

高校本科专业评估指标体系的信度与效度研究

王庆辉¹, 杨荣昌¹, 孙莱祥²

(1. 上海商学院 高等教育研究所, 上海 201400; 2. 复旦大学, 上海 200433)

摘要:文章采用上海市两所高校123个专业的专业评估实测数据,考察上海市高校本科专业达标评估指标体系的信度和效度。通过克朗巴赫(Cronbach's) α 系数、相关分析和因子分析(factor analysis)等方法检验指标体系的信度和效度。数据分析结果表明,本科专业达标评估指标体系的克朗巴赫(Cronbach's) α 系数为0.828;20项指标得分都与总分的相关系数在0.01的水平上显著相关;因子分析67.66%的总方差可以由7个因子解释。该本科专业达标评估指标体系具有较好的信度与效度,适合作为高校本科专业评估的测量工具。

关键词:本科专业;评估指标体系;信度分析;效度分析

中图分类号:G647

文献标志码:A

文章编号:1003-2614(2017)07-0059-04

《国家中长期教育改革和发展规划纲要(2010-2020年)》指出,“提高质量是高等教育发展的核心任务,是建设高等教育强国的基本要求”,要“健全教学质量保障体系,改进高校教学评估”^[1]。《上海市中长期教育改革和发展规划纲要(2010-2020年)》指出,要“制定教育质量标准,建立健全科学、多元的教育评价体系”^[2]。因此,建立健全质量保障体系并推动其有效运行,是高校人才培养过程中应有且重要的环节。

专业是人才培养的基本单位,专业教学质量直接关系到高校的教育教学质量和社会声誉。从专业人才培养的视角看,高校质量保障体系可以分为大循环系统和小循环系统两类。高校评估可以看作质量保障体系的大循环系统,而小循环系统中的专业评估是专业建设过程中的自循环评估,即在完成最初的新专业评估之后,定期接受或主动对自身的建设状况进行评估。专业评估制度已经成为学校质量保障体系最基础、最重要的组成部分。

我国高校的本科专业评估工作尚无统一的评价标准,各地根据各自的情况从不同侧面和角度去开展本科专业评估。鉴于这种情况,建立一个综合的评价指标体系显得越发重要。上海市教育评估院课题组在认真研究国家和上海的相关文件精神^[3]的基础上,通过文献调研和广泛征求专家意见,采用改进的德尔菲法(Delphi method),建立上海高校本科专业达标评估指标体系。该指标体系共设计了6个一级指标,即专业定位与培养目标、教师队伍、教学资源、专业教学、教学管理、教学效果;20个二级指标(见表1)和46个观

测点。本研究拟检验该指标体系的信度和效度,为其推广应用提供科学依据,同时为建立科学的专业评估指标体系提供方法借鉴^[3]。

表1 高校本科专业达标评估指标体系

一级指标	二级指标
1. 培养目标与培养方案	1.1 专业定位与人才培养目标
	1.2 培养方案
	1.3 课程体系
	2.1 数量与结构
	2.2 教学工作
	2.3 科研情况
2. 教师队伍	2.4 培养培训
	3.1 实验室与实习基地
	3.2 图书资料
	3.3 教学经费
4. 专业教学	4.1 课程教学
	4.2 实践教学
	4.3 教学改革
	4.4 毕业设计(论文)
5. 教学管理	5.1 组织机构及规章制度
	5.2 学生服务
	5.3 质量监控
6. 教学效果	6.1 学风
	6.2 人才培养目标实现情况
	6.3 就业与社会评价

一、对象与方法

(一) 研究对象

本次研究以上海市两所本科院校的123个本科专业作为评价对象。数据来源为专家先根据资料查阅、听取汇报、

收稿日期:2017-05-09

基金项目:教育部人文社会科学青年项目“中美研究型大学系主任领导效能及其影响因素研究”(编号:15YJC880087);上海市哲学社会科学一般项目暨教育科学规划重点项目“美国公立研究型大学教师评议会制度对上海高校的启示研究”(编号:A1208);上海商学院教师教学发展中心课题“基于中本贯通的教学质量保障研究”(编号:15JFBS07)。

作者简介:王庆辉,上海商学院高等教育研究所副研究员,管理学博士,研究方向:比较高等教育、高等教育管理;杨荣昌,上海商学院教学质量监控与评估中心,高等教育研究所教授,教育学博士,研究方向:高等教育管理;孙莱祥,复旦大学教授,研究方向:高等教育管理、数学研究。

教师访谈、学生座谈、现场观察、工作记录查看等情况而进行评分,再通过专家分组讨论形成专业的最后得分。指标体系的信度和效度以专家对各专业的打分为基础。本次研究采用的指标体系包括一级指标6个、二级指标20个。

(二) 信度和效度的检验方法

1. 信度评价

信度包括内在信度与外在信度两类。本研究主要测量指标体系的内在信度,即各个项目之间是否具有一致性。具体方法是通过SPSS19.0软件“分析—度量—可靠性分析”命令计算克朗巴赫(Cronbach's)α系数来对得分情况进行可靠性分析^[4]^[5]。

2. 效度评价

效度通常可分为内容效度、结构效度和效标效度等三类^[6]。本研究主要运用的是前两类效度分析:内容效度是指所设计的指标项能否代表所要测量的内容或主题;结构效度是指测量结果体现出来的某种结构与测值之间的对应程度^[7]。

内容效度的测量一般可通过计算单项指标得分与总分的相关系数的方法实现,通过运用SPSS19.0中“分析—相关”路径的方法来计算单项指标得分与总分的皮尔逊相关系数来判定结果是否显著^[8]^[9]。

测量结构效度主要是通过因子分析法提取公因子,考察提取出的公因子是否与问卷假设中的结构一致。如果一致,则表明公因子较好地反映了问卷结构。我们也是通过SPSS19.0中的“分析—降维—因子分析”方法来实现结构效度的分析^[10]^[11]。根据因子的累积贡献率和因子个数的选择情况,累积贡献率反映公因子对量表的累积有效程度^[12]。

从因子的累积贡献率是否满足因子个数对累积贡献率的要求,选择若干因子。根据不同指标与相关因子的相关程度,可以组成不同模块,进而分析指标体系是否具有良好的结构效度。

二、结果

(一) 信度检验结果

对样本数据进行处理,得到的结果见表2。

表2 内部一致性信度(克朗巴赫α系数)

Cronbach's Alpha	基于标准化项的Cronbach's Alpha	项数
0.828	0.828	20

如表2所示,对于二级指标内部一致性信度系数的计算结果,基于标准化项的克朗巴赫α系数值为0.828,信度系数大于0.8,说明该指标体系具有较高的内在一致性,信度较好。

(二) 效度检验结果

1. 内容效度

在社会科学领域,往往是在实施前由研究者自己或聘请相关学者、专家依据一定的理论对内容效度进行判断。学者专家包括有实际工作经验者、与此相关研究经验者,以及有相关学术背景的学者等^[13]。

相关分析也是确定内容效度的方法之一。现采用单项与总和相关分析法对指标体系的内容效度进行分析。

对样本数据进行处理,得到的结果见表3。

表3 内容效度统计——项目与整体的相关系数

项目	Pearson相关性	显著性(双侧)	项目	Pearson相关性	显著性(双侧)
专业定位	0.700 **	0.000	课程教学	0.398 **	0.000
培养方案	0.720 **	0.000	实践教学	0.599 **	0.000
课程体系	0.424 **	0.000	教学改革	0.541 **	0.000
师资队伍	0.532 **	0.000	毕业设计	0.559 **	0.000
教学工作	0.247 **	0.006	规章制度	0.377 **	0.000
科研情况	0.528 **	0.000	学生服务	0.303 **	0.001
师资培训	0.501 **	0.000	质量监控	0.472 **	0.000
实验实习	0.412 **	0.000	学生学风	0.511 **	0.000
图书资料	0.410 **	0.000	目标实现	0.690 **	0.000
教学经费	0.348 **	0.000	就业评价	0.345 **	0.000

** 在 0.01 水平(双侧)上显著相关。

从表3我们可以看到,各项指标与总分之间基本上呈显著相关($p < 0.01$)表明该指标体系具有较好的内容效度。

2. 结构效度分析

现采用因子分析法对指标体系的结构效度进行适当性(Kaiser - Meyer - Olkin KMO)检验,得到的结果见表4。

如表4所示,KMO统计量为0.753($KMO > 0.7$)。原始变量适用因子分析,可放心进行因子分析^[14]。巴特利特(Bartlett)球形检验中相关系数矩阵与单位矩阵具有显著性差异($\chi^2 = 704.480$, $p < 0.01$)因子分析有效。

表4 结构效度统计——KMO 统计量和巴特利特(Bartlett) 球形检验

取样足够度的 Kaiser – Meyer – Olkin 度量。	0.753
Bartlett 的球形度检验	近似卡方 704.480
	df 190
	Sig. 0.000

表5 结构效度统计——因子分析总方差解释情况

成分	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	4.843	24.216	24.216	4.843	24.216	24.216
2	2.365	11.825	36.042	2.365	11.825	36.042
3	1.511	7.557	43.599	1.511	7.557	43.599
4	1.295	6.477	50.076	1.295	6.477	50.076
5	1.287	6.433	56.509	1.287	6.433	56.509
6	1.230	6.149	62.658	1.230	6.149	62.658
7	1.001	5.003	67.661	1.001	5.003	67.661

提取方法:主成分分析。

通过方差最大旋转后,得到二级指标的因子载荷矩阵(见表6)。因子1主要受课程体系、专业定位、教学改革、培养方案指标的影响,反映一级指标中培养目标与培养方案的要求;因子2主要受实验实习、实践教学、就业评价指标的影响,体现实践能力的要求;因子3主要受学生服务、学生学风、目标实现指标的影响,反映一级指标中教学效果的要求;因子4主要受师资队伍、师资培训、科研情况指标的影响,反映一级指标中教师队伍的要求;因子5主要受教学工作、课程教学、毕业设计

因子分析总方差解释情况表明因子的累积贡献率和因子个数的选择情况,累积贡献率反映公因子对量表的累积有效程度^[15]。按因子尽量少、特征值 α 大于1.0、累积贡献率至少达到65%的原则,取前7个因子,其累积贡献率为67.661%(见表5)。

表6 结构效度统计——旋转成分矩阵

二级指标	成分						
	1	2	3	4	5	6	7
课程体系	0.722						
专业定位	0.628						
教学改革	0.619						
培养方案	0.605						
实验实习		0.853					
实践教学		0.755					
就业评价		0.477					
学生服务			0.845				
学生学风			0.615				
目标实现			0.457				
师资队伍				0.632			
师资培训				0.618			
科研情况				0.549			
教学工作					0.665		
课程教学					0.639		
毕业设计					0.627		
质量监控						0.774	
规章制度						0.668	
图书资料						0.549	
教学经费							0.861

提取方法:主成分。旋转法:具有 Kaiser 标准化的正交旋转法。a. 旋转在 9 次迭代后收敛。

三、讨论

本文对专业达标评估指标体系进行初步的信度、效度分析。本研究中该指标体系的克朗巴赫系数(Cronbach's) α 系

指标的影响,反映了一级指标中专业教学的要求;因子6主要受质量监控、规章制度、图书资料指标的影响,反映了一级指标中教学管理的要求;因子7主要受教学经费指标的影响。通过将因子分析结果和指标体系的设计结构进行对比,发现二者较为一致,可以认为该评价表具有良好的结构效度^[16]。分析结果表明,指标体系中个别指标的归类可做一些调整,如因子2中的实验实习、实践教学、就业评价等指标体现实践能力的要求,可归为实践教学模块。

指标体系的20项指标与总分的相关系数在0.01的显著

水平呈显著相关。该指标体系具有良好的内容效度。对指标体系 20 个二级指标进行主成分分析,提取的 7 个主成分的累积贡献率达 67.66%,说明结构效度较好,结构基本符合最初设想。

然而,指标体系结构效度分析主成分分析结果显示,指标体系中个别指标的归类可做一些调整,如因子 2 中的指标实验实习、实践教学、就业评价等指标体现实践能力的要求,可归为实践教学模块。此外,指标教学经费指标单独成为主成分,可能是在设计上存在问题,需要进一步研究。

注释:

①上海高校本科专业达标评估指标体系的确定依据主要有《上海市新专业检查的通知》(沪教委高〔2011〕19号)、《教育部关于合格评估的通知》(教高厅〔2011〕2号)、《教育部关于印发〈普通高等学校设置暂行规定〉》(教发〔2006〕18号)、《教育专业基础办学数据统计工作的通知》(沪教委〔2013〕56号)、《教育部关于普通高等学校本科教学评估工作的意见》(教高〔2011〕9号)、《上海市教育委员会关于开展本科专业评估工作的通知》(沪教委高〔2012〕72号)和《教育部关于提高本科教学质量的若干意见》(教高〔2012〕4号)等。

参考文献:

- [1] 教育部. 国家中长期教育改革和发展规划纲要(2010—2020 年)(2010—07—29) [EB/OL]. (2015—08—26). http://www.moe.edu.cn/srcsite/A01/s7048/201007/t20100729_171904.html.
- [2] 上海市教委. 上海市中长期教育改革和发展规划纲要(2010—2020 年)(2010—09—10) [EB/OL]. (2015—08—26). <http://www.shmec.gov.cn/html/xxgk/201009/>
- [3] 王芳,等. 社区卫生服务绩效评价指标体系的信度与效度评价[J]. 中国全科医学, 2008(1): 77—79.
- [4] [8] [美]罗纳德·约克奇. SPSS 其实很简单 [M]. 刘超, 吴铮, 译. 北京: 中国人民大学出版社, 2014: 62, 193.
- [5] [9] [11] 张怡婕, 鲍桂莲. 教学质量评价表的信度和效度分析——以北京联合大学为例 [J]. 人力资源管理, 2012(2): 109—110.
- [6] [10] 夏怡凡. SPSS 统计分析精要与实例详解 [M]. 北京: 电子工业出版社, 2010: 93, 101.
- [7] [13] [14] 杜智敏, 樊文强. SPSS 在社会调查中的应用 [M]. 北京: 电子工业出版社, 2015: 412, 413, 419.
- [12] 薛薇. SPSS 统计分析方法及应用 [M]. 北京: 电子工业出版社, 2014: 276.
- [15] 邱皓政. 量化研究与统计分析 [M]. 重庆: 重庆大学出版社, 2013: 303.
- [16] 张庆文, 马苓. 大学绩效评估指标系统的信度与效度分析 [J]. 河北大学学报: 哲学社会科学版, 2009(3): 88—92.
- [17] 吴明隆. SPSS 统计应用实物——问卷分析与应用统计 [M]. 北京: 科学出版社, 2003: 109.
- [18] 贺鹭, 等. 婴儿康复远期效果评价指标体系的信度和效度分析 [J]. 中国卫生统计, 2014(12): 955—957.
- [19] Fekkes M, Theunissen NCM, Brugman E, et al. Development and psychometric evaluation of the TAPQOL: A health-related quality of life instrument for 1—5—year-old children [J]. Qual Life Res, 2000(9): 961—972.
- [20] Connolly MA, Johnson JA. Measuring quality of life in pediatric patients [J]. Pharmacoeconomics, 1999(16): 605—625.

Reliability and Validity of Valuation Indicator System for the Undergraduate Programs

WANG Qing-hui¹, YANG Rong-chang¹, SUN Lai-xiang²

(1. Institute of Higher Education, Shanghai Business School, Shanghai 201400, China;
2. Fudan University, Shanghai 200433, China)

Abstract: In this study, the professional evaluation of 123 specialties in two universities from Shanghai was used to examine the reliability and validity of the undergraduate evaluation indicator system of undergraduate education in Shanghai. The reliability and validity of the index system were tested by means of Cronbach's alpha coefficient, correlation analysis and factor analysis. The result of data analysis shows that Cronbach's alpha coefficients of the second class indicators were 0.828; The scores of the 20 indexes were significantly correlated with the coefficient of total score at the level of 0.01. The factor analysis of the second class indicators showed that 7 factors were extracted, the accumulative variance reached 67.661%. The evaluation indicator system has good reliability and validity, which can be used for evaluating the undergraduate programs.

Key words: undergraduate programs; evaluation indicator system; reliability; validity