

# 基于标准分的高校学生评教数据统计分析与结果应用研究

于长英<sup>1</sup>,高旭<sup>2</sup>

(1. 沈阳建筑大学教学咨询与督导工作办公室, 辽宁 沈阳 110168;  
2. 沈阳建筑大学学报编辑部, 辽宁 沈阳 110168)

**摘要:**基于目前大部分高校对学生评教数据的统计分析与处理过于简单的现状,对评教结果缺乏足够的说服力、评教结果运用不充分、无法充分发挥学生评教作用的缺陷进行分析,提出了基于标准分的学生评教数据处理的新思路和方法,通过提高评教结果的可比性,使评教结果信度和效度大幅提高,从而充分发挥学生评教的积极作用。

**关键词:**标准分;高校;学生评教;统计分析;结果应用

**中图分类号:**G64;C81      **文献标志码:**A

学生评教作为高校评价教师教学质量的一个重要手段和教学质量保障体系的重要环节,越来越受到各高校的重视。学生评教,不仅为学校了解整体教学情况,总结和推广先进教学经验,改进教学管理工作,制定提高教学质量有效措施提供了重要依据,更重要的是可以使教师了解自己的教学状况和教学效果,总结教学经验,改进教学工作,不断提高教学质量和教学水平<sup>[1]</sup>。

随着高等教育大众化的不断深入,教学质量日益成为影响高等教育可持续发展的重要因素。在这个过程中,学生主体作用的发挥、学生对教学工作的看法和评价越来越受到重视,依据学生对教师教学工作的评价,即学生评教结论来改进教学工作已逐步成为各高校调控教学质量的重要手段。因此,从20世纪80年代开始,越来越多的高校开展了针对教师教学情况的评教活动,充分发挥了学

生的主体作用<sup>[1]</sup>。管理人员利用学生评教评测教师的教学质量和效果,利用评教结果对比教师的教学质量好坏,并进行排序,用于教师的年终考核和职称评聘中一票否决。然而,教师对此反映较大。很多教师认为学生评教随意性很大,很多学生根据自己的喜好、本门课程成绩及老师要求的宽严等来评价老师,有的甚至评的是老师的外貌、年龄等,对严格、认真、年龄较大的老师不公平。与之相对的是,教学管理者则试图通过各种方式来证明学生评教的信度和效度,认为当学生数量达到一定程度时,学生评教的结果是可靠的。例如,马秀麟等<sup>[2]</sup>“从大数据挖掘的视角分析学生评教的有效性”,通过研究证明评教数据具有较好的内部一致性;通过跟踪分析同质教学班的多轮教学,证明同质教学班在不同学期的评教成绩存在显著性相关,督导的评价也与学生评教结论高度相关。即

学生评教数据具有很高的信度和效度。尽管如此,在教师群体内部和学术界,对于“学生评教数据的客观性、评教结论反映教学质量的科学性,以及应该如何将评教数据运用在实际教学管理中”等问题,仍然存在着激烈争论<sup>[3]</sup>。赞成的观点高度肯定了学生评教在教师考核中的作用,认为在教师考核和职称晋升中运用学生评教结果来反映教师的教学质量,充分体现了学生的主体地位,有利于教师改进教学工作,提高自己的教学质量和水平。而反对者则认为,由于学生个体认知能力不同,在知识结构和学科水平方面也存在差异,导致学生无法从自己的视角对教师的教学质量作出准确评价;同时,学生在评教过程中的随意性、甚至对老师的报复行为,很难保证学生评教数据的客观、科学、公正。之所以会造成教师和管理者反应不一的现象,其原因在于目前各高校对评教结果的统计分析过于简单,缺乏可比性<sup>[4-6]</sup>。

## 一、高校学生评教结果统计分析及应用现状与存在问题

### 1. 学生评教结果的统计方式

目前,绝大部分高校已采用网上评教系统进行学生评教活动。从统计学角度和对教师认真负责的角度考虑,为保证评教结果的客观、公正,一般高校都要求所有学生对给自己上课的所有任课教师均进行教学评价打分,而且理论课评价指标体系和实践课、体育课也有区别,学生按不同的评价指标体系对相应课程的任课教师评价打分,学校通过对学生评教数据的统计进行分析,才能得到每位老师的评教结果。而统计分析的方法和质量不仅会直接影响到评教结果,而且还会影响到评教结果的应用价值。这可能就是造成教师对学生评教工作意见很大,从而使评教结果不能得到充分运用的一个重要原因。

目前,绝大多数高校对评教结果的统计都是采用受评教师各门课程的加权平均分,即各门课程评价分数与评价学生人数乘积之和除以学生总人数。其中每一门课程的分数

是该门课程所有参评学生评分的算术平均值。算术平均分充分运用了不同学生的评教分数,综合反映了各参评学生对被评教师教学情况的评价,将评分的多少作为评价教师教学质量的标准似乎合情合理,无疑也是各高校应用最广泛的统计方法,然而也是最简单的统计方法。笔者通过多年对学生评教结果的仔细研究发现,直接应用学生原始评价数据的算术平均值作为评判教师教学质量的标准,并以此来比较教师教学质量好坏,其实有很明显的缺陷。

(1) 评教主体的影响。不同的教师,给他们打分的学生的背景完全有可能不一样。例如班级不同、年级不同,甚至学院也不同。不同的评价者尽管评价指标体系相同,但掌握的评价标准不可能完全相同,甚至差距很大<sup>[5]</sup>。造成虽然评价分数相同但由于评价标准不同,相同评价分数的教师教学质量和效果不仅不同,还可能差距较大。例如,在同一学期,A老师的学生评教分数的平均值是90分,而B老师的学生评教分数平均值是88分。如果这两位老师在不同的学院,因教师群体不同,一般不会去比较,但如果同一学院,通常按学生评教平均分的高低来评判教师教学质量的优劣,很容易得出B老师的教学质量和效果较差,A老师更受学生认可的结论。然而得出这一结论的前提,必须是所有班的学生对每个教师的评价标准是一样的,而问题恰恰是如果这两位老师的授课对象并不是同一个班级甚至不同年级、不同学院,而由于不同群体的学生对教师的评价标准不尽相同。很有可能出现的情况是:A老师的授课班级为P班,该班同一学期内有6位授课老师,除A老师外其余5位老师获得的评教分数分别是95,94,93,92,91;B老师的授课班级为Q班,Q班也有6位授课老师,除B老师外该班学生对其余5位老师的评教分数分别是89,88,86,85,84。在这种情况下,还会有人说A老师比B老师教学质量好吗?当然不会。很简单,在这个例子中,我们可以明显地获得两个信息:一是P班学生给

P班上课的授课教师评教分数普遍比Q班学生给Q班上课的授课教师的评教分数要高;二是A老师在P班各位授课老师的评教分数中排名最后,而B老师虽然评教原始分数低于A老师,但其在Q班各位授课老师的评教分数中却排名第二。对于第一点,由于不能将不处于同一参照系下的分数强行放在一起进行比较,因此,不能因为P班任课教师的评教分数高,就简单地认为P班授课老师的教学水平普遍比Q班高,更受学生认可。同理,也不能简单地用P班学生给A教师打出来的原始分数,直接与Q班学生给B教师打出来的原始分数进行比较。因为两个不同班级在对教师进行教学质量评价时,可能评分标准在整体把握上会有较大差异。这类似于许多项目评审中不同的评委对同一项目给出的评分不同,尽管评价指标体系相同,但各自的评价标准不同一样。如此一来,不同评分标准下,评教的原始分数就很难准确反映学生对所评教师的认可程度,即教师教学质量。而一般来说,一个高校中某个学院基本都有十几甚至几十个班级,学院对教师教学质量评价往往是通过学生评教结果对学院内所有教师进行排序比较,难免会遇到这样“低分高质”的反例;更何况有些基础学科和公共学科的学院,给其他各学院的学生上课,不同学院之间学生的评教标准更可能是千差万别,从而造成这样的情况相当普遍。如果扩大到全校性教师比较中,如教学优秀奖的评选中,不同学院的教师,用学生评教的原始分数去说明各参评教师教学质量的高低,不仅不科学,而且可以说对评选没有任何实用价值。

(2)课程性质的影响。不同的课程,所属的学科领域、专业等不同,在培养方案中的要求和定位也不同,学习要求与授课形式也不同,不同性质的课程,对学生的重要性和吸引力也不同,如是通识课还是专业课,是必修课还是选修课,是考试课还是考查课,是理论课还是实践课等等,这些都会对学生评教结果产生影响<sup>[6]</sup>。

(3)教师和学生个体的影响。不同的教师讲课方式、方法不同,对学生要求宽、严不同,考试的难、易程度不同,教师的年龄、性格、性别、外貌、气质等,都会或多或少影响学生评教的结果。同时,不同学生对不同教师的讲授风格接受程度不同,学生的兴趣、爱好不同,这些都可能影响到对该教师教学质量的客观评价<sup>[7]</sup>。

可见,即便是对给同一班级授课的几位老师之间进行比较,也受课程性质、难易程度、教师讲课风格等因素影响,更何况是不同班级、不同专业、不同学院之间的课程,因此,通过对学生评教结果进行简单的平均分统计来确定教师的教学质量排序是不科学的<sup>[8]</sup>。

## 2. 学生评教结果的应用现状

尽管各高校把提高教学质量放在学校工作的首要位置,由一把手主要负责,也一直强调学生评教的重要性,然而衡量一所高校是不是真的重视了学生评教,关键要看有没有使用学生评教结果,是如何使用的;也就是说是否把学生评教结果真正作为衡量教师教学质量的重要指标;是否将评教结果向教师反馈,使教师以此来改进教学,从而提升教学质量;是否将评教结果与教师职称评聘、评奖、评优挂钩等<sup>[9]</sup>。经了解,大部分高校,基本均是各学院内部按教师评教平均分排序,学校组织校级督导组对排在后面的教师集中听课督导,帮助其改进教学方法和教学艺术,促进其教学质量的提高。对反复听课却仍然教学效果极差的个别教师有调离教学岗位的,也有在职称评聘中一票否决的,这是评教结果用得较好的高校。当然也有一些高校,学生评教也是每学期期末进行一次,正常反馈给各学院,虽然学校教学管理部门也对各学院学生评教结果的运用有一定要求,但各学院使用情况不尽相同。有的学院在教师年终考核时占一定比例,并与评价排名靠后的教师谈话,帮助其分析原因,查找不足,促进其教学质量的提高;也有些学院,领导拿到结果后就锁在抽屉里了,对评教结果不采取任何措施也不加以利用,使评教完全成了走



形式和过场。

可见,从学校到学院,对“学生评教”结果的应用并不充分,这种状况不仅反映了对此项工作的重视不够,还存在一定客观因素,与评教结果的可比性不高有直接关系。因此,必须改进评教结果的统计方法,使其具有可比性。

二、基于标准分的学生评教结果统计方法探索

从对高校学生评教结果统计与应用现状分析中,不难看出,学生评教结果之所以没有得到广泛、有效应用,主要是因为目前的评教数据统计与处理方法使得评教结果的直接可比性不高,那么如何增加学生评教结果的可比性呢?笔者认为可以借助统计学中标准分的概念<sup>[10]</sup>。

1. 标准分数的概念

上文中已经阐述了由于直接使用原始评教分数可能造成的“高分低质”现象,引入标准分就可以弥补这些不足吗?那么可以先来看看标准分的含义。

标准分数是一种由原始分数推导出来,以一组分数的标准差为单位,以这组分数的平均值作为参照点,表示一个原始分数在团体中偏离平均值几个标准差的位置,从而表明该分数在这组分数中的相对位置的量数,又称基分数或 Z 分数。由数值和正负号两部分组成,数值表示原始分距离平均值的远近程度,正负号说明原始分是大于还是小于平均值。Z 分数准确地反映了一个分数在一组分数中的相对位置。简单地说,标准分就是原始分与平均分的差,除以标准差所得的商。计算公式为

$$Z = (X_i - X) / S \tag{1}$$

式中: $X_i$  代表一组分数中的任意一个原始分数; $X$  代表这组分数的平均值; $S$  为这组分数的标准差,是反映原始分数离散程度的一个量,是所有数与平均值的差的平方和除以数的个数(当数足够大时是个数减一)的算术平方根。标准差是方差的算术平方根。计算公式为

$$S = \sqrt{\sum_{i=1}^N (X_i - X)^2 / N} \tag{2}$$

2. 标准分数的特点

(1)可以相互比较大小。从标准分的概念可以看出,标准分与原始分不同,由于它以标准差为单位,以团体平均分( $X$ )作为比较的基准,反映的是某个数据在一批团体数据中的相对位置,因此不同背景下的成绩,只要转换为标准分数,同一分数就具有相同的含金量,就相当于处于同一背景下,从而可以相互比较大小。

(2)可以进行加减运算。由标准分数的概念同时也可以看出,标准分数是一个抽象化数值,不受原始分数团体影响,能使不同背景下的原始分数转化成同一背景下,因而可以进行加减运算。

(3)意义更加明确、直观。知道了某一原始分数的标准分数,就可以很明确地知道该分数在一组分数中的位置,所以,标准分数比原始分数更能直观反映出数据的意义。

由标准分的这些特点可以看出,原始分数转换成标准分数后,既具有可比性,又具有加和性。因此,可使批量分数的汇总与比较更直观、更科学。用标准分数来处理学生评教的结果应该更科学、更客观、更公正。

3. 基于标准分的学生评教数据处理

以前文提到的 A、B 两位老师学生评教情况为例,根据公式(1)和(2)进行计算。

(1)计算 P 班各授课教师评教分数  $X_i$  (指 P 班各参评学生对该班某位授课教师评价打分的平均分)的平均数  $X_p$  (即 P 班授课教师的  $X_i$  的平均数)、标准差  $S_p$  (P 班授课教师的  $X_i$  的标准差)和标准分数  $Z_p$ :

$$X_p = (95 + 94 + 93 + 92 + 91 + 90) / 6 = 92.5$$

$$S_p = \sqrt{\sum_{i=1}^6 (X_i - X_p)^2 / 6} = 1.87$$

$$Z_p = (X_i - X_p) / S_p \tag{3}$$

(2)按公式(3)计算 P 班 6 位授课教师各自的标准分数:95,94,93,92,91,90 这 6 个原始分数对应的标准分数依次是:1.34,0.82,0.27,-0.27,-0.82,-1.34。可知 A 老师的标准分数为 -1.34,根据前文所述标准分数的概念,负

值表示 A 老师受学生认可的程 度在 P 班 6 位老师受学生认可的 平均水平之下,与平均水平相差 1.34 个标准差(1.87)。

(3) 计算 Q 班各授课教师评教分数的平均数  $X_q$ 、标准差  $S_q$  和标准分数  $Z_q$ :

$$X_q = (89 + 88 + 88 + 86 + 85 + 84)/6 = 86.67$$

$$S_q = \sqrt{\sum_{i=1}^6 (X_i - X_q)^2/6} = 1.97$$
$$Z_q = (X_i - X_q)/S_q \tag{4}$$

(4) 按公式(4) 计算 Q 班 6 位授课教师学生评教原始分数 89,88,88,86,85,84 对应的标准分数分别是 1.18, 0.68, 0.68, -0.34, -0.85, -1.36。从这一结果不仅可以看出 B 老师的学生评教标准分数是 0.68。同时也表明,B 老师受学生认可的程 度在 Q 班 6 位老师平均水平之上 0.68 个标准差(1.97),在 Q 班并列第二。

(5) 根据标准分数可以相互比较大 小的特点进行比较,B 老师学生评教的标准分 0.68 明显高于 A 老师的标准分 -1.34。因此,可以认为,B 老师的教学比 A 老师更受其授课班级学生认可,其教学效果更好。

4. 标准分与 Z'分数的转换

综上所述,由于标准分数的可比性,将学生评教原始分转换成标准分后,使得对给不同班级授课的教师的评教分数进行比较成为可能。但是,如此计算出来的标准分数有一半左右不是正数,看起来既不直观,也难于理解,所以,需要对标准分做适当处理。最科学有

效又不改变标准分优越性的方法就是将标准分数进行线性转换。具体公式如下:

$$Z' = aZ + b \tag{5}$$

其中  $a$ 、 $b$  为常数,一般取  $a = 10$ 、 $b = 100$  经线性转换后的标准分数,不仅保持原始分数的分布形态,并具有原来标准分数的一切优点,同时变成了正数,方便比较。

例如,将 P 班 6 位教师的学生评教标准分数按式(5)线性转换后得到的  $Z'$  分数分别为:113.4,108.2,102.7,97.3,91.8,86.6。Q 班的 6 位教师学生评教标准分的  $Z'$  分数分别为:111.8,106.8,106.8,96.6,91.5,86.4。A 老师的  $Z'$  分数为 86.6,同样比 B 老师的  $Z'$  分数 106.8 要低很多,和原比较结果一致。可见  $Z'$  分数不仅保持了  $Z$  分数原有可比性、可加和性的特点,而且更方便、直观。

5. “多对多”的处理

前面讨论的是简单的一对一的情况,即两位老师分别只给一个班级授课。而在实际工作中这种情况较少,经常会出现一个老师同时给多个班授课或讲授多门课程的情况。而教师某门课程的学生评教成绩是上这门课的多名学生评教的平均结果,那么当这位教师同时给多个班授课或讲授多门课程时,就形成了“多对多”的情境。例如,如果在某学期 A、B 老师都各上了 5 个不同班的课,那么如何比较他们的教学在该学期受学生认可的程度呢? 假设 A、B 两位老师分别从 5 个教学班得到的学生评教原始分数以及各教学班学生为该班所有任课教师所打评教分数,计算得到的平均数和标准差如表 1 所示。

表 1 A、B 两位老师从各班获得评教分数标准分情况

教师	授课班级	给 A、B 老师打的原始评教分数	评教分数的算数平均值	给本班所有任课教师评教分数的平均分	对本班所有任课教师评教分数的标准差	对 A、B 老师评教分数的标准分
A 老师	A1	93	91.8	95	4	-0.50
	A2	91		90	4	0.25
	A3	93		92	2	0.50
	A4	90		92	2	-1.00
	A5	92		93	4	-0.25
B 老师	B1	88	89.0	87	4	0.25
	B2	89		89	4	0.00
	B3	91		89	2	1.00
	B4	90		91	2	-0.50
	B5	87		86	4	0.25

如果不采用标准分数,分别计算各人的评教平均分,可知 A 老师 5 门课程学生评教原始平均分 91.8 分,B 老师得分为 89.0 分。乍看起来,A 老师学生评教平均分比 B 老师高将近 3 分,应该教学质量更高,更受学生认可。但如果我们按上述方法分别计算各教学

班给 A、B 两位老师打的评教分数的标准分数,可以明显看出,即使是原始评教分数相同,但各班给所有任课教师打分的评均值和标准差不同,则计算得到的标准分也会截然不同。为方便比较,进一步计算 A、B 两位老师的标准分 Z' 分数、标准分平均数、标准分平均值的 Z' 分数如表 2 所示。

表 2 各班给 A、B 两位老师所打评教分数的各种形式平均分统计情况

教师	授课 班级	学生评教分 数的 Z' 分数	Z' 分数的 平均数	评教分数标 准分之和	评教分数的标 准分平均值	标准分平均值 的 Z' 分数
A 老师	A1	95.0	98.0	- 1.00	- 0.20	98.0
	A2	102.5				
	A3	105.0				
	A4	90.0				
	A5	97.5				
B 老师	B1	102.5	102.0	1.00	0.20	102.0
	B2	100.0				
	B3	110.0				
	B4	95.0				
	B5	102.5				

可见,经过线性转换后的 Z' 分数的平均数,仍然是 B 老师更高,结果再一次出现了反转。同时,从表 2 还可以看出,先求标准分的平均值再线性转换和先对各标准分进行线性转换后再求平均值,结果是一样的,即 B 老师 Z' 分数更高。这也同时验证了前文所提到的标准分数的可比性和可加和性,从而实现对教师教学效果进行科学、公正地比较。其实,不用线性转换,直接用标准分平均值的大小,一样可以进行比较(之所以不能用标准分总和进行比较,因为班数可能不同),只是经过线性转换后数据更直观、更易理解。

三、通过增加参评学生数量消除师生个人和课程性质等因素的影响

根据统计学理论,当参与评教的学生数量足够大时,由师生个人因素、课程性质等因素影响造成的随机误差其正负会相互抵消,其影响几乎为 0,可以忽略不计<sup>[2,8]</sup>。大量有关学生评教的研究也表明,当学生数量达到足够多时,这种随机误差的干扰因素可以忽略不计。所以,学生评教时要求尽可能多的学生参评,以消除来自于师生自身或课程性

质等方面随机因素的影响。现在各高校基本都应用网上评教系统,方便采取措施要求学生对自己上课的所有教师都要进行评价,否则不能查看成绩。

四、学生评教结果的应用

以前学生评教结果之所以没有得到充分、有效的利用,是因为受评者本人——广大教师对学生评教结果存在质疑,学校教学管理人员鉴于评教结果统计分析不到位使评教分数可比性较差,也感觉底气不足。而现在引入标准分数对学生评教结果进行相关的后续处理后,教师在某一学期无论是上几门课,给多少个班上课,都可以通过学生评教标准分对教师的授课效果进行相互之间科学、公正、无障碍地比较。可见,标准分的引入,不仅可以使评教结果作为各级教学管理部门评估教师教学质量、与教学相关的评优评奖等工作中可靠的参考指标,更重要的是,还可以使教师本人在相互比较中更清楚地了解自己的教学状况、在团体中所处的位置,发现不足,进行相应地改进和提高,进而促进教学质量的整体提升,这才是学生评教最重要的目的。

## 五、结 语

学生评教虽然已在各高校开展很多年,但其促进教学质量提高的目的还没有得到充分发挥,这与评教结果统计分析的简单化有直接关系。只有科学合理地处理和分析评教结果,使教师教学质量相互比较成为可能,也使教师更能正视自己存在的问题,从而为改进和提高教学质量而努力,才能使学生评教真正达到促进教学质量提高的目的。

## 参考文献:

- [1] 于长英. 基于柯式模型的高校教师教学效果评价机制探索[J]. 山东高等教育, 2015(11): 48-54.
- [2] 马秀麟, 袁克定, 刘立超. 从大数据挖掘的视角分析学生评教的有效性[J]. 中国电化教育, 2014(10): 78-84.
- [3] 曹超, 张莲. 高校“学评教”现状分析及问题探讨[J]. 中国成人教育, 2010(4): 21-22.
- [4] 张进. 高校学生个体因素对学生评教结果影响的实证分析[J]. 前沿, 2013(16): 182-183.
- [5] 秦勇, 罗明薇, 于洁. 影响大学生评教结果客观性的实证研究[J]. 吉林大学学报, 2013, 3(29): 14-15.
- [6] 梁进, 鲁映青, 孙利军, 等. 教师背景特征对学生评教结果影响的实证分析[J]. 中华医学教育探索, 2014, 13(6): 550-553.
- [7] 赵卿. 影响学生评教结果有效性的因素分析及对策思考[J]. 科教导刊, 2015, 4(11): 159-160.
- [8] 吴星. 高校学生评教方法的固有限制、不当应用及其改进探析[J]. 重庆理工大学学报(社会科学版), 2016, 30(3): 116-124.
- [9] 唐云, 张丽雯, 寸丽仙. 论高校“学生评教”结果的统计与应用[J]. 学校管理, 2015(4): 205-207.
- [10] 刘子龙, 张卓. 基于标准分的学生评教结果的优化处理[J]. 现代教育管理, 2010(8): 52-54.

# Study on the Data Statistical Analysis and the Results Application of Students' Teaching Evaluation in Higher Education Institutes Based on the Standard Score

YU Changying<sup>1</sup>, GAO Xu<sup>2</sup>

(1. Office of Teaching Consultancy and Supervision, Shenyang Jianzhu University, Shenyang 110168, China; 2. Editorial Department of Journal, Shenyang Jianzhu University, Shenyang 110168, China)

**Abstract:** Since the statistical analysis and data processing of students' evaluation of teaching are too simple in most higher education institutes, this paper analyzes the following problems: the lack of sufficient conviction in the evaluation results, the insufficient use of the evaluation results and giving no full play to the students' evaluation. Moreover, this paper puts forward the new ideas and methods of data processing of students' teaching evaluation based on the standard score. By improving the comparability of the evaluation results, the reliability and validity of the evaluation results will be greatly improved and the positive effects of the students' evaluation of teaching will be fully achieved.

**Key words:** standard score; higher education institutes; students' evaluation of teaching; statistical analysis; the application of the results