

● 教育理论

基于概化理论的医科院校教师教学水平 学生评价量表信度分析

孟琼¹ 陈平² 王萍³ 陈莹¹ 何利平¹ 喻箴¹

(1. 昆明医科大学公共卫生学院, 云南 昆明 650031; 2. 云南省健康教育所, 云南 昆明 650018;
3. 厦门大学公共卫生学院, 福建 厦门 361005)

【摘要】 通过计算概化系数和可靠性指数来评价医科院校教师教学水平学生评价量表的信度。结果发现: 教学内容、教学方法、教学态度、教学效果领域信度较好, 教学组织领域信度尚可, 若要修改, 可考虑增加教学组织的条目数, 减少其他领域的条目数。

【关键词】 概化理论; 教学水平; 量表; 信度

DOI: 10. 3969/j. issn. 1002-4701. 2016. 08. 008

【中图分类号】 G451 **【文献标识码】** A

【文章编号】 1002-4701(2016)08-0017-02

可信可靠的评价工具是整个教学评价的前提, 本研究组前期制定出了《医学高校教师教学水平学生评价表》^[1], 并应用经典测量理论对其信度进行了评价。概化理论 (Generalizability Theory, GT) 是一种现代测量理论, 被广泛应用于教育领域, 相对于经典测量理论^[2]。其优势在于可以具体地提出条目数达到多少时可以有较好的信度, 为改进评价量表的条目设置提供依据。因此, 本文旨在报道如何采用概化理论的方法来分析自制的《医学高校教师教学水平学生评价表》的信度, 并提出了一些改进量表条目设置的建议。

一、材料与方法

1. 测量工具。测量所采用的自制评价量表是由 24 个条目组成, 条目被划分为 5 个领域, 分别为教学组织 (3 个条目)、教学内容 (7 个条目)、教学方法 (5 个条目)、教学态度 (5 个条目)、教学效果 (4 个条目), 每个条目按“一点也不、有一点、有些、相当、非常”分为 5 个等级, 并分别记作 1、2、3、4、5 分。

2. 测量目标、测量侧面、调查人群。测量目标: 学生评价的昆明医科大学公共卫生学院一名教师的教学水平。测量侧面: 《医学高校教师教学水平学生评价表》的条目。调查人群: 昆明医科大学由该教师授课的不同专业、年级的 3 个班的全体学生。设计方案: 采用随机单面交叉设计, 记作 $p \times i$ 。随机, 是指假定学生和条目均是随机抽取的。单面, 是指只有一个测量侧面即评价量表的条目, 因为仅对一名教师的教学水平进行评价, 教师这个测量侧面固定了。交叉设计, 就是测量目标与测量侧面之间的一一碰面的相互关系, 即每个学生均要回答每一个条目。

3. 概化理论研究过程。概化理论研究过程分为 G 研究 (Generalizability Study, 概化研究) 和 D 研究 (Decision Study, 决策研究)^[2]。G 研究是在一定的测量情境下进行测量设计, 尽可能多地挖掘出各种潜在的测量误差来源, 并估计

出他们的变异分量大小。D 研究是以 G 研究得到的变异分量估计值为基础, 先计算测量侧面样本含量与 G 研究相同时的相对误差、绝对误差、概化系数和可靠性指数等, 再改变测量侧面 (条目) 的样本容量, 重新计算相对误差、绝对误差、概化系数和可靠性指数等, 进而估计出条目数量对测量误差和信度指标的影响。

4. 统计学方法。应用 SPSS 17.0 软件中 General linear Model 菜单的 UNIANOVA 命令, 计算出由学生解释的方差、由条目解释的方差及两者交互作用解释的方差, 分别用 $\sigma^2(p)$ 、 $\sigma^2(i)$ 、 $\sigma^2(pi)$ 表示, 然后在计算以上三者的变异分量: 学生解释的变异分量 $\sigma^2(P)$ 等于 $\sigma^2(p)$, 条目解释的变异分量 $\sigma^2(I)$ 等于 $\sigma^2(i)$ 除以条目数, $\sigma^2(PI)$ 交互作用解释的变异分量 $\sigma^2(PI)$ 等于 $\sigma^2(pi)$ 除以条目数。相对误差就等于交互作用解释的变异分量 $\sigma^2(PI)$, 而绝对误差等于条目解释的变异分量 $\sigma^2(I)$ 和交互作用解释的变异分量 $\sigma^2(PI)$ 之和。概化系数 E_p^2 是由条目解释的变异分量占其与相对误差之和的比例, 可靠性指数 Φ 是由条目解释的变异分量占其与绝对误差之和的比例^[2]。

二、结果

1. G 研究结果。在 G 研究, 分别对教学组织、教学内容、教学方法、教学态度、教学效果等 5 个领域计算了各变异来源的方差分量及其占总变异的百分比。5 个领域中, 变异的最大来源是学生侧面 (见表 1)。

2. D 研究结果。按目前该量表的条目数, 教学方法、教学内容、教学态度、教学效果等 4 个领域的概化系数和可靠性指数均大于 0.8, 只有教学组织领域的概化系数为 0.781, 可靠性指数为 0.750。若要概化系数和可靠性指数均大于 0.8, 需将教学组织的条目数增加至 4 个 (见表 2)。

三、讨论

信度主要是指测量结果的可靠性、一致性和稳定性, 即测验结果是否反映了被测者的稳定的、一贯性的真实特

表 1 G 研究中 p × i 设计量表 5 个领域的变异分量估计及所占总变异的百分比

领域	学生侧面 (p)		条目侧面 (i)		学生和条目间交互作用 (p × i)	
	变异分量	百分比 (%)	变异分量	百分比 (%)	变异分量	百分比 (%)
教学组织	0.261	50	0.042	8	0.219	42
教学内容	0.132	55	0.008	3	0.099	42
教学方法	0.300	63	0.008	2	0.170	35
教学态度	0.141	56	0.002	1	0.107	43
教学效果	0.247	61	0.017	4	0.140	35

表 2 D 研究中 p × i 设计不同条目数下量表 5 个领域的概化理论系数和可靠性指数

领域	条目数	$\sigma^2(P)$	$\sigma^2(I)$	$\sigma^2(PI)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$	E_p^2	Φ
教学组织	2	0.261	0.021	0.110	0.110	0.131	0.704	0.666
	3	0.261	0.014	0.073	0.073	0.087	0.781	0.750
	4	0.261	0.011	0.055	0.055	0.065	0.827	0.800
教学内容	2	0.132	0.004	0.050	0.050	0.053	0.728	0.713
	3	0.132	0.003	0.033	0.033	0.036	0.800	0.788
	7	0.132	0.001	0.014	0.014	0.015	0.903	0.897
教学方法	2	0.300	0.004	0.085	0.085	0.089	0.779	0.771
	3	0.300	0.003	0.057	0.057	0.059	0.841	0.835
	5	0.300	0.002	0.034	0.034	0.036	0.898	0.894
教学态度	3	0.141	0.001	0.036	0.036	0.036	0.798	0.795
	4	0.141	0.000	0.027	0.027	0.027	0.840	0.838
	5	0.141	0.000	0.021	0.021	0.022	0.868	0.866
教学效果	2	0.247	0.008	0.070	0.070	0.078	0.779	0.759
	3	0.247	0.006	0.047	0.047	0.052	0.841	0.825
	4	0.247	0.004	0.035	0.035	0.039	0.876	0.863

注: $\sigma^2(\delta)$ 相对误差, $\sigma^2(\Delta)$ 绝对误差, E_p^2 概化系数, Φ 可靠性指数
征^[3]。G 研究结果发现: 五个领域测量结果变异的最大来源均是被试者效应, 被试者效应和被试者与条目的交互效应已经解释了得分总变异的 90% 以上, 这是比较合理的也正是我们期望的。

有学者提出^[4] 概化系数大于等于 0.8 信度为好, 概化系数在 0.4 到 0.8 之间信度尚可, 概化系数在 0.4 及以下信度较差。若按此标准, 研究结果中教学内容、教学方法、教学态度及教学效果的信度较好, 教学组织的信度尚可。考虑到在填表时条目数越多, 学生所耗费的时间越长, 容易产生随意填写或对后面条目的填写失去耐心, 这样反而会降低量表信度, 因此, 建议保证一定可信度的情况下适当减少量表条目数。按概化系数大于 0.8 的标准来进行选择, 对各领域条目数可如下调整: 教学组织由 3 个增至 4 个; 教学内容由 7 个减少至 3 个; 教学方法由 5 个减少至 3 个; 教学态度由 5 个减少至 4 个; 教学效果由 4 个减少至 3 个。

[参考文献]

[1] 孟 琼, 张美霞, 陈 莹, 等. 医学高校教师教学水平学生评价表的信度分析[J]. 卫生软科学, 2016, 30(7): 46-48.
[2] 杨志明, 张 雷. 测评的概化理论及其应用[M]. 教育科学出版社, 2003: 48-76.
[3] Bachman, Lylef. Fundamental Considerations in Language Testing [M]. Oxford: Oxford University Press, 1990: 177-181.
[4] Winterstein B P, Willse J T, Kwapil T R, et al. Assessment of score dependability of the Wisconsin Schizotypy Scales using generalizability analysis [J]. Journal of Psychopathology and Behavioral Assessment, 2010, 32(4): 575-585.

[收稿日期] 2015-07

[作者简介] 孟 琼, 女, 在读博士生, 副教授, 主要研究方向: 生命质量量表研制, 概化理论的应用, 卫生筹资公平性。

[基金项目] 昆明医科大学教研改项目(2013JY-Y39)。

A reliability study of students' rating on teachers' teaching on the basis of the generalization theory

Meng Qiong¹, Chen Ping², Wan Ping³, et al

(1. Department of Epidemiology and Health Statistics, School of Public Health, Kunming Medical University, Kunming 650031, Yunnan, China; 2. Institute of health education in Yunnan province, Kunming, Yunnan 650018, China; 3. School of Public Health, Xiamen University, Xiamen 361005, Fujian, China)

Objective: This study uses the generalization theory to check the reliability of students' rating on teachers' teaching on the basis of the generalization theory. While the findings indicate good reliability in the areas of teaching content, method, attitude and outcome, the reliability of teaching arrangement indicates changes need to be made probably by increasing questions in this survey item.

Key Words: Generalizability theory; Teaching level; Scale; Reliability