教师才是好老师，而另外有些同学认为获得 4.0 分的教师就是好老师。这种定位不一致的情况，直接导致了“多位评价者针对同一课程的协同系数较低，而同一学生面向不同教学班的协同系数较高”这一现状。另外，由于 B 校的评教指标不是一种等级制的体系，每个指标项都有自己的满分分值。部分粗心的学生常常因没有记住某些指标项的满分值，导致评教结果出现较大的偏差。再者，B 校指标体系中还存在着一个超出学生知识判定水平的指标项“能够结合教学需要充实本学科新进展和新成果”。这一原因也会导致 B 校学生评教的协同系数降低。

尽管如此，这并没有影响学生评教的有效性。因为对于每个学生来讲，针对自己面对的各门课程，其内心都有一个相对的、阶梯性的标准，只要这种居于学生内心的阶梯性分值满足排序一致性（即大家都认为某位教师的教学质量可排名第一、而另一位教师的教学质量可排名第二，……）就不会影响到学生评教数据的有效性，不会导致评教数据的混乱。由于这种具有层次性的评教分值已经能够区分优秀教师和存在教学缺陷的教师，所以，获得的评教结论仍然是有效的。

二 学生评教数据偏差原因分析

1 针对学生评教偏差原因的调查

2009 年初，课题组在 B 校选取了 200 多位学生进行了两

次网络调查，通过调查问卷征集了学生对生评教的想法、态度。在实际的测试中，共有 224 名学生参与了前测调查，回收有效数据 207 份；210 名学生参与了后测调查，回收有效数据 198 份。调查问卷的统计结果如表 1 所示。表 1 中前测数据较明显地呈现出评教偏差存在的原因。

表1 学生对评教活动看法的调查结果

题干	选项	前测		后测	
		人数	频率	人数	频率
您认为目前的学生评教实践对改进我校教学效果的作用	非常高	18	8.7%	25	12.6%
	高	42	20.3%	46	23.2%
	有点效果	88	42.5%	91	45.9%
	没有效果	59	28.5%	36	18.1%
针对每学期末开展的评教活动，您的态度是	一定要认真填写，真正地把自己的意见反馈出来	121	58.4%	163	82.3%
	这是学校布置的任务，随便填填得了	70	33.8%	9	4.5%
	不要和老师过不去，尽量填写高分就是了	10	4.8%	20	10.1%
如果您没有获得对评教结果的反馈，您认为自己的意见	受到了重视，只是没有告诉我罢了	59	28.5%	82	41.4%
	老师已经知道了，会在以后的教学中改进	32	15.4%	65	32.8%
	教务处会重视，但教师们未必会重视	31	14.9%	49	24.7%
	没有受到过重视，意见白提了	85	41.0%	2	1.0%
根据您的了解，在每年的学生评教活动中，能够认真填写评教数据的比例有多高	95%以上	5	2.4%	4	2.0%
	85%~95%	85	41.1%	80	40.4%
	50%~85%	87	42.0%	82	41.4%
	30%~50%	22	10.6%	23	11.6%
	很少	8	3.9%	9	4.6%
您一般给予比较满意的教师（ ）分（满分 5 分）。	5.0 分	35	16.9%	42	21.2%
	4.5 分	120	57.9%	156	78.8%
	4.0 分	52	25.2%	0	0%
	3.5 分	0	0%	0	0%

2 导致学生评教产生偏差的主要原因

从表 1 前测数据反映出的观点看，导致评教发生偏差的原因很多，有态度方面的问题，也有个体的评分习惯方面的问题。另外，学生评分的偏差度还受学生知识水平与判定能力的影响。

(1) 态度问题

态度是影响学生评教质量的重要因素。由于有的学校规定评教是学生的基本义务，不进行评教的学生不能选课、甚至不能获得奖学金，反而引发了学生的反感，导致部分学生把评教作为一项必须去应付的差使来做。极端的现象是：某些学生在登录评教系统后，根本不看评教指标项，胡乱选择一个分数了事。在这种情形下，常常导致评教数据缺乏离散度，对所有指标项、甚至所有课程都给予了相同的分值，基本就是无效数据。

(2) 个体评分习惯

由于对评教指标标准理解的差异，评教学生可能按照自己内心的基准给教师打分。有的学生以 4.5 作为基准分，对于不同课程上下浮动，也有学生以 4 分作为基准分，上下浮动。这一问题会导致学生个体对不同课程的评价呈现出较好的协

同性，即导致他们给出的分值普遍偏高或普遍偏低。

(3) 学生的知识水平与判定能力

学生评教是一个要求学生做出判定的过程，这个过程受学生知识水平和判定能力的影响。特别是某些评教指标中要求学生“对教学覆盖前沿知识”、“教学知识性、系统性”方面做出判定，已经超出了学生现有的知识水平，也会导致学生的评教结论出现较大的偏差。

总之，导致学生评教产生偏差的原因很多，但比较关键的因素是学生对评教的态度、对课程的期望、内心的评判标准、对大学学习模式的适应度。学生评教中的一些不良行为是否及时地得到纠正也是影响学生评教偏差度的重要原因。

另外，学生个体在评教过程中也会受到性格、情感、评教时刻情绪等方面因素的影响。当然，某些学生意志不坚定、喜欢不严格要求自己的教师，或者缺课严重，不能全面地了解教师的授课状况，也会导致在评教时出现较大的偏差^[2]。

三 学生评教控制的模型设计

1 针对学生评教开展控制的主导思想

(1) 激发学生的评教积极性

从学生评教数据的结果看,学生对评教活动的轻视和填写评教数据的随意性是影响评教信度的重要原因,而影响学生评教态度的关键因素在于学生是否已感受到自己的评教结论受到了校方的重视、学生是否及时地获得了关于评教状况的反馈。为激发学生评教的积极性,需要及时地对学生评教实施元评价、并及时地向学生反馈元评教结论,使学生感受到校方对评教的重视,进而激发学生评教的积极性;

(2) 借助教育信息平台,创建学生评教控制模型

随着信息技术的发展,基于信息平台的学生评教已经广泛应用。但对评教控制与评教数据应用的研究还较少。事实

上,由于信息平台的强大运算能力,可以保证在较短的时间内完成海量数据的统计分析,并能及时地把分析结论反馈给相关人员。因此,依托信息技术,建构评教控制模型来实现对学生评教质量的管理是完全可行的。即学生评教控制模型应该借助网络技术和数学模型,建构在信息平台之上。

2 学生评教控制模型

通过对近百万份评教记录的分析,笔者认为从学生评教数据的离散度、相似度和均值等角度对数据进行学生评教的元评价是可行的。其控制模型的基本结构如图3所示。

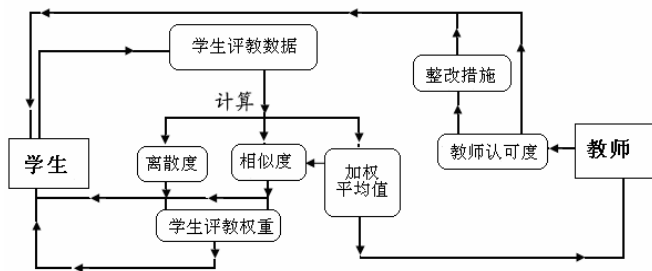


图3 学生评教控制模型流程图

为了能够清晰地说明控制流程,对本模型中涉及到的概念进行必要的说明:

(1) 离散度

用于反应学生每学期填报评教数据的离散程度,主要用于监测学生在评教过程是否全部填写了相同的分值。例如,有的学生对指标中的选项,全部填写为5分或3分,是一种典型的不负责任现象。通过测算此数据的离散度,并及时地把测量结果反馈给学生,给部分不重视评教工作的学生恰当的警示,尽量提高调查数据的信度。

(2) 加权平均值

由于对教师的每个测评指标都有多名学生给出测评结论,因此需要根据多名学生的测评结论计算出其加权平均值,作为教师在该指标上的得分(即标准分)。在计算过程中,新生的权重恒定为1,但老生的权重可由前一学期的测评数据计算获得^[3]。以加权平均值作为学生评教的效标。

(3) 相似度

针对教师的一学期授课,首先以每个测评指标的加权平均值作为该教师在此指标上的标准分,然后检测每个学生在全体指标上的评价分与教师标准分的相似程度。即检测该学生针对某教师的测评曲线与该教师的加权均值曲线的拟合程度^[4]。也可称之为偏差程度。

(4) 学生评价权重

通过每个学生的相似度结论和离散度结论,计算其评教权重。对评教权重很高的学生,可通过评教系统、院系辅导员或学生工作部门给予一定的奖励。

(5) 教师认可度

教师应及时地对学生评教结论进行评议,对每一个指标的分值都给出具体的认可度,并针对得分分值极低的指标提出具体的整改措施。

四 学生评教控制模型实践及其效果

对基于教务信息系统设计的评教控制模型,必须通过实践的检验并依据检验结果对控制模型进一步完善与改进。根据上述设计思路,课题组在B校开展了对评教控制模型的实践活动,并通过调查问卷、座谈会和数据分析等方式对评教控制模型的有效性进行了检验。

1 对学生评教控制模型的实践

根据上述设计思路,课题组按照学生评教控制模型算法对B校中文系2009年的评教数据展开了重新计算,获取了每个学生评教的偏差度和离散度,并进而得到了每个学生的评教权重,最后以权重作为反应学生评教质量的重要标志。这一过程的基本步骤包括:

- (1) 依据学生权重计算教师的评教标准分,作为效标;
- (2) 计算每个学生与效标的偏差度(即效标距),同时计算每个学生的评教离散度;
- (3) 依据学生的评教偏差度和离散度,计算本学期每位学生评教的权重值;
- (4) 把针对学生评教的再评价信息通过教务系统反馈给每一个学生。

教务信息系统向学生反馈评教质量的界面如图4所示。

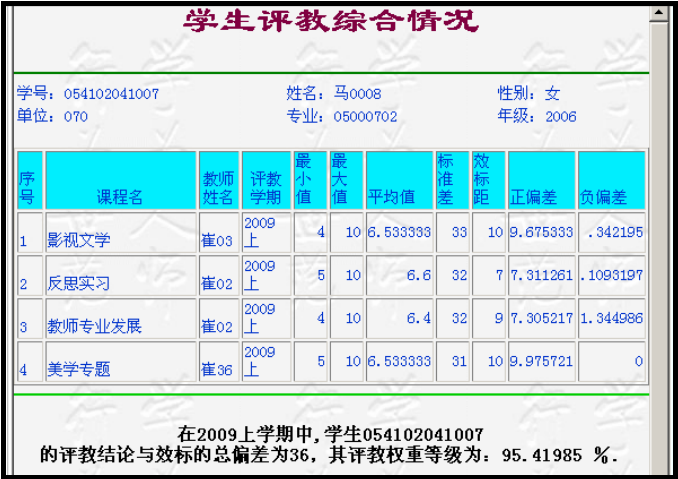


图4 向学生反馈评教质量的示意图

2 评教控制模型应用效果

(1) 应用评教控制模型能够提升评教数据的信度和效度

按照评教控制模型, 以加权平均分思路重新计算教师的评教得分, 然后对新的评教结论进行内部一致性检验。分析结果证实: 检验结论与前述结论相同, 数据的可靠性系数为0.986, 显示出极好的可靠性, 而且其 F 检验对应的概率值p=0.00, 各个指标项的加权均值在总体上存在显著性差异, 能够较好地反应教学质量的各个方面。

(2) 对比应用评教控制模型前后, 学生的评教积极性有了明显提升

为获得评教控制模型应用效果的第一手资料, 课题组在实施学生评教控制后对相关学生再次进行了问卷调查, 其统计结果如表 1 的后测部分所示, 其中 5 个指标项的统计值变化明显地证实了应用评教控制模型实施数据校正与及时反馈的必要性。

从表 1 的后测数据列可知: 在“针对学期末学生评教活动的态度”栏目中, 选择“一定要认真填写, 把自己的意见正确表达出来”的比例从 58.4%提升到了 82.3%, 认为对学生评教要认真对待的比例提升了接近 25 个百分点, 而把“评教作为一个任务进行应付的”比例则从 33.8%下降到了 4.5%,

这是一个非常可喜的变化; 另外, 在“您认为自己的意见会受到如何对待”栏目下, 有更多的学生认为自己的意见得到了重视(从 28.5%上升到了 41.4%), 认为“自己的意见白提了”的比例则从 41%下降到了 1%。

调查结论证实, 在学生接触到评教控制系统并获得了校方对自己评教的反馈后, 学生针对评教的态度有了较大的变化, 而且学生针对评教标准的理解也更为清晰, 减少了基准不一致现象的发生。

参考文献

[1] 衷克定.SPSS for Windows 数据统计分析工具应用教程[M]. 北京:北京师范大学出版社,2001:86-88.
[2] 张臻,王建军.对影响学生评教因素的研究与分析[J].教育管理,2009,(3):129-130.
[3] 王孝玲·教育统计学(第二版)[M].上海:华东师范大学出版社, 2001:41-48,52-54.
[4] ZHAO Hui-tang 著,邝祝芳译·数据挖掘原理与应用(SQL Server 2008 数据库)[M].北京:清华大学出版社,2007,(1): 109-253.

Research of Excursion Analysis & Control Model on SETQ

MA Xiu-lin ZHONG Ke-ding ZHANG Qian

(School of Education Technology, Beijing Normal University, Beijing 100084,China)

Abstract: The problem of the SETQ Data excursion is found when we analysis the validity of the SETQ, So we reassess the data of SETQ by these method: Discrete Degree, Weighted Average, Similar Degree etc, and then prove the control model is useful by the network investigation.

Keywords: SETQ(the Students Evaluation of the Teaching Quality); excursion analysis; control model

收稿日期: 2011 年 1 月 10 日

编辑: 小禾