

◆ 本文引用: 徐红彦, 孟莉, 郭志坤. 中青年教师教学竞赛评委的信度检验[J]. 新乡医学院学报, 2016, 33(6): 552. 封三. DOI: 10.7683/xyxyxb.2016.06.030.

【医学教育研究】

## 中青年教师教学竞赛评委的信度检验

徐红彦<sup>1,2</sup>, 孟莉<sup>1</sup>, 郭志坤<sup>1,2</sup>

(1. 新乡医学院基础医学院, 河南 新乡 453003; 2. 新乡医学院教务处, 河南 新乡 453003)

**摘要:** 教学竞赛是提高教师教学水平, 加强高等院校师资队伍建设的措施。新乡医学院已成功举办9届中青年教师教学竞赛活动, 竞赛的评价是否客观、科学、公平、公正成为全校师生关注的焦点。本研究以第9届教学竞赛的评分数据为对象, 进行竞赛评分的信度检验, 从统计学角度检验竞赛评委评价工作的可靠性。结果表明, 整个评分工作具有较高的信度。

**关键词:** 教学竞赛; 评委; 评价; 信度

**中图分类号:** G642.1 **文献标志码:** A **文章编号:** 1004-7239(2016)06-0552-02

如何迅速提高青年教师的教学水平, 是高等院校师资队伍建设的重大课题。教学竞赛是科学审视和检阅高校教学工作的重要举措, 是夯实教学基础、强化教学质量、提高教学水平、促进青年教师成长、推动高校教学工作的进步和教育事业可持续发展的有效途径<sup>[1]</sup>, 只有做好教学竞赛的评价工作, 才能起到教学竞赛的预期作用<sup>[2]</sup>。作者以新乡医学院第9届中青年教师教学竞赛的决赛评分数据为研究对象, 对评委之间评分一致性程度和单个评委评分与选手最终得分之间的一致性程度进行分析, 从统计学角度检验本次教学竞赛的评价质量。

### 1 对象与方法

**1.1 研究对象** 本研究以新乡医学院第9届中青年教师教学竞赛的决赛评分数据为研究对象, 此评分数据完整且不缺项。本次比赛共分6个比赛组, 每组评委分别由4个教师评委和12个学生评委组成, 其中, 每个教师评委的评分为1个有效评分, 学生评委以4个评分为1组, 每组的平均分为1个有效评分。去掉最高分和最低分, 其余得分平均分为参赛选手的比赛得分。

**1.2 比赛方法** 比赛分为综合教学和临床教学2个类别, 每个类别又分为助教组、讲师以上(含讲师)组和双语组, 各组比赛同时进行。抽签决定参赛教师出场顺序, 每人讲课限时20 min。每组各设一

等奖1名、二等奖2名、优秀奖若干名。

**1.3 评分标准** 评分以《新乡医学院中青年教师教学竞赛评分标准》为依据, 共计10个指标(其中双语组为9个指标), 满分10分, 每位评委的评分小数点后保留2位有效数字。

**1.4 评价指标** 多个评委的评分一致性程度由肯德尔和谐系数  $W$  (Kendall's  $W$ ) 与克朗巴赫  $\alpha$  系数衡量。在经典测量理论中, 肯德尔和谐系数  $W$  是计算多名评分者一致性程度最常用的统计量, 应用 SPSS 软件进行计算, 然后进行  $\chi^2$  检验<sup>[3-4]</sup>。克朗巴赫  $\alpha$  系数一般用于评估测试项目内部的一致性, 本研究中用于评估评委评分的一致性。应用 SPSS 软件计算本研究的克朗巴赫  $\alpha$  系数, 如果  $\alpha$  系数  $< 0.6$ , 一般认为内部一致信度不足;  $0.7 < \alpha$  系数  $< 0.8$  时表示量表具有相当的信度;  $0.8 < \alpha$  系数  $< 0.9$  时说明量表信度非常好<sup>[5-6]</sup>。单个评委评分与选手最终得分之间的一致性程度用皮尔逊积差相关系数衡量, 皮尔逊积差相关系数是用来反映2个变量线性相关程度的统计量, 绝对值越大表明相关性越强。在本研究中把各个选手的最终得分作为效标, 计算各个评委与选手有效成绩的皮尔逊积差相关系数, 得到该位评委的效标校度, 即评价质量, 相关系数越高, 说明该评委的评分越准确。

**1.5 统计学处理** 应用 SPSS 16.0 软件进行数据分析, 计算肯德尔和谐系数  $W$ 、克朗巴赫  $\alpha$  系数以及皮尔逊积差相关系数,  $P < 0.05$  为差异有统计学意义。

### 2 结果

**2.1 肯德尔和谐系数  $W$**  结果见表1。6组评委间的肯德尔和谐系数  $W$  均在0.01水平上显著一致, 其中5组评委的肯德尔和谐系数  $W$  在0.001水

DOI: 10.7683/xyxyxb.2016.06.030

收稿日期: 2016-02-29

基金项目: 河南省教育厅五年制临床医学人才培养模式改革项目(编号: 2014JGLX048)。

作者简介: 徐红彦(1978-), 男, 河南许昌人, 硕士, 讲师, 研究方向: 高等教育评价和教育政策研究。

通信作者: 郭志坤(1956-), 男, 河南商丘人, 博士, 教授, 博士研究生导师, 研究方向: 高等教育人才培养和教育政策研究; E-mail: gzk@xxmu.edu.cn。

平上显著一致,表明评分者之间的一致性程度很高。

表1 各组评委评分的肯德尔和谐系数 W

组别	肯德尔和谐系数 W	$\chi^2$	P
1 组	0.660	51.090	<0.001
2 组	0.565	31.150	<0.001
3 组	0.355	62.985	<0.01
4 组	0.479	29.080	<0.001
5 组	0.410	47.640	<0.001
6 组	0.658	36.120	<0.001

2.2 克朗巴赫  $\alpha$  系数 本次6组比赛中,各组评委的克朗巴赫  $\alpha$  系数分别为 0.894、0.861、0.676、

表2 单个评委评分与选手最终得分的相关系数

组别	教师评委				学生评委		
	1	2	3	4	1	2	3
1 组	0.617 <sup>a</sup>	0.707 <sup>b</sup>	0.778 <sup>b</sup>	0.847 <sup>b</sup>	0.912 <sup>b</sup>	0.845 <sup>b</sup>	0.933 <sup>b</sup>
2 组	0.800 <sup>b</sup>	0.586 <sup>a</sup>	0.632 <sup>b</sup>	0.840 <sup>a</sup>	0.753 <sup>b</sup>	0.833 <sup>b</sup>	0.916 <sup>b</sup>
3 组	0.003	0.169	0.549	0.717 <sup>a</sup>	0.952 <sup>b</sup>	0.910 <sup>b</sup>	0.840 <sup>b</sup>
4 组	0.516	0.788 <sup>a</sup>	0.453	0.843 <sup>b</sup>	0.783 <sup>a</sup>	0.456	0.861 <sup>b</sup>
5 组	0.173	0.879 <sup>b</sup>	0.250	0.572 <sup>a</sup>	0.886 <sup>b</sup>	0.679 <sup>b</sup>	0.937 <sup>b</sup>
6 组	0.880 <sup>b</sup>	0.723 <sup>b</sup>	0.766 <sup>b</sup>	0.798 <sup>b</sup>	0.711 <sup>b</sup>	0.836 <sup>b</sup>	0.747 <sup>b</sup>

注: <sup>a</sup>  $P < 0.05$ , <sup>b</sup>  $P < 0.01$ 。

### 3 讨论

本研究通过对第9届中青年教师教学竞赛结果进行信度检验,验证竞赛的评价工作是否可信可靠。本次竞赛评委有教师评委和学生评委,其中,教师评委由教学经验丰富、学术水平高、办事公道的专家教授组成,他们的优点是评分经验丰富,对比赛的节奏掌握较好,缺点是可能存在打人情分的因素。另外,教学竞赛的内容涉及各个学科的专业问题,而教师评委只能覆盖少数几个专业,由于专业的差异,对讲课内容的理解程度和欣赏角度不同,打分时可能会出现较大的差别<sup>[6]</sup>。本次教学竞赛中,教师评委中共有7人次出现了相关性较差,这和比赛中教师评委的删除人次较多相一致。将教师评委和学生评委对各组选手的平均评分进行排序,分别与选手的实际排名顺序进行比较,检验2组评委对特定选手的排名差异,结果发现,教师评委的表现对前3名选手的排名没有影响,但是,对获得优秀奖选手的排名有不同程度的影响。

本次教学竞赛引入了学生评委,学生是教师授课的对象,他们能够真实反映教师的授课效果,评价较客观,可以避免人情分的因素,缺点是评分缺乏经验。对此,采取以下校正措施:(1)要求所有学生评委的考试平均成绩位列本专业前20名并至少获得过1次学校奖学金,以保证他们掌握了必要的专业知识;(2)在比赛之前对学生评委和教师评委一起进行严格培训,详细介绍评分标准,统一扣分原则和

0.762、0.726、0.890,其中3组克朗巴赫  $\alpha$  系数 > 0.8,2组 > 0.7,1组 > 0.6,这也印证了肯德尔和谐系数 W 的结论,即评委们之间的评分一致性很高,本次教学竞赛评委的评价工作具有较高的信度。

2.3 皮尔逊积差相关系数 结果见表2。本次比赛评分的所有相关系数中,教师评委有7人次相关性较差,学生评委有1人次相关性较差,其余34位评委的评分均与选手比赛成绩呈显著正相关,说明大部分评委的评分比较可靠。

尺度;(3)把4个学生评委的平均分作为1个有效评分,以有效弥补学生评委评分经验的不足,提高学生评委的评分质量。本研究结果证实,引入学生评委是非常成功的尝试,有利于客观真实地反映选手的授课效果。

本研究结果表明,新乡医学院第9届中青年教师教学竞赛的评委评价工作具有较高的质量,本次教学竞赛结果的信度检验为今后进一步优化教学竞赛程序和方法、制定更加合理的评判标准提供了参考依据。

#### 参考文献:

- [1] 张忠华,李婷婷.论大学青年教师的教学能力结构与发展策略:基于实证的结论与逻辑的分析[J].中国高教研究,2013(4):51-55.
- [2] 徐扬,张清,徐辰武.提高课堂教学质量的积极探索与期待:以扬州大学2011年中青年教师讲课比赛为分析对象[J].扬州大学学报(高教研究版),2012,16(5):73-77.
- [3] 马天梅.中青年教师课堂教学竞赛的信效度检验[J].中国高等医学教育,2006(2):40-42.
- [4] 高普梅,杨世昌.无聊倾向性量表在大学生中的信度效度分析[J].新乡医学院学报,2014,31(2):114-116.
- [5] 王孝玲.教育统计学[M].上海:华东师范大学出版社,2001:256.
- [6] 徐翠霞,樊小东.关于高校讲课竞赛评审机制的思考:由青年教师讲课竞赛数据引发[J].黑龙江教育(高教研究与评估版),2012(8):52-54.

(本文编辑:张艳丽)