

文章编号: 2095-1663(2020)01-0080-05 DOI: 10.19834/j.cnki.yjsjy2011.2020.01.12

博士学位论文评审指标体系可靠吗

——基于全数据的信度和效度研究

刘春荣^a, 郭海燕^b, 吴瀚霖^c

(北京师范大学 a. 心理学部; b. 学位办; c. 信息科学与技术学院, 北京 100875)

摘要: 论文评审是学位论文质量评价的重要环节。为了使评审结果能真实地反映博士学位论文的质量和水平, 必须首先保证论文评审指标体系的准确性、可靠性和有效性。信度和效度分析是验证指标体系准确性、可靠性和有效性的重要方法。文章利用北京师范大学五年博士学位论文匿名评审全数据量化结果, 对评审指标体系的信度和效度进行实证研究。结果表明, 目前广泛采用的评审指标体系具有良好的信度和效度, 能够真实地反映博士学位论文的质量和水平。

关键词: 博士学位论文评审; 指标体系; 信度分析; 效度分析

中图分类号: G643

文献标识码: A

博士学位论文质量是衡量博士生培养质量的核心指标。论文评审是对学位论文质量进行评价的重要环节, 虽然各高校具体实施办法和要求略有差异, 但均为学校依据《中华人民共和国学位条例》^[1] (简称《条例》) 中规定的博士学位授予标准, 采纳、设计或修订一套评审指标体系, 请同行专家对论文进行评审, 并就论文是否达到博士学位授予标准做出明确判断的过程。可见, 论文评审指标体系作为质量评价的工具, 其准确性、可靠性和有效性对评价结果的真实性具有决定作用。

一、研究背景

为提高对博士学位论文质量的监控力度, 从2013年开始, 北京师范大学(简称: 北师大)实施了全部博士学位论文由校学位办委托教育部学位与研

究生教育发展中心(简称: 学位中心)学位论文评审平台进行匿名评审。同时, 采纳了学位中心推荐的评审指标体系(见表1)。全国博士学位论文抽检以及委托学位中心进行论文评审的众多高校同样采用这套体系。可见, 该指标体系已广泛用于我国博士学位论文评审。

论文评审作为博士生培养过程的一种终结性评价, 对是否能授予博士学位具有重大决定意义。因此, 必须保证论文评审指标体系的准确性、可靠性和有效性, 必须保证评审结果能真实地反映博士学位论文的质量和水平。信度和效度分析是验证指标体系精确性和可靠性的重要方法。对此, 尚未见有相关的研究。因此, 有必要利用北师大五年博士学位论文匿名评审全数据量化结果, 对评审指标体系的信度和效度进行实证研究, 以确立该指标体系的公信力。

收稿日期: 2019-06-10

作者简介: 刘春荣(1981—), 女, 河北衡水人, 北京师范大学心理学部副部长, 助理研究员, 博士。

郭海燕(1962—), 女, 辽宁辽阳人, 北京师范大学研究生院副处长, 研究员。

吴瀚霖(1993—), 男, 黑龙江哈尔滨人, 北京师范大学信息科学与技术学院博士研究生。

基金项目: 中国学位与研究生教育学会重点课题“‘三位一体’研究生学位论文质量评价与监测体系研究”(编号: A2-2015Y0407-011)

二、研究方法

(一) 信度检验方法

一个好的教育测验首先必须保证测验的结果可靠,测验结果的可靠性称为测验的信度^[2]。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。内在信度反映的是调查问卷各个项目(问项)间的相关程度,这些项目应该反映同一独立概念的不同侧面。^[3]

因论文式测验的评分没有严格的评分标准,在这种情况下,我们可以采用克朗巴赫(Cronbach)创造的 α 系数检验信度^[2]。通过调用 SPSS 中的信度分析模块,计算 α 系数,对评审结果进行信度检验。

(二) 效度检验方法

效度即有效性,是一种测验能够正确地测量出它所测量的特性或功能的程度。因此,效度是教育测验中最基本、最重要的问题,是选择和评鉴教育测验质量的重要指标^[2]。效度分为三种类型:内容效度、准则效度和结构效度。效度检验有多种方法,不同方法检验结果反映效度的不同类型。

①内容效度又称表面效度或逻辑效度,它是指所测验的内容与预定要测的内容之间的一致性程度^[2]。对内容效度常采用逻辑分析与统计分析相结合的方法进行评价。逻辑分析一般由研究者或专家评判所选题项是否“看上去”符合测量的目的和要求。统计分析主要采用单项和总和及相关分析法获得评价结果,根据相关是否显著判断是否有效,相关系数越大,内容效度越高^[3]。②准则效度又称效标效

度,是指测验结果与作为准则的另一独立测验结果之间的一致性^[2]。若测验结果与准则测验结果相关显著,则该题项为有效题项。③结构效度指一种测验所能衡量到理论上期望的特征的程度。最理想的结构效度分析方法是利用因子分析测量量表的结构效度。其目的是想了解属于相同概念的不同测验项目是否如理论预测那样集中在同一公共因子里。其中因子负荷反映了测验项目对该概念的贡献,因子负荷值越大说明与该概念的关系越密切。^[4]

本文将从如上这三个方面分别对指标体系的效度进行检验。所有数据均通过 SPSS 23.0 进行统计分析。

(三) 统计分析数据来源

统计分析研究数据来源于北师大 2013~2017 年共计五年的全部博士学位论文评审结果。根据《北京师范大学学位授予工作细则》的规定,博士学位论文的评审由校学位办委托学位中心,采取匿名(隐去导师名和评审专家名)的方式,按照博士生的学科、专业,以及学校优先选择高校的学科层次,聘请三位与论文选题相关专业领域的专家评阅。同时,采纳了学位中心推荐的评审指标体系(见表 1)。专家根据评审指标体系,从分评 1 到分评 4 按照“优秀、良好、一般、较差”的评价标准,分别给出 4 个分项评价指标的评定等级,并在此基础上做出总评评定结果。为便于定量分析,对“优秀、良好、一般、较差”4 类评价标准分别赋值为:10、8、6、4 分;只考虑首次申请评审且非增评的评阅意见,形成共计 3173 篇论文的 9519 份评审结果,也即 9519 条有效分析数据。

表 1 博士学位论文评审指标体系^①

评价指标	内涵	
分评 1	选题	选题的前沿性和开创性;研究的理论与现实意义;对国内外该选题及相关领域发展现状的归纳、总结情况。
分评 2	创新性 及 论文价值	对学科发展的贡献,对解决社会发展中重要理论问题和现实问题的作用(文科);对有价值现象的探索、新规律的发现、新命题和新方法的提出等作用(理科);对解决人文社会科学、自然科学或工程技术中重要问题的作用;论文及成果对文化事业发展或科技发展与社会进步的影响和贡献。
分评 3	基础知识及 科研能力	论文体现的学科理论基础坚实宽广程度和专门知识系统深入程度;论文研究方法的科学性,引证资料的翔实性;论文所体现的作者独立从事科学研究的能力。
分评 4	规范性	引文规范性、学风严谨性;论文结构的逻辑性;文字表述准确性、流畅性。
总评		论文的总体评价。

三、检验结果分析

学位论文评审指标体系的信度和效度分析就是考查评审结论反映学位论文真实学术水平的程度。

(一)信度分析

论文评审指标体系的信度描述的是评审结果的可靠程度,包括内在信度与外在信度两类。本研究主要检验论文评审指标体系的内在信度,即各个分项评价指标之间是否具有 consistency。采用 Cronbach's α 系数评价指标体系的信度。计算公式为:

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum S_2_i}{S_2_x} \right)$$

其中, α 为信度系数; K 为指标体系中分项的总数,在此, $K=4$; S_i 为第 i 项得分的项内方差($i=1, 2, 3, 4$), S_x 为全部分项总得分的方差。

从公式中可以看出,Cronbach's α 系数评价的是量表中各分项评价指标之间的一致性,属于内在一致性系数。 α 系数值介于 0 与 1 之间, α 值越高,表明分项之间的一致性越强,内部一致性可信度越高。一般认为 α 在 0.7~0.8 之间表示信度相当好,在 0.8~0.9 之间表示信度非常好^[3]。

经计算,本研究中学位论文评价指标体系的 Cronbach's α 系数结果为 0.828,说明各分项评价指标的内部一致性极好,指标体系的信度是非常令人满意的。

更进一步,还可以评估每个单项指标的信度,以便识别那些与整个指标体系关联性不大的分项,为进一步优化指标体系提供统计理论依据。计算删除 4 个指标中某一个分项后的 Cronbach's α 系数,结果见表 2。可见,删去 4 个分项评分中任意之一, α 系数均变得小于 0.828,说明 4 个分项指标每一个都是可信的,每一个与总体的关联性都较大。

表 2 带已删除变量的 Cronbach's α 系数

序号	删除变量	α 系数
1	选题	0.804133
2	创新性 & 论文价值	0.771311
3	基础知识 & 科研能力	0.753952
4	规范性	0.796957

(二)效度分析

论文评审指标体系的效度描述的是评审结果反映论文质量的准确程度。效度是衡量指标体系最重

要的因素,直接影响指标体系的价值。效度分为三种类型:内容效度、准则效度和结构效度。

1. 内容效度分析

论文评审指标体系的内容效度就是指所设计的各项指标反映博士学位论文应达到的学术标准的程度。对内容效度采用逻辑分析与统计分析相结合的方法进行评价。

逻辑分析主要依据《条例》第六条对于博士学位授予的标准“(一)在本门学科上掌握坚实宽广的基础理论和系统深入的专门知识;(二)具有独立从事科学研究工作的能力;(三)在科学或专门技术上做出创造性的成果”,^[1]从逻辑推演的角度判断分析四项分评指标的内容究竟在体现博士学位学术标准方面达到多大程度。选题是论文价值的基础,是“做出创造性成果”的前提,所以,分评 1(选题)和分评 2(创新性 & 论文价值)两项的内涵准确地反映了博士学位授予标准的第(三)方面。规范性反映了研究生培养过程中对学术规范素养训练的程度,也是保障学位论文质量的基础知识和独立从事科研工作的基本能力。所以,分评 3(基础知识 & 科研能力)和分评 4(规范性)的内涵对应标准的第(一)、第(二)方面,总评是评审专家对论文整体水平的衡量。同时,还兼具了不同学科共性和特性的表达。因此,从逻辑分析的角度看,该指标体系达到了评价博士学位论文学术质量的目标,有理想的内容效度。

统计分析采用计算每个分评与总评的相关系数,根据相关是否显著判断是否有效,相关系数越大,内容效度越高。考虑到数据集为有序数据,所以考察变量相关关系时用 Spearman 秩相关系数,Spearman 秩相关系数定义见公式。

$$\theta = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \cdot \sum_{i=1}^n (S_i - \bar{S})^2}}$$

其中, R 与 S 分别为两个变量观测的秩。

相关系数 θ 计算结果见表 3。可以看出,总体评价与各分项评价相关系数在 0.63~0.75 之间,呈现明显正相关关系,表明指标体系的内容效度较高。

2. 准则效度分析

由于论文评审指标体系中的各分项指标与总体评价是相互独立做出的,因此可以选择总体评价为准则,分析指标体系中各分评与准则的相关性。若分评与总体评价相关显著,则该分评指标为有效选

项。所以,由上述相关分析结论——总体评价与各分项评价呈现明显正相关关系,也可得出准则效度也是较高的。

表3 Spearman 秩相关系数

指标	总体评价	选题	创新性	基础知识	规范性
总体评价	1				
选题	0.63173	1			
创新性 及 论文价值	0.75275	0.53695	1		
基础知识及 科研能力	0.74889	0.48044	0.58103	1	
规范性	0.64658	0.43083	0.46843	0.58239	1

同时,从相关分析还可以看出,各分项之间也均呈现显著正相关关系,相关系数的值在 0.43~0.58 之间。各分评之间的相关实际上标志着指标体系的内部一致性的信度,如果相关较高,表明各个分评之中有公共因子的存在,也可以作为内部一致性信度的证据。

3. 结构效度分析

论文评审指标体系的结构效度指评审指标的内涵能够体现出博士学位授予标准的程度。评审指标体系要具有较高的结构效度,应保证对于来自相同概念(科研能力或创新性)的分项指标是否如理论预测那样集中在同一个公因子中。利用因子分析法检验指标体系的结构效度。

因子分析的主要功能是从指标体系的全部分项指标中提取一些公因子,各公因子分别与某几个分评指标高度相关,因此这些公因子可代表指标体系的基本结构,通过因子分析可以考察研究者设计指标体系时假设的某种结构。在因子分析之前,首先进行适应性检验,测量 KMO(Kaiser-Meyer-Olkin) 值和并进行 Bartlett 球体检验。KMO 测度的值越高(接近 1.0 时),表明项目间的相关性越强,越适合进行因子分析。Bartlett 球体检验的目的是检验相关矩阵是否是单位矩阵。一般说来,显著水平值越小(<0.05)表明原始变量之间越可能存在有意义的关系,指标体系的结构效度越好。KMO 和 Bartlett 球体检验的结果见表 4。本研究 KMO 指标为 $0.787 > 0.5$,说明原始变量适合做因子分析。Bartlett 球形检验的显著性水平为 $0.000 < 0.001$,

说明变量间的相关矩阵与单位矩阵有显著性差异,因子分析的结果有效。

表4 KMO 和 Bartlett 球体检验的结果

KMO 取样适切性量数		0.787
Bartlett 的球形度检验	近似卡方	14027.087
	自由度	6
	显著性	0.000

对四个分项评价指标进行因子分析,采用主成分分析法提取因子,计算相关矩阵特征值,结果见表 5。

表5 相关矩阵特征值

成分	特征值	比例	累积贡献率
1	2.63973194	0.6599	0.6599
2	0.58036038	0.1451	0.8050
3	0.44087105	0.1102	0.9152
4	0.33903663	0.0848	1.0000

累计贡献率反映了公因子的选择个数对总方差的解释程度。按照累计贡献率大于 80% 的准则,前两个因子累计贡献率已达到 80.5%,所以,应该选取前两个公因子。

为了使因子负载便于解释,决定各分指标归属哪个因子,需要对因子负载矩阵进行旋转,通过方差最大旋转可以得到因子载荷矩阵(见表 6)。因子载荷反映了各分项指标对该概念的贡献,因子载荷值越大说明与该概念的关系越密切。对于因子 1,分评 3(基础知识及科研能力)、分评 4(规范性)贡献明显高于分评 1、2,所以,因子 1 体现出了“科研能力”的概念;对于因子 2,分评 1(选题)和分评 2(创新性 & 论文价值)贡献明显高于分评 3 与分评 4,所以,因子 2 体现出了“创新性”概念。

表6 因子载荷矩阵

分评指标	因子 1	因子 2
选题	0.20181	0.90115
创新性 & 论文价值	0.44481	0.73335
基础知识 & 科研能力	0.75558	0.44682
规范性	0.90547	0.20303

因子 1 解释了方差 1.6293549,因子 2 解释了方差 1.5907375。因此,两个公因子对 4 个变量的方差累计解释达到 3.22(见表 7),其累计贡献率能达到 80.5%。

从《条例》第六条博士学位授予的标准分析可见,三条标准之间存在着递进的逻辑关系,只有“(一)在本门学科上掌握坚实宽广的基础理论和系统深入的专门知识”,才可能“(二)具有独立从事科学研究工作的能力”,进一步才可能“(三)在科学或专门技术上做出创造性的成果”。即博士学位授予标准的核心就是“科研能力+创造性”。

表 7 两个因子对 4 个变量的方差解释

最终的公因子方差估计:总计=3.220092			
选题	创新性 及 论文价值	基础知识及 科研能力	规范性
0.85278785	0.73566278	0.77054032	0.86110138

由因子分析可见,来自相同概念“科研能力”的 2 个分项指标“基础知识和科研能力”“规范性”评审结果集中在因子 1 里;来自相同概念“创新性”的 2 个分项指标“选题”“创新性和论文价值”的评审结果集中在因子 2 里。这个结果与内容效度的逻辑分析结果也是完全吻合的。即博士学位论文评审指标体系的内涵能够很好地衡量到《条例》第六条博士学位授予标准,结构效度良好。

四、结论与讨论

博士学位论文评审结果是授予博士学位的终结

性依据。而科学完善的评审指标体系是保证评审结果真实地反映博士学位论文质量和水平的基础和前提。信度和效度分析是验证指标体系精确性和可靠性的重要方法,只有经过评估证明可信度和有效度高的指标体系才具有公信力和生命力。利用北京师范大学 5 年博士学位论文评审结果分析博士学位论文评审指标体系的信度和效度,结果表明,目前用于博士学位论文评审的评价指标体系具有良好的信度和效度,可以采纳并被广泛推广。

致谢:感谢北京师范大学统计学院张淑梅教授的指导。

注释:

① 见:《北京师范大学博士论文评阅书》

参考文献:

- [1] 全国人民代表大会常务委员会. 中华人民共和国学位条例[Z]. 2004-08-28.
- [2] 朱德全,宋乃庆. 现代教育统计与测评技术[M]. 重庆:西南师范大学出版社,1998.
- [3] 孙雅波,范厚明,刘益迎,等. 基于信度和效度分析的海运强国评价指标体系构建[J]. 上海海事大学学报, 2014(4): 26-31.
- [4] 曾五一,黄炳艺. 调查问卷的可信度和有效度分析[J]. 统计与信息论坛,2005,20(6):11-15.

On Reliability of Doctoral Dissertation Review Indicator System: A study based on reliability and validity of all data

LIU Chunrong^a, GUO Haiyan^b, WU Hanlin^c

(a. Faculty of Psychology; b. Office of Academic Degree Committee;
c. College of Information Science and Technology, Beijing Normal University, Beijing 100875)

Abstract: Dissertation evaluation is an important link in the dissertation quality evaluation. In order to ensure that the evaluation results can truthfully reflect the quality of the doctoral dissertation, we must ensure the accuracy, reliability, and validity of the dissertation evaluation indicator system and the reliability and validity analysis is an important method to verify the accuracy, reliability, and validity of the indicator system. The authors use the outcomes of the full data quantification of the anonymous doctoral dissertation evaluation at Beijing Normal University in the past five years to carry out an empirical study on the reliability and validity of the system. The results show that the evaluation indicator system widely applied is fairly reliable and valid, which can truly reflect the quality and standard of doctoral dissertations.

Keywords: doctoral dissertation review; indicator system; reliability analysis; validity analysis