

基于 SEM 的研究生学位论文质量评价研究

孔祥沛^{1,2}, 孙继红¹

(1. 南京航空航天大学 经济与管理学院, 江苏 南京 210016; 2. 江苏省教育评估院, 江苏 南京 210024)

摘 要 研究生学位论文质量评估是保证研究生培养质量的重要手段。以“论文选题与综述”“论文水平”和“论文写作”为指标构成的研究生学位论文质量评价指标体系, 基于结构方程模型, 通过验证性分析、效度和信度检验、复合效度检验等多种方法检验结构方程模型的拟合度。检验表明, 研究生学位论文质量评价指标体系是合理的, 能够有效、可信地对研究生学位论文质量进行评估。

关键词 研究生学位论文 质量评价 结构方程模型

中图分类号: G643.8

文献标识码: A

文章编号: 1671-4970(2008)04-0070-04

一、问题的提出

研究生学位论文质量是学位与研究生教育质量的重要组成部分。我国的学位与研究生教育评估从 20 世纪 80 年代中期就开始出现, 各种评估活动对保证我国学位与研究生教育质量发挥了重要作用。但总的来说, 由于缺乏完整的制度背景和相对完善的评估制度, 整个学位与研究生教育评估工作尚处于初级阶段。尽管如此, 相关教育管理部门仍然在加强学位与研究生教育评估的法规建设, 规范各类评估工作方面进行了探索和研究。例如, 由国务院学位委员会办公室、教育部研究生工作办公室在一系列调查和研究工作的基础上形成了《学位与研究生教育评估暂行规定》(征求意见稿), 教育部强调“建立健全研究生教育评估制度、不断完善质量保证体系, 形成有效的激励机制”等。另外, 随着研究生教育的迅猛发展, 越来越多的硕士研究生考上博士或从学校走向社会, 在一定程度上影响着博士生的培养质量和各行各业发展的速度, 研究生的质量已成为社会各界普遍关注的一个大问题, 而研究生培养的质量也就集中体现在其学位论文的水平上, 因此, 开展学位与研究生教育评估, 根据评估结果, 采取相应的措施, 是学位与研究生教育管理部门对研究生培养质量进行监控的重要手段之一。

在现有的各类学位与研究生教育评估中, 学位与研究生教育自我评估是重要评估方式, 也是国家评估和社会评估的基础, 是各学位授予单位科学、合理规划学科建设的主要手段, 是建立自我管理、自我

约束、自我发展机制的关键环节。教育部也提出相应要求“培养单位应开展经常性的自我评估工作, 及时发现和解决培养工作中出现的问题, 建立起具有自我完善功能的质量保证和监控机制”。因此, 从中国目前研究生学位论文评估实践看, 除了全国百篇优秀博士论文具有全国性评估标准外, 各个学校均制定了相应的评估办法, 各种评估方法差异大, 评估方法是否科学、合理等都有待进一步探讨。

目前学位论文评价普遍存在着以下几个方面的问题。首先, 一般学位论文的评价均采用定性评价的办法, 即主要以论文评阅书和答辩决议书的形式进行。由于各专家评阅标准与宽严程度并不一致, 因而往往导致同一篇论文的评价结果差异较大, 致使论文的质量优劣难辨, 例如清华大学对学位论文评估后发现, 现在的学位论文评阅意见书甚至答辩决议书对博士论文的评价普遍存在着有意拔高、用词雷同倾向, 致使从这些材料中难以判别学位论文水平的高低, 甚至对一篇优秀的博士论文与一篇基本合格的博士学位论文, 单纯从评语上难分伯仲^[1]。其次, 有的学校在论文评阅过程中, 采用定量的记分评价法, 但其所用方法只是一种简单的数学统计法, 因而其结果缺少一定的客观性和可靠性^[2]。再次, 在学位论文评价中还存在一个学科差异的问题, 对学位论文的要求和侧重点也就有所差别, 仅仅用简单的记分法和评阅法进行评价, 其结果往往难以横向比较^[3]。最后, 更重要的是无论是定性评价还是定量评价都没有对评价指标体系的科学性引起足够的重视, 而这个问题却是研究生学位论文质量

评估的核心和基础。本文即通过构建研究生学位论文质量的结构方程模型对现有的学位论文质量评价指标体系进行验证性分析和合理性评估。

二、研究生学位论文质量评价指标体系的构建

江苏省学位委员会、江苏省教育厅为了进一步加强研究生教育的培养工作,切实提高研究生教育质量,进一步加强学位论文管理工作,提高论文质量,从 2002 年开始对研究生培养单位的学位论文进行抽检评议。为了保证抽检评议结果的科学性,江苏省教育评估院还制定相应的‘研究生学位论文评议指标体系’(表 1)。

表 1 研究生学位论文评议指标体系

一级指标	二级指标	权重 (博/硕)
A1 论文选题与综述	B1 选题的理论意义或实用价值	0.1/0.1
	B2 对本学科及相关领域研究状态的了解与评述	0.1/0.1
A2 论文水平	B3 对基础理论、专门知识的掌握与运用	0.2/0.3
	B4 论文在理论、方法或解决生产实际问题上的创新	0.4/0.3
A3 论文写作	B5 独立开展科研的能力	0.1/0.1
	B6 论文的逻辑、结构、文笔及学风等	0.1/0.1

研究生学位论文评议指标体系共有 3 个一级指标和 6 个二级指标,其中一级指标是 :论文选题与综述、论文水平以及论文写作。

“论文选题与综述”下设二级指标为‘选题的理论意义或实用价值’和‘对本学科及相关领域研究状态的了解与评述’,前者是考察学位论文在学术方面的创新性和前瞻性,技术应用方面的先进性,以及研究课题是否具有较高的理论意义、学术水平和实用价值 ;后者主要考察占有和阅读的文献资料的数量和质量,判断研究者驾驭文献资料的能力,初步了解研究者对基础理论和专门知识的掌握情况。

“论文水平”下设二级指标为‘对基础理论、专门知识的掌握与运用’和‘在理论、方法或解决生产实际问题上的创新’,前者主要考察研究者在攻读学位过程中是否掌握基本的理论和专门知识,对硕士研究生而言,必须在本门学科上掌握坚实的基础理论和系统的专门知识,对博士研究生而言必须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 ;后者主要考察学位论文的创新性,与本科教育不同,研究生更强调的是培养研究生的创新能力,这也是研究生教育的根本任务。

“论文写作”下设二级指标为‘独立开展科研的能力’和‘论文的逻辑、结构、文笔及学风等’,前者是指研究生在导师的指导下,是否能够根据科研课题任务,拟定研究计划,根据学科属性,运用调查法、文献法等多种研究方法,搜集、整理各种信息,并对信息加以正确的分析、评价,确定研究的主攻方向和难点,提出开展研究的方法和途径,建立有关的假设、模型和计算方法,拟定试验方案等,并最终取得科研成果 ;后者则是基本的学术规范和规格等问题^①。

三、基于 SEM 的研究生学位论文质量评价指标体系的实证检验

结构方程模型 (Structural Equation Modeling, SEM) 是一种综合性统计分析技术和模型方法,通过把一系列假设的变量间的因果关系反映成统计依存模式的综合假设,反映这种因果关系的参数表示解释变量(可观测变量和潜在变量)对被解释变量(可观测变量和潜在变量)的影响(直接的或间接的)程度。通过把这种假设的关系转换成可检验的数学模型,SEM 提供了量化理论假设和检验理论模型的综合分析工具。在研究中,经常会有一些潜在变量,潜在变量是无法直接观测的、隐含的假设或理论变量(结构),一般缺乏明晰或精确的尺度,但往往可以通过一些可观测的指标(变量)来反映其结构。而 SEM 即是被用来研究这些潜在变量之间的联系以及这些潜在变量和相关指标之间的联系。潜在变量在模型中既可以是解释变量也可以是被解释变量^[4]。因此,结构方程模型可用来验证上述研究生学位论文质量评价指标体系。具体的,本部分将首先构建学位论文质量评估的结构方程模型,然后对该评价模型进行验证性分析,在此基础上验证模型的效度和信度,最后是对模型的复合效度分析,从 3 个方面来验证上述评价体系的合理性。

1. 学位论文质量评价的结构方程模型的构建与拟合结果

根据学位论文评价指标体系,学位论文质量的 3 个影响因素 :选题与综述、论文水平、论文写作都是潜变量,无法直接测量,因此必须为每个潜变量选择测量指标。3 个潜变量的测量指标为上述指标体系的 6 个二级指标。由此构建学位论文质量评价的结构方程模型如图 1 所示。

本课题对 2007 年授予硕士学位的论文进行了分层随机抽样,共抽取 110 份硕士学位论文,将每份硕士论文所得专家打分成绩输入 SPSS11.5 统计软件进行数据整理,并采用结构方程模型软件

① 对该指标体系的更详细介绍,参见江苏省教育评估院《研究生学位论文评议指标诠释(讨论稿)》。

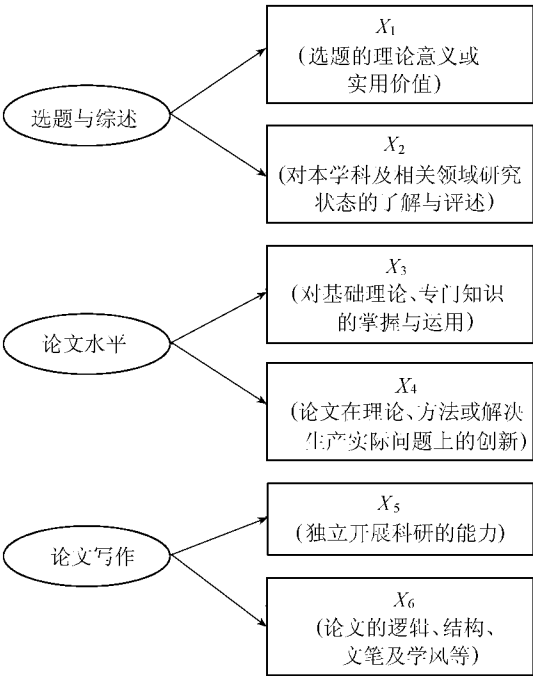


图1 学位论文质量评价的结构方程模型
LISREL8.70 对数据进行分析处理。

处理得到 3 个指标与其因子的相关矩阵(表 2),每个参数对应 3 个数值,第一个是(未标准化的)参数估计值,第二个是标准误,第三个是 T 值。由表 2 可见,因子负荷以及因子相关系数(协方差)都是显著的。进一步,考察结构方程模型的误差方差,结果表明,除了指标 5 的误差方差不大显著外(1.91,已经接近临界值),其余 5 个误差方差均显著。再通过 PHI 矩阵反映学位论文质量的 3 个方面因子(选题与综述、论文水平、论文写作)相关关系。根据上述结果,便可得到估计参数后的学位论文质量评价指标体系的模型路径图,如图 2 所示。

表 2 研究生学位论文评议指标体系

参数	KSI 1	KSI 2	KSI 3
VAR 1	0.76	-	-
	(0.09)		
	8.70		
VAR 2	0.88	-	-
	(0.08)		
	10.45		
VAR 3	-	0.91	-
		(0.07)	
		12.14	
VAR 4	-	0.93	-
		(0.07)	
		12.77	
VAR 5	-	-	0.97
			(0.07)
			13.37
VAR 6	-	-	0.84
			(0.08)
			10.63

注 表中每个参数对应 3 个数值,第一个是未标准化的参数估计值,第二个是标准误,第三个是 T 值。

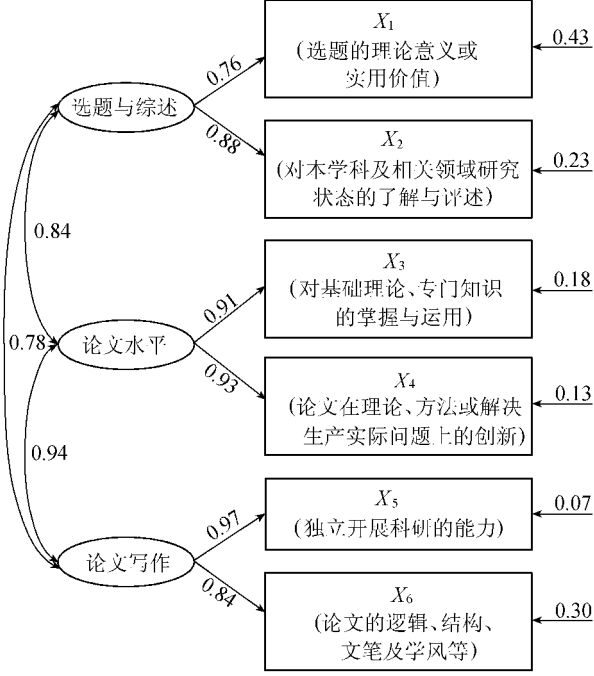


图2 学位论文质量评价的结构方程模型拟合结果

2. 学位论文质量评价模型的验证性分析

为了进行结构模型的验证性分析,学者们构建了不同的指数,包括卡方检验(Chi-Square)即 χ^2/df 、标准化的残差均方根(SRMR)、非正态拟合指数(NNFI)和比较拟合指数(CFI)、拟合优度指数(GFI)以及近似误差的均方根(RMSEA)等。这些指标的拟合结果如表 3 所示。结合表中各指标拟合值和各个指标的拟合标准对比分析可见,学位论文质量评价的结构模型拟合比较理想。

3. 学位论文质量评价模型的效度和信度分析

信度和效度指标是用来衡量各个因子的度量指标选择是否合理。信度是指测量和评价结果的稳定性或可靠性程度,也就是测量或评价结果是否真实、客观反映了被测量对象的真实水平。效度是指测量和评价结果的准确性和有效性的程度,也就是测量和评价是否达到预期的目的[5]。

利用结构方程模型来进行变量的效度分析首先要检查 3 个方面内容:其一,检查有无明显不合理的参数,如负方差、负测量误差、太大标准误等;其二,检查整体拟合指数是否达标;其三,需要进一步考察潜变量的构建效度。

根据上述 3 个方面检查,模型首先不存在明显的不合理参数,其次从表 2 的拟合指标结果看,学位论文评价的结构方程模型效度也是良好的。进一步重点考察第三方面。根据前述指标与因子矩阵,矩阵中如果 T 值大于 1.96,那么表示该参数已经达到 0.05 的显著性水平;如果 T 值大于 2.58,表示该参数已达到 0.01 显著性水平,显然表 2 结果可见,所有估计参数的标准差均很小,且显著,因此效度也是良好的。

表 3 研究生学位论文评议结构模型拟合指标

拟合指标	χ^2	df	χ^2/df	SRMR	NNFI	CFI	GFI	RMSEA
模型	14.78	6	2.46	0.026	0.97	0.99	0.96	0.082
拟合标准*	-	-	<5 可接受 <3 较好	越接近 0 越好		> 0.95	≥0.85	0 表示完全拟合 0.05 表示接近拟合 0.05 - 0.08 表示相当拟合 > 0.1 表示不相拟合

* 各个拟合标准均根据文献整理而得。

信度根据 LISREL 软件计算得到的每个观测指标的 R^2 可以测量。各个潜变量观测指标的信度如表 4 所示。由此可见,所有信度指标值都在 0.50 以上,这里的信度表示每个观测指标能够被潜变量解释的程度,如 X_1 (选题理论意义或实用价值)能够被‘选题及综述’解释 57%。

表 4 学位论文质量评估各潜变量观测指标的信度分析

因子	指标	标准化 载荷	测量 误差	R^2	组合 信度	平均方差 抽取量
选题及 综述	X_1	0.76	0.43	0.57	0.80	0.67
	X_2	0.88	0.23	0.77		
论文 水平	X_3	0.91	0.18	0.82	0.92	0.85
	X_4	0.93	0.13	0.87		
论文 写作	X_5	0.97	0.07	0.93	0.90	0.82
	X_6	0.84	0.30	0.70		

* 各个拟合标准均根据文献整理而得。

从以上对模型的信度和效度的分析可见,本文提出的研究生学位论文评价结构方程模型与数据拟合良好,说明整个模型的信度和效度都很高,表示模型具有很强的现实解释力和评估效果。

4. 学位论文质量评价模型的复合效度分析

所谓复合效度分析是在对整体模型拟合指数分析以及信度和效度评估基础上的进一步检验工作,主要是考察所构建的结构方程模型是否仅仅适用于所选择的特定样本还是对总体依然适用。国外学者从复合样本和模型数量两个纬度把复合效度划分为 4 类:模型稳定、模型选择、模型效度延展以及模型普遍化^[6]。

结构方程模型主要有两种方法完成这 4 种模型的复核:第一种方法是同多样本比较来考察,这要求有两个或者两个以上的样本方能进行;第二种方法是借助 ECVI 指标来考察。最好的方法是二者结合,即以 ECVI 值为标准,通过多样本比较来选择既有良好内部有效性,又有良好的外部推广能力的模型。一般而言,ECVI 值越小,模型的预测效度越好。实践中,拟合模型的 ECVI 值应当小于独立模型和饱和模型的 ECVI 值。

由于结构方程模型所需样本数量较大,虽然本研究有效样本 110 个,但分割为校验样本和复核样本,则单个样本的数量就不能很好地满足模型的拟合要求,所以采用第二种方法。通过 LISREL 软件分

析,模型的 ECVI 值为 0.41,而独立模型的 ECVI 值为 7.37,饱和模型的 ECVI 值为 1.39,由此可见模型的 ECVI 值小于饱和模型和独立模型的 ECVI 值,因此从复核角度分析,本文构建的学位论文质量评估模型是恰当的。

四、结论与建议

本文构建了研究生学位论文质量评价指标体系,并以 110 份硕士学位论文为样本,基于结构方程模型对模型的合理性和科学性进行了全面的评估。由‘论文选题与综述’、“论文水平”和‘论文写作’为一级指标;选题的理论意义或实用价值”、“对本学科及相关领域研究状态的了解与评述”、“对基础理论、专门知识的掌握与运用”、“论文在理论、方法或解决生产实际问题上的创新”、“独立开展科研的能力”、以及‘论文的逻辑、结构、文笔及学风等’为二级指标构成的研究生学位论文质量评价指标体系是合理的,能够有效、可信地对研究生学位论文质量进行评估。在未来的研究生学位论文质量评估中,可以考虑适当增加样本,从而能够将样本分割为校验样本和复核样本,从而能更好地进行复合效度分析。另外,由于学位论文质量的评价是一个典型的多因素综合评价问题,各个因素对评价结果的影响程度通常是人们根据主观判断来确定的,并且这一评价的结论还具有模糊的特性^[7],因此如何保证各个因素的评估的准确性有待进一步探讨。

参考文献：

[1] 刘颖,孙火斤,吴振一.博士学位论文评估的实践与思考[J].学位与研究生教育,2000(4):26-29.
[2] 王战军.学位与研究生教育评估技术与实践[M].北京:高等教育出版社,2000:28-45.
[3] 雷亚萍.研究生学位论文的模糊综合评价方法[J].西安科技学院学报,2003(9):331-334.
[4] 侯杰泰,温忠麟,成子娟.结构方程模型及其应用[M].北京:教育科学出版社,2004:15-17.
[5] 胡中锋,李芳.教育测量与评价[M].广州:广东高等教育出版社,1999:31-46.
[6] ADAMANFIOS D,JUDY A S.Introducing LISREL:A guide for the Uninitiated[M].Thousand Oaks,2000:130.
[7] 关志民,束军意,马钦海.学位论文质量的多层次模糊综合评价模型及其应用[J].科研管理,2005(5):153-157.