

博士学位论文双盲评阅风险分析

武小悦

(国防科学技术大学 信息系统与管理学院, 湖南 长沙 410073)

【摘要】 在现行博士学位论文评阅制度下, 博士论文双盲评阅主要面临两类风险: 一类是将合格的论文评审为不合格论文, 另一类是将不合格论文评审为通过。基于专家评阅水平和博士学位论文的质量水平, 建立了博士论文双盲评阅风险与评阅专家数量、评阅通过规则之间的关系模型。最后, 通过运用模型进行计算分析, 得出了不同假设情况下两类风险水平的变化规律。

【关键词】 博士学位论文; 评阅; 双盲; 风险分析

【中图分类号】 G643 **【文献标识码】** A **【文章编号】** 1672-8874 (2014) 03-0047-04

Risk Analysis for Double-Blind Review of Doctoral Dissertation

WU Xiao-yue

(College of Information System & Management, NUDT, Changsha 410073, China)

Abstract: Under current peer review mechanism for doctoral dissertation, two main risk types exist: one is the risk that a qualified dissertation is judged as unqualified. Another is that an unqualified dissertation is judged as qualified. In this paper, based on the quality level of expert review and the doctoral dissertation, a model is built for analyzing the relationship between the double-blind review risk, and the number of reviewing experts, rules for passing review. Finally, by numerical computation using the model, the risk level change of the two types of risk are obtained under different hypothesis conditions.

Key words: undergraduate; curriculum system; management engineering; Comprehensive university

一、引言

博士学位论文的水平是一个学校研究生教育和科学研究水平的重要标志。2012 年《教育部关于全面提高高等教育质量的若干意见》(教高〔2012〕4 号) 中要求, 加大博士学位论文抽检范围和力度, 每年抽查比例不低于 5%。2000 年教育部开始了对博士学位论文质量随机抽查工作。为了保证博士学位论文的质量和水平, 我国各高校已普

通采用了双盲评阅机制^[1] (即, 在论文送给评阅人时隐去论文作者与导师的相关信息, 同时, 评阅人的信息对外保密)。采用专家双盲评阅为保证博士学位论文质量、避免评阅过程中的“人情”、“面子”影响起到了积极的作用^[2,3]。但是, 双盲评阅增加了论文评阅处理时间和评阅成本。对于涉密论文, 还容易造成了一定的保密隐患。由于专家学术评价的主观性、专业水平局限、宽严要求和责任心等不同^[4,5], 还存在论文质量误判的风险。

【收稿日期】 2013-08-15

【基金项目】 国防科学技术大学“十二五”研究生教育教学改革研究课题 (yjsy2013016)

【作者简介】 武小悦 (1963-), 男, 山西平遥人, 国防科学技术大学信息系统与管理学院教授, 博士, 主要研究方向为系统评价与决策分析。

双盲评阅机制存在有两类论文质量误判风险。第1类风险(称为 α 风险)是将合格的论文评为不合格论文(包括修改后答辩、未达到要求等)。造成申请人论文推迟答辩和重新申请。第2类风险(称为 β 风险)是将不合格论文评为通过(包括达到要求,同意答辩)。为了降低双盲评阅的这两类风险(特别是 β 风险),通常的做法是根据多个盲评专家的评阅结论按一定的判决规则(如有2位专家有异议时本次申请取消)做出是否通过评阅的决定。各个高校对于盲评专家的数量确定及评审规则也有不同的规定和要求^[6]。有的学校规定只要有2个专家提出异议,本次学位申请过程即中止。有的学校按3个专家的评审等级进行组合做出是否中止申请的决定。如何制定科学合理的双盲评阅规则,需要从保证博士学位论文质量水平以及质量保证的经济和时间成本(包括评阅成本、博士生推迟答辩等)等多个方面进行综合权衡。为此,必须分析不同的评阅规则、论文质量水平及通过评阅风险之间的量化关系。

关于博士学位论文评阅问题已开展了一些研究。在评阅机制研究方面,张宇航^[7]较系统地比较了国内外学位论文的评阅机制,说明了采用双盲评阅的必要性和有效性,并基于国防科技大学的多年的论文评阅数据,就评阅机制的改革提出了建议。周珞晶等^[1]结合国防科技大学2001-2003年博士学位论文评阅的统计数据,发现普评与盲评成绩存在2.7-3.6分的明显差异。刘晓元等^[2]基于国防科技大学2001-2008年博士学位论文的17102个评阅数据进行了研究,发现盲评中出现异议的比例为5.4%明显高于明评。也有的文献基于专家评阅结果对论文质量存在的问题进行了分析。魏赤^[8]对2006-2010年北京大学中文系博士学位论文专家评审的数据进行了分析。对论文选题、文献综述、创新成果、基础理论和专业知识、科研能力、学术道德、写作规范等7个方面的专家评分进行了汇总统计,发现创新成果分值明显低于其它分项。施亚玲等^[3]对华南理工大学2005-2010年3666份评阅结果进行了统计分析。发现创新性成果评分较其它项低,并分析了总评成绩与分项成绩的相关度。另外有的文献还研究了当评阅结果不一致时进行处理的方法。孙忻等^[5]对清华大学2001-2004年的980份博士学位论文的隐名评审意见进行了统计,分析了非共识意见的产生原因,提出了建立复审或申辨机制等处理建议。

综上所述,目前关于博士学位论文评阅的相关研究尚处于定性或简单的描述性统计分析阶段,缺少对于评阅机制与不合格论文通过评审的风险之间

的定量关系研究。

评审专家人数以及评审通过规则的制定主要是根据各高校的情况和管理经验确定,缺少充分的理论指导和定量决策依据。本文根据统计检验的原理,基于一定的假设,建立了博士学位论文评阅风险的量化分析模型,为改进和完善博士学位论文评阅的机制提供决策参考依据。

二、评阅风险分析模型

(一) 研究假设

为了简化研究,本论文提出如下的研究假设。

1) 博士学位论文的真实质量水平 θ 分为合格与不合格两个等级。 $\theta=1,0$ 分别表示合格与不合格。

2) 各个专家的评阅结果分为合格与不合格两种情况。第 i 个专家对于博士学位论文的评阅结果为 x_i 。 $x_i=1,0$ 分别表示评阅结果为合格与不合格。

3) 盲评专家的数量为 N 。每个专家的评分是独立、客观公正进行的,各个盲评专家对于学位论文的判断能力水平相同。专家评阅的评阅判断水平用误判概率表示。假设

专家将真实质量水平 $\theta=i$ 的论文评判为 j 的概率是 p_{ij} 。

4) 假设盲评规则为:若 N 个盲评专家中有大于或等于 d 个认为论文不合格,则论文不能通过本次评审。

(二) 评阅风险的计算方法

根据前面的假定,当论文真实质量水平为合格时(即 $\theta=1$ 时),盲评专家将合格论文评阅为不合格论文的概率为 p_{p10} ,将论文评阅为合格论文的概率为 p_{11} 。由概率理论,当 N 个盲评专家中有 s 个特定的专家将论文评为合格,其余 $N-s$ 个特定的专家将论文误评为不合格的概率为 $p_{11}^s p_{10}^{N-s}$ 。考虑所有 s 个专家和 $N-s$ 个专家的不同组合,评阅中有 s 个专家将论文评为合格,而 $N-s$ 个专家将论文误评为不合格的概率为 $C_N^{N-s} p_{11}^s p_{10}^{N-s}$ 。其中, C_N^{N-s} 表示从 N 个对象中抽取 $N-s$ 个对象的不同组合数。

根据2.1中关于评阅机制的假设,若盲评专家中评阅结时为合格的专家数量 s 小于等于 $N-d$ (即,大于或等于 d 个认为论文不合格),则论文不能通过本次评审。出现这种情况的条件为 $s \leq N-d$ 。

这样,考虑所有使一篇合格论文被评为不合格的情况,可以得到使该论文按盲评规则被误判为不合格的情况出现的概率 P_α 为

$$P_{\alpha} = \sum_{s \leq N-d} N p_{11}^s p_{10}^{N-s}$$

类似地, 可以推导出当论文真实质量水平为不合格时 (即 $\theta = 0$ 时), 论文按盲评规则被误判为合格的情况出现的概率 P_{β} 为

$$P_{\beta} = \sum_{m \leq d-1} N p_{00}^m p_{01}^{N-m}$$

上述公式中, p_{01} 表示盲评专家将不合格论文评为合格论文的概率。 p_{00} 表示盲评专家将不合格论文评为不合格论文的概率。

由上面两个计算 P_{α} 与 P_{β} 的公式即可得到博士学位论文评阅中两种类型的评阅风险、盲评专家数量与水平之间的关系。

三、风险的计算分析

通过编制计算机程序, 可以由前面的公式计算得到在阅专家数量 N 与评阅规则临界值 d 给定时的两类风险。以下通过算例进行说明上述模型。

图 1、2 给出了 $N = 1 \sim 10$, $d = 1 \sim 10$ 时的两类风险水平变化情况。

图中, α 风险用实线表示, β 风险用虚线表示。图 1 表示的情况是: 假设双盲专家评阅时, 犯第 1 类错误 (将合格论文评为不合格论文) 的概率小于犯第 2 类错误的概率 (将不合格论文评为合格论文), 即考虑到通常评阅人将一篇送审的博士学位论文评为不合格是十分慎重的^[7]。误判概率分别为 $p_{01} = 0.2$, $p_{10} = 0.1$ 。图 2 表示的情况是: 假设双盲专家评阅时, 犯两类类错误的概率相等, $p_{01} = 0.1$, $p_{10} = 0.1$ 。

表 1 给出了在盲评专家的总数和专家评阅水平给定的条件下, 两类评阅风险最为接近时的值及对应的两类风险值。

通过计算分析, 结合图 1、2 和表 1, 可以得到如下规律。

1) 当盲评专家总数 N 一定时, 随着评阅规则 d 的增加, α 风险变小, 而 β 风险变大。因此, 仅改变 d 不能同时降低两类风险的水平。这是由于当 d 变大时, 要将一篇论文根据专家评阅意见判为不合格论文的难度变大, 而不合格论文通过评阅的可能性变大。

2) 当盲评专家数量 N 和决策规则参数 d 一定时, 盲评专家对于论文质量的判断水平越高 (如, p_{01} 由 0.2 降低为 0.1), 对应的风险水平越小 (参见图 2)。

3) 通过增加盲评专家数量和专家评阅判断水平, 可以同时降低评阅机制的两类风险。特别是专家判断水平的提高将可使风险明显下降。如表

1 所示, 在 $p_{01} = 0.2$, $p_{10} = 0.1$, $N = 7$, $d = 3$ 时, 对应的两类风险水平分别为 $P_{\alpha} = 0.0257$, $P_{\beta} = 0.0047$ 。在 $p_{01} = 0.1$, $p_{10} = 0.1$, $N = 7$, $d = 4$ 时, 对应的两类风险水平分别为 $P_{\alpha} = 0.00027$, $P_{\beta} = 0.0027$ 。这时, 1000 篇论文中被误判的数量约为 3 篇。

4) 当盲评专家评阅质量水平为 $p_{01} = 0.1$, $p_{10} = 0.1$ 时, 为使两类风险均降低到 0.001 以下的水平 (1000 篇论文中被误判的数量不大于 1 篇), 盲评专家不应少于 9 个, 评阅规则对应的 $d = 5$ 。为使两类风险均降低到 0.01 以下的水平, 盲评专家不应少于 5 个, 对应的 $d = 5$ 。

5) 当以控制 β 风险 (不合格论文被评阅机制误判为通过) 为主时, 若 $p_{01} = 0.1$, $p_{10} = 0.1$, 为使 β 风险降低到 0.001 以下, 盲评专家不应少于 3 个。这时的评阅规则为 $N = 3$, $d = 1$ 。若 3 个评阅专家中有 1 个专家认为不合格, 则该论文被判为不通过。对应的两类风险水平分别为 $P_{\alpha} = 0.2710$, $P_{\beta} = 0.001$ 。这时 α 风险变大, 合格论文被误判的概率较大。

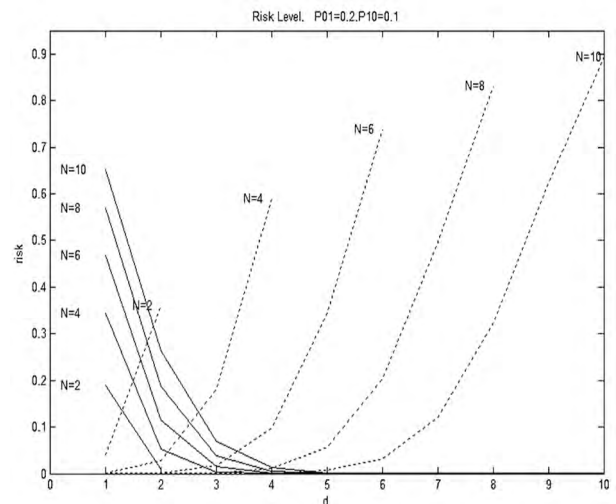


图 1 两类风险水平的变化 ($p_{01} = 0.2$, $p_{10} = 0.1$)

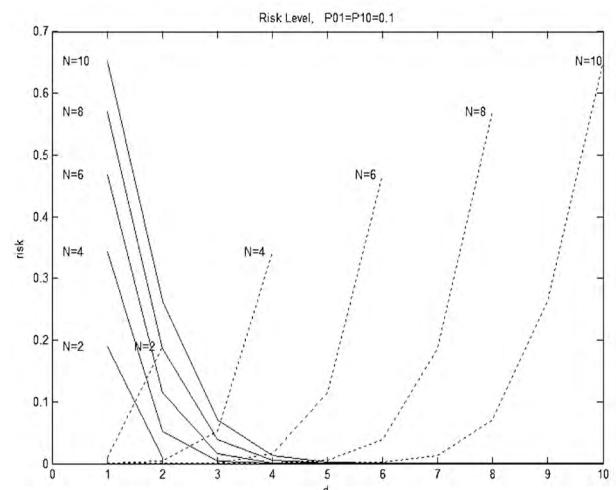


图 2 两类风险水平的变化 ($p_{01} = 0.1$, $p_{10} = 0.1$)

表1 两类风险水平接近时不同规则对应的风险值

N	$p_{01} = 0.2, p_{10} = 0.1$			$p_{01} = 0.1, p_{10} = 0.1$		
	d	α	β	d	α	β
1	1	0.1000	0.2000	1	0.1000	0.1000
2	1	0.1900	0.0400	1	0.1900	0.0100
3	2	0.0280	0.1040	2	0.0280	0.0028
4	2	0.0523	0.0272	2	0.0523	0.0037
5	3	0.0086	0.0579	3	0.0086	0.0086
6	3	0.0159	0.0170	3	0.0159	0.0013
7	3	0.0257	0.0047	4	0.0027	0.0027
8	4	0.0050	0.0104	4	0.0050	0.0004
9	4	0.0083	0.0031	5	0.0009	0.0009
10	5	0.0016	0.0064	5	0.0016	0.0001

四、结束语

为了维护博士学位论文水平的声誉和保障博士学位申请人的正当权益,迫切需要建立科学、公正和规范的博士学位论文质量评价体系和监督机制,以便合理地控制不合格论文通过评审的风险水平,同时又避免不必要的评阅成本与评阅时间投入。

基于数理统计理论,建立博士学位论文评阅中

(上接第41页)

与实际脱节严重,论文写出来后便束之高阁,派不上用场,这种状况必须改变。

其次,在教学方法上,要适应研究生教育转变的需要。一是要大力开展研究讨论,减少知识传授性的教学,增加对部队现实问题的研究,促进研究生研究能力的提高;二是要增大实践教学比例,将所学理论运用到实践中,提高学员适应岗位任职的能力;三是要加强与部队的联系,可以采用到部队调研或聘请部队领导担任兼职导师的方法,使学员熟悉部队情况、摸准部队“脉搏”,加强学习和研究的针对性。

(四)军队研究生教学要注重方法论的指导

研究生与教员、科研人员不同,重要的不是形成科研产品,而是掌握科研方法、培养创新精神和创新能力。因此,研究生培养必须注重方法论的教育。

首先,要注重研究生课程的教学方法,大力提倡研讨式教学。通过教学,不仅让学员掌握必需的知识,更重要的是培养学员的问题意识和质疑精神,掌握分析和研究问题的思路和方法,获得方法论的启示。研究生教学,一定不要过多采用大课讲授的方法,用集体教学代替个性发展。

的两类风险的分析模型,通过模型计算,可以给出博士论文评审的误判风险与盲评专家数量、盲评专家评阅水平、评审通过规则之间的定量关系。这些关系可为高校博士学位管理部门制定学位论文评审规则提供定量的决策参考依据。

[参考文献]

- [1] 周珞晶,张春元,方毅,张宇航. 博士学位论文评阅制度改革成效分析[J]. 高等教育研究学报, 2005, 28(1): 47-51.
- [2] 刘晓元,周珞晶,方毅. 博士学位论文评阅制度改革的设想[J]. 学位与研究生教育, 2010, (10): 24-27.
- [3] 施亚玲,李若英,杜娟,何源. 关于博士学位论文评价体系若干问题的思考[J]. 华南理工大学学报(社会科学版), 2011, 13(5): 105-110.
- [4] 刘朝晖,万宇. 提高博士学位论文质量之我见[J]. 西南交通大学学报(社会科学版), 2003, 4(2): 115-118.
- [5] 孙妍. 谈博士学位论文匿名评审中非共识评阅意见的处理[J]. 中国高教研究, 2004, (10): 40-42.
- [6] 周珞晶. 关于我校博士学位论文评阅和答辩规定的思考[J]. 高等教育研究学报, 2007, (3): 93-94.
- [7] 张宇航. 博士学位论文评阅制度改革研究与实践——以国防科技大学为例[D]. 长沙: 国防科技大学研究生院, 2006.
- [8] 魏赤. 关于博士学位论文创新问题的探讨[J]. 海南师范大学学报(社会科学版), 2011, 24(1): 152-154.
- [9] 武小悦,刘琦. 装备试验与评价[M]. 北京: 国防工业出版社, 2008: 286-287.

(责任编辑: 胡志刚)

其次,要注重研究生开题过程的指导。研究生开题不是简单的指定论文题目,而是要让学员明确研究目标,理清研究思路,确定研究要点,掌握研究方法。换句话说,开题过程实质上也是一个方法论的教育过程。因此,开题过程一定不能简单化,要扎扎实实进行。实践证明,开题过程进行得扎实有效,论文撰写过程会减少很多盲目性。

再次,要注重研究生课题研究期间的指导。学位论文撰写既是研究生科研方法训练的重要过程,也是研究生毕业考核的重要方式。作为导师,要帮助研究生制定严密的研究计划,实行定期、不间断的指导,及时发现和解决问题,确保研究过程的有序进行。作为研究生培养单位,也要定期检查、督促,发现问题及时纠正。

[参考文献]

- [1] 董会瑜. 军校教育“特点”问题研究[M]. 北京: 解放军出版社, 2004: 186.
- [2] 黄宝印. 我国专业学位教育发展的回顾与思考[J]. 学位与研究生教育, 2007(6): 16.
- [3] 任海泉. 外军院校教育研究[M]. 北京: 国防大学出版社, 2008: 135.
- [4] 赵军. 研究生个性化教育扫描: 问题、原因及对策[J]. 学位与研究生教育, 2012(6): 19.

(责任编辑: 陈勇)