

研究生学位论文盲审分歧及评审改进研究

杜学领

(贵州理工学院 矿业工程学院, 贵州 贵阳 550003)

摘要:学位论文盲审是提升研究生培养质量的重要手段,公平盲审是发挥盲审裁定权的重要前提。采用统计调查法,以B高校两专业45位博士生合计135份学位论文盲审数据为样本,对样本中分项内容的盲审分歧、盲审结果的最大分差、熟悉程度对盲审结果的影响、盲审结果与论文后期表现的关系等方面展开分析,并对盲审制及改进策略展开讨论。研究表明:盲审有利于自由裁量权的发挥,在盲审中普遍存在轻微分歧,个体遭遇的中等分歧比重偏高,可接受的盲审结果却未必是公正的结果。不公正的评价更容易发生在平均成绩较差的学位论文中,权威的可谬性更容易发生在高估的场景中。对评审内容自我感觉越熟悉,越敢于给出负面或正面评价,而熟悉程度仅为“一般”的评阅人,其评审结果未必精准。从学位论文的后期表现而言,对于被延期学生而言,否定性的盲审结果是相对客观的,但对于非延期学生的评价,盲审分歧之中有可能存在不当评价。未来,有必要在建立统一的网络化评审平台的基础上,改善盲审评审前、评审中、评审后的各个环节,让盲审更大程度地回归学位论文本体的评价之中。

关键词:学位论文;盲审;研究生;高等教育;评审公平

中图分类号:G643;G647;G521

文献标志码:A

文章编号:1674-5884(2021)02-0110-11

学术评价广泛存在于学术活动之中,在科研项目评选、期刊论文同行评议、学位论文评审之中均可见学术评价的身影。但正如有学者指出的,因学术评价本身具有自由裁量权,造成公众监督困难、司法介入有限,学术评价符合程序但未必程序正当、程序正当未必结果公正、基于公平正义但差别性原则未必得到贯彻,以致于学术评价结果的权威性和公信力受到质疑^[1]。近70年来,我国研究生教育招生规模不断扩大、培养学制多次变革、教育政策多次变迁^[2-4],研究生体量规模化增长的同时也存在一系列问题,如培养目标不清、导师制流于形式、实践能力弱、偏重于论文发表评价、研究生教育的结构性失衡等^[5,6]。研究生延期毕业及相关问题引发社会广泛关注,甚至一项分层抽样的实证研究发现博士研究生平均延期率高达39.68%^[7]。尽管有学者对研究生延期的影响因素和应对策略展开过分析,并提出提高生源质量、强化过程管理、健全教育投入等改进措

施^[8],但在教育实践中,学位论文外审评价对学生延期与否具有关键性作用。

研究生学位论文外审一般分为明审、单盲审、双盲审三种类型,其中于永凤等根据论文评价载体、评审专家选择、送审渠道、送审机构等将双盲审细分为多种类型,并认为“第三方平台式”送审要优于“直接送达式”和“单位互助式”送审^[9]。已有的实证性研究也表明,明审存在虚高现象,盲审相对更客观、要优于明审^[10,11]。在近年来重视学术诚信、提高人才培养质量、教育部学位论文抽查等背景下,学位论文盲审成为众多培养机构的重要选择。以往的研究偏重于论证盲审的正当性、必要性及与明审的差异,认为盲审客观的结论是建立在与明审的对比之后,缺少对盲审结果本体公正性的探讨。但基于权威的可谬性,评审专家同样会犯错^[12],以致于产生不公正的评审结果,这种不公平可以通过盲审结果进行评价。基于盲审结果反思盲审过程,对于改进盲审评审机

收稿日期:2020-11-08

作者简介:杜学领(1986-)男,满族,河北承德人,副教授,博士,主要从事高等教育、工程教育、通识教育研究。

制针对性更强,可靠的评审结论是执行延期的必要依据。本文以B高校两个专业45位博士生合计135份学位论文盲审意见为分析样本,由评价结果反思盲审的结果公正性,并进一步反思研究生盲审制的改进策略,研究内容有助于改进现有的学位论文盲审制度。

1 研究背景概述

1.1 研究对象

B高校在2016年春季对当年送审某二级学院的博士学位论文全部实施“第三方平台式”双盲审送审,本文以本次送审中该学院S和M专业盲审反馈数据为样本(含一级学科下的不同方向及延期毕业同期送审的数据),其中每篇学位论文均由3名外审专家评阅并给出评审意见,三位评审中有一位认为总体质量较差,一般即做延期处理。当年参与盲审送审的S和M两个专业的学生数分别为19,26人,合计135份学位论文盲审意见。

1.2 评审指标及数据处理说明

本次评审均采用相同的评价量表,分项评价内容主要包括“选题”“创新性及论文价值”“基础知识和科研能力”“论文规范性”四个方面,每一项的评价等级分为“优秀”“良好”“一般”“较差”四级,并要求评阅人给出总体评价;在百分制的条件下,“优秀”“良好”“一般”“较差”对应的分数段分别为“100~90”“89~75”“74~60”“60以下”。评审人还要就“是否同意答辩”给出意见,其中“同意答辩”“修改后直接答辩”“不同意答辩”的分数指标分别为“80分以上”“79~60分以上”“59分以下”;评审人还应给出“是否推荐该论文参加优秀论文评选”,并注明自己对论文的熟悉程度。此外,评审人还应给出学术评语、指出论文的不足,并评价论文的创新点,创新点的评价同样分为“优秀”“良好”“一般”“较差”四级。

在上述评价内容中,并没有给出每一项评价

内容的比重,也没有要求评审人对每一个分级有明确的评分,就有可能造成评价内容与总体评价不对应。以往的研究中对于盲审结果一般从总体分布特点上进行分析,潜在地默认了评审结果是被接受的。但这种评价方法并没有考虑不同评审之间的分歧及盲审结果是否可信。在盲审的情况下,不同评阅人实施自由裁量权的结果越接近,说明评阅人之间的分歧越小,一定程度上可以反映评阅群体对同一学位论文的评价相近,这种默认达成的结果可以视为相对公平。反之,若评阅人之间存在分歧,则分歧越大,对学生的影响越显著。分歧大,会直接影响到学生的平均成绩,较高的分数会影响到学位论文的推优,较低的分数则有可能导致学生被延期,二者均会影响到评审的公平性。为了便于量化分析,本文采用相对距离法,对于非定量评价的不同分级,认为相邻等级之间的距离为1,对同一学位论文的三份盲审结果进行两两比较,将比较结果的加和值定义为分歧指数。分歧指数越高,表明评阅人之间分歧越大,对于同一被评阅人而言评阅结果的公平性越差。分歧指数为0时,表明不同评阅人之间的评价结果相同,公平性相对最优。为了便于统计分析,本文分专业按照总评成绩平均分由低到高对每一位学生进行编号,并利用Microsoft Excel对盲审结果进行统计和处理。

2 盲审结果分歧及讨论

2.1 分项内容的盲审分歧

一般而言,评审人的专业素养和辨识力要高于学生,正常情况下评审人之间虽然处于盲审状态,但也应达成一定共识。因此,根据前述分歧指数的界定和分歧指数的大小将盲审分歧程度分为无分歧、轻微分歧、中等分歧、严重分歧四级,相对应的分歧指数分别为0,2,4,≥6。表1为统计结果,图1为不同个体的总体分歧指数。

表1 盲审分歧程度统计

单位:人数

指标等级	选题		创新性及论文价值		基础知识和科研能力		论文规范性	
	S专业	M专业	S专业	M专业	S专业	M专业	S专业	M专业
无分歧	4	6	0	3	3	6	5	3
轻微分歧	11	16	16	16	12	15	9	15
中等分歧	4	3	3	6	4	5	4	6
严重分歧	0	1	0	1	0	0	1	2

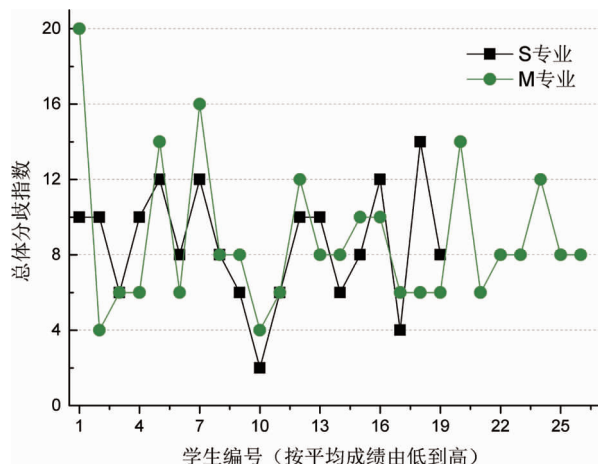


图1 不同个体的总体分歧指数

结合统计情况及表1、图1可知:(1)从分歧程度而言,两专业不同分项评价内容比重最高的均为轻微分歧,这说明轻微分歧在当前的盲审条件下是普遍存在的。(2)因两专业的样本容量均较小,可近似认为无分歧和中等分歧的比重相当,但中等分歧的比重要略高于无分歧,二者的比重均低于轻微分歧的情况。(3)严重分歧的情况相对较少,S专业编号为12、M专业编号为7的严重分歧均在“论文规范性”方面,M专业编号为1的学生除“基础知识和科研能力”为轻微分歧外,其余三项均存在严重分歧,综合其评语可知,该学生的学位论文确实存在被高估的问题,个别盲审专家未尽到公平公正评阅的义务。(4)S专业各分项分歧的总和近似相等,而M专业“创新性”及“论文价值”的分歧要高于其他两项内容。对于“创新性”及“论文价值”,可能与评审人本身的专业素养和判断力有关,在135份对“创新性”及“论文价值”的评价中,“优秀”“良好”“一般”“较差”的数量分别为18,63,50,4;较多的评审人认为创新性处于“良好”或“一般”,因二者为临近等级,这种分歧是可接受的分歧。但对比相应的评语可知,关于“论文规范性”的分歧,可能和不同单位对学位论文的格式要求有关,甚至一些关于“论文规范性”的建议本身就是错的,以致于产生误判。这种误判,应通过改善评审环节予以避免。(5)从个体的总体分歧指数而言,单项分歧经过分项加和后,施加在个体身上的分歧效应进一步被放大,S和M专业分别有47.36%和30.76%的学生遭遇中等分歧,两者均高于表1中分项中等分歧的比重。中等分歧比重过高,显然是不利

于评审公平的,应采取措施降低中等及严重分歧。

之所以存在这种分歧,笔者认为,主要的原因在于以下几点:(1)本轮盲审分项内容仅采用了分级评价的方法,且分级只有四级,有限的分级使得分级的区分度也较为有限;(2)每份学位论文只有3位评阅人,评阅人数量较少,可能造成分歧被放大;(3)分级评价中为定性评价,而总体评价中为定量加定性评价,二者不完全对应,造成存在分歧;(4)分项评价结果在不同评价主体间两两对比后经过累加,造成施加在个体上的总的分歧被放大。

2.2 盲审结果的最大分差

图2为盲审结果的最大分差分布情况,其中最大分差由三份评审意见中的最高总体评价分数与最低总体评价分数做差值所得,人数统计为两专业45名博士生的总体统计情况。由图2可知:(1)盲审条件下,自由裁量权得到淋漓尽致的发挥,45名被评价对象中仅有1人最大分差为0,盲审有利于实现差异性评价,从而更全面地反映评阅人的真实意见。(2)最大分差的分布基本符合正态分布特征,最大分差极大和极小的情况均属于少数,也从侧面说明评价结果总体是可接受的。但这种可接受的结果却未必是公正的结果,本轮盲审最大分差高达50分,对于该学生的评审意见,其中应有并不恰当的评审意见,以致于评审结果并不公正。(3)最大分差在6~20分的人数高达32人,占总人数的71.11%,这一部分最大分差的平均值为12.59。而总体的最大分差平均值为13.31,135份总评成绩的平均值为79.99,考虑到“优秀”“良好”“一般”“较差”的分级差值一般在10~14分,在总体当中最大分差不低于10分的高达27人,占比60%,说明在本次盲审中,盲审分歧较大,且盲审分歧会影响到盲审结果的公正性。(4)尽管最大分差的分布具有一定随机性,但结合学生的平均成绩可知,学生平均成绩越低,往往最大分差越大;学生平均成绩越高,往往最大分差越小。两专业中,平均成绩最高的10名学生的最大分差平均值仅为10.15,而平均成绩最低的10名学生的最大分差平均值则高达17.3。这说明,不公正的评价更容易发生在平均成绩较差的学位论文中。由于极低的分数可能给学生带来惩罚性、灾难性的后果,可以推知,在学位论文评审中,权威的可谬性更容易发生在高估的场景当中,即

一些高分评审结果中,不排除包含错误评审结果。对于样本中最大分差达50分的学生,其三份总评成绩分别为90,60,40,结合学术评语部分,可以很明显地看出其90分的高分评审结果是存疑的,甚至60分的评审结果也有放水的嫌疑。若无盲审中40分的制衡,反而该学生当年有可能顺利毕业。此外,由于总评成绩的平均值为79.99分,而高达60%的学生最大分差都不低于10,这意味着这次盲评不仅分歧较大,而且盲评的精准度也有待于进一步提高。

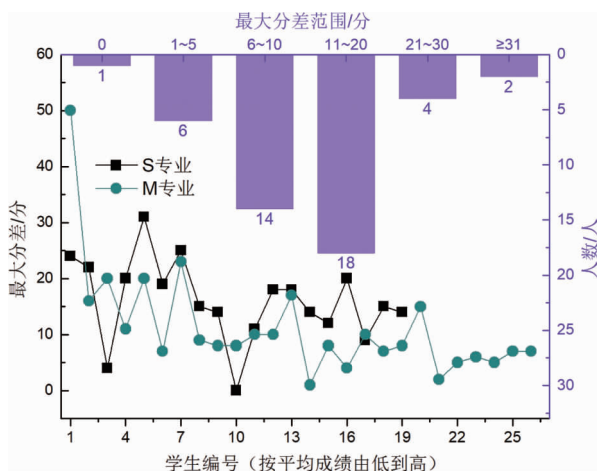


图2 盲审结果的最大分差

2.3 熟悉程度对盲审结果的影响

因两专业单独分析时样本容量较小,在熟悉程度的分析中,将两专业的数据进行总体分析,统计结果如表2所示。

结合统计过程及表2可知:(1)盲审评阅人自评中,“很熟悉”“熟悉”“一般”的比例分别达到22.96%(31人次)、65.19%(88人次)、11.85%

(16人次),从总体而言,评阅人自我感觉良好。但从前述分析可知,高达60%的学生最大分差都不低于10,评阅人自我感觉良好与真实评价结果之间不完全对应,以致于其中必然存在不公平的评审。(2)在所有熟悉程度中,不论是总评还是分项评价,“良好”所占比例均相对较高,“优秀”和“一般”次之,这一结果与总评结果基本吻合,总评中“优秀”“良好”“一般”“较差”的占比分别为14.81%,58.52%,22.96%,3.7%。(3)选拔优秀与剔除劣质论文与熟悉程度有较强的相关性,对评审内容自我感觉越熟悉,越敢于给出负面或正面评价,而对于熟悉程度仅为“一般”的,无论是分项评价还是总体评价,给出“较差”的情况为0,给出“优秀”的比例也相对较低。在打击学术不端、倡导学术诚信、提升学位论文和培养质量的现状下,如果评阅人的熟悉程度仅为“一般”,其评审结论的重要程度不可与“很熟悉”“熟悉”同日而语。特别是在仅有3位评阅人的情况下,因熟悉程度为“一般”的评阅人无法给出极端差评,造成其评审结果会影响到评审的公平。(4)从分项内容而言,熟悉程度为“很熟悉”“熟悉”的评阅人认为选题“一般”的比例要高于熟悉程度为“一般”的评阅人,而熟悉程度为“一般”的评阅人仅能在“论文规范性”方面给出相对较高比例的“优秀”评价,这说明熟悉程度为“一般”的评阅人对于分项内容评价的把握未必精准。在评阅人选择方面,应尽可能多地选择熟悉学位论文内容的同行作为评阅人,并避免同一论文全部送到熟悉程度仅为“一般”的评阅人手中。

表2 不同熟悉程度对评审结果的影响

(a)“很熟悉”31份评审结果详情					单位: %
评价内容	选题	创新性及论文价值	基础知识和科研能力	论文规范性	总评
优秀	19.35	19.35	22.58	32.26	29.03
良好	61.29	48.39	54.84	35.48	41.94
一般	16.13	25.81	19.35	25.81	22.58
较差	3.23	6.45	3.23	6.45	6.45
(b)“熟悉”88份评审结果详情					单位: %
评价内容	选题	创新性及论文价值	基础知识和科研能力	论文规范性	总评
优秀	15.91	12.50	19.32	12.50	10.23
良好	67.05	42.05	48.86	48.86	64.77
一般	15.91	43.18	30.68	35.23	21.59
较差	1.14	2.27	1.14	3.41	3.41

(c)“一般”16份评审结果详情

单位:%

评价内容	选题	创新性及论文价值	基础知识和科研能力	论文规范性	总评
优秀	37.50	6.25	6.25	18.75	12.50
良好	56.25	68.75	68.75	56.25	56.25
一般	6.25	25.00	25.00	25.00	31.25
较差	0.00	0.00	0.00	0.00	0.00

2.4 盲审结果与论文后期表现的对比

本轮盲审中,凡三位评审中有任意一人给出的总评分数低于60分,被评价人会被延期半年或以上。除延期学生外,其余大部分学生均在本次盲审后按期毕业,2016年下半年其学位论文即可在中国知网下载,论文在中国知网发布距今至少3年以上,延期学生的学位论文在中国知网发布距今至少1年以上。由于当年盲审的结果在5月中下旬返回,而学生一般在6月上旬就要参加毕业答辩,实际上在此期间对学位论文内容进行实质性、大范围改动的属于小概率事件,因此可以认为,盲审前后论文的主体内容是变动不大的。论文的下下载量一定程度上可以反映同行对选题的关注,而论文的被引用次数一定程度上可以反映出论文的价值。与有限的盲审专家相比,这种公开的数据在更大范围反映了论文的后期表现。

图3为论文后期表现统计,统计数据源于中国知网的公开数据,统计时间均为2020年5月20日,其中M专业两名学生因论文未上线而数据缺失。根据统计结果可知,S专业平均成绩与下载量、被引用量的相关系数分别为0.384,0.407;M专业平均成绩与下载量、被引用量的相关系数分别为0.331,-0.074。无论是从下载量还是从被引用量而言,二者与盲审的平均成绩不具有明显的相关性,反而离散性较强。如S专业编号为3、M专业编号为2和3的学生,尽管这三位学生盲审的平均成绩在同期当中表现较差,但后续其学位论文下载量、被引用量在同期学生当中均达到较好水平。反之,S专业盲审中平均成绩最高的学生,至今其论文下载量仅为389、被引用次数仅为2,后期表现与盲审结果形成强烈反差。四位被延期的学生,论文平均下载量为308、平均被引用次数为2,而总的平均下载量为689、总的平均被引用次数为7,即便考虑论文上线时间的差异,二者间的差异也是较大的,说明对于被延期学生而言,盲审否定性的评审结果是相对客观的。同时,这种否定对于学位论文的后期表现并没有起

到显著的改善作用,以致于被延期后提升论文质量有可能变得名存实亡;但对于非延期学生的评价,盲审分歧之中有可能存在不当评价。甚至可以认为,盲审中的公平是相对的,而不公平是绝对存在的。但因不公平的评价并没有引发延期这种恶性后果,使得过程合规、但结果不公正并未引起充足重视。但若后续的优秀论文、优秀毕业生评选要依据盲审成绩,由于连锁效应,会造成不公平的影响持续扩大。因此,盲目地寄希望于盲审而不考虑盲审结果的公平性意义,是不可取的。

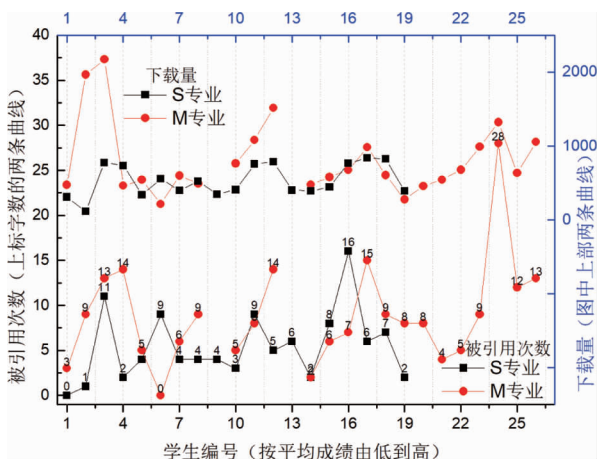


图3 学位论文后期表现统计

2.5 盲审结果分歧的其他方面

对创新点的评价方面也存在较大争议,主要表现为三方面:(1)不同评审人对评审表的理解差异,导致评审分歧。原评价表中对创新点标有数字1~5,本意为对每一创新点进行单独评价,但一些评审人只在第一条中给出评价,有误评、懒评之嫌疑,135份盲审结果中有14份出现这种情况,占比达10.37%。(2)对分条创新点部分的误评,表现为评价数量与被评价论文的创新点数量不符,如对某篇学位论文创新点的评价,三位评阅人分别给出了1,3,5条评价结果,其中必然有误评。而S、M两专业中,在分条创新点评价中分别有11(57.89%),19(73.07%)位学生遇到这种评价结果,说明评价的不公平是客观存在的。造成

这一结果的原因,一方面和评价人本身的理解有误相关,后续盲审时应对评价指标给予必要的说明并提出明确的评审要求,以减少这种误评;另一方面,也与评审人本身的素养和治学的严谨程度关系紧密,说明本轮盲审中一些评阅人极其不负责任,以致于出现驴唇不对马嘴的盲审结果。如S专业编号14的一位盲审结果中,对于博士学位论文的评价竟然出现了“符合硕士学位论文的要求”这样荒谬的评语。在后续盲审中,应对评价结果建立反馈机制,反向评价评阅人的评审能力,并建立退出和惩罚机制。很难想象,对博士学位论文评审都如此草率的专家,日常如何教书育人、为人师表?(3)对创新点的评价与对分项评价中对“创新性 & 论文价值”评价之间的分歧和自我矛盾。由于对“创新性 & 论文价值”的评价只能勾选一个选项,而对创新点的评价是对多条创新点的单独分级评价,会形成相应的多条评价结果,造成二者之间可能存在分歧和自我矛盾。如对“创新性 & 论文价值”的评价结果为“良好”,但对三条创新点的评价却为“良好”“一般”“一般”,这种评价结果是自相矛盾的。由于“良好”和“一般”本身属于近似评价,在后续的盲审中,可以进一步在细化“创新性 & 论文价值”评价的基础上,取消对创新点的单独评价,或将其融入“创新性 & 论文价值”评价中。

盲审中推优结论的问题。在135份评价结果中,有99份结论为“不推荐”,其中总评成绩最高分为95分,有5份的总评成绩不低于90分;有30份结论为“推荐学校优秀”,其中总评成绩有23份低于90分、7份总评成绩不低于90分;结论为“推荐全国优秀”“推荐省级优秀”的均为3份,其中最高总评成绩为95分。很明显,上述评价结果是存在较大分歧的,甚至总评成绩与推优结论在不同评审人之间的尺度并不相同。在此情况下,实际上其推优结论的可参考性已经大大降低。

此外,主观部分的“学术评语”部分与分项评价之间也存在一些矛盾之处,如“学术评语”中认为“论文研究内容较为丰富,研究思路较为清晰,研究方法得当,具有一定的理论意义和实践意义”,但在分项评价中却认为“选题”“创新性 & 论文价值”“基础知识和科研能力”均为“一般”,类似的分歧及评审矛盾也并不鲜见。

综合以上分析可以看出,盲审虽然发挥了自

由裁量权,但限于现有的盲审规模和评价指标体系,自由裁量权反而导致不同评审之间分歧较大、对论文内容不熟悉的评审无法发挥出应有的裁定作用,以致盲审结果的公正性受到影响。对于现有的盲审制,还应进一步改进。

3 关于盲审制的讨论及改进

3.1 关于盲审制的讨论

关于学位授予的评价标准,龚向和等^[13]认为与西方国家主要以学位论文和课程学分作为主要评价依据相比,我国的学位授予中存在重结果轻过程、要求发表非学位论文等问题。在稍早的研究中,甚至有学者提出学生发表权威期刊或高影响因子国际期刊论文可以申请论文免审^[14],这无疑也过度夸大和助长了“唯论文”导向。关于盲审制的改进,汪小会等^[15]提出抽检、完善评分标准、选择送审院校、允许复审、签订“公正评审承诺书”、基于盲审结果进行院校的后续管控、建立动态更新的盲审专家库、优化网上评审等策略。罗泽胜^[16]详细讨论了硕士学位论文盲审意见的异议机制,但同时指出异议机制仍面临利益相关者不重视、缺乏评审资源、评审时间仓促等问题。类似的关于评价标准、评审改进的研究还有很多,其中不乏合理的建议,甚至可以认为策略层面已经较为健全。但通过本文的盲审数据可以看出,实际盲审结果却不尽如人意,提高盲审制的实效性变得尤为重要。限于篇幅,本文主要就与本文相关的几个问题展开讨论。

3.1.1 导师和评阅人,哪个更重要

盲审制之下,一些导师为了规避责任,直接将评审的决定权交给盲审专家,甚至出现先盲审、后导师评审的情况。盲审全部通过,导师也不否定;盲审不通过,导师也不给予肯定。这种倒置的评审顺序看似“小聪明”,但当盲审结果并不公正时,导师却未能捍卫学生应得的评价。单纯依托盲审制的评价,有可能导致导师的义务缺失。

导师和评阅人,哪个更重要?从责权而言,导师在培养周期内履行监督、指导义务,导师负责的是过程管理,而评阅人在终端附近行使裁定建议权。但无论是对培养对象,还是对学位论文内容,导师的熟悉程度显然要高于评阅人,因为有了过程的看在眼里,才有了对结果的全面解读。从后期评阅结果的反向影响而言,导师也是利益相关

人员。例如,若同一导师连续多年培养的学生难以通过盲审,其培养质量就会受到质疑,进而影响到其招生资格。盲审制之下,不能将导师评审后置,而应该将导师评审前置。前置不能仅有定点式把关,如现有的开题答辩、中期检查、论文预评审等关键节点,会有导师的介入,但仅仅依靠有限的单点监管,显然是难以胜任培养任务的。导师还应该完善过程性指导环节,如定期召开组会、组织小组专题讨论、日常指导等。导师不能仅仅充当签字的机器,而应该对学生思路上有启发、方向上有指导、平台上有支持、进度上有追踪、质量上有要求,导师是学生量变引起质变的见证人。很显然,对于这些过程性的管理,导师比评阅人更重要。盲审制之下,导师评审的前置,一方面要将过程管理前置,并落到实处;另一方面,导师要对学生学位论文的送审提出全面过程性评价,包括导师对论文终稿指导了几次、每一次有哪些意见、针对每一条意见学生有哪些修改、经过多轮指导之后导师是否同意送审等。在近年来的一些研究生与导师之间的矛盾中,当学生发生恶性后果后,往往导师成为众矢之的,甚至百口莫辩。之所以存在这种现象,与导师拿不出过程性证据密切相关。当家长指责导师故意刁难学生的时候,不妨把所谓的“卡人”条件公开,是非曲直也会一目了然。而在盲审条件下,显然,盲审评阅人有限的评价无力调和这种矛盾。盲审制之下,必须将导师的过程性评价前置。为了避免导师评价前置后确实存在导师刁难学生的情况,还应该健全监管机制,避免导师评价前置后的权力滥用。

3.1.2 评审和时间,都去哪儿了

在现有对学位论文评审的讨论中,评审资源缺失和学生集中毕业导致评审任务多而时间不足引起部分研究人员注意。但造成这些问题与院校本身的评审选择和现有的管理体制有关,主要表现在四个方面:

(1)盲目认同评审专家。在选择评审专家时,一些院校喜欢同类或水平高于本校的评审人做盲审专家,如在前述汪小会等^[15]的研究中就提出扬州大学送审院校会选择学科专业排名略高于或与扬州大学水平相同的院校,这就造成排名越高的高校,可能评审任务越重,直接的不良反应就是评审任务重而评审质量低劣。实际上,对于学位论文的评审,应选择能够公正评价、有助于学生

改进论文质量的人,而不是攀比学科排名。在近年来持续扩招的条件下,具备审理学位论文内容的同行专家数量不在少数,但因学科专业排名低而被排除在评审圈以外。在学科排名过多的倚重非学位论文内容进行评价时,这种对所谓名校专家的趋同是盲目的,也人为地造成部分评审人任务重。

(2)本硕博毕业季扎堆。我国本硕博的毕业时间一般在每年6月,这就造成毕业季之前的评审任务重。在盲从于学科排名的情况之下,名校专家的评审质量更是堪忧。解决这一问题,一是改进专家库遴选机制。目前的专家遴选一般以推荐制为主,未来可适当向自主申请制转移,并根据工作质量而非专家名气来动态调整专家库人员。二是扩大评审规模。如本科的毕业设计或论文,在全省、全国范围内相同专业领域选择评审人,一般教学经验在3年以上拥有博士学位的人审理此类论文问题都不大;硕博学位论文的评审,可在高级职称人员及优秀青年人才中更精准地选择评阅人。限于行业小领域的研究人员有限,若仅仅按照学科专业排名选择,显然是不合理的。笔者建议,硕士研究生学位论文盲审人数不低于3人,博士研究生学位论文盲审人数不低于5人。评审规模扩大,既有利于学生获得更多的评审意见改进论文,并从总体平均的视角看待评审成绩,还有利于出现异议时可通过多份评审意见形成共性可借鉴结论,为异议处理提供依据。三是可以考虑改革现有的学制管理,如鼓励部分优秀研究生申请提前半年毕业,或将硕士的学习年限更改为2.5年,以错开现有的毕业季评审高峰。

(3)无序的送审。现行的送审,往往是知其人而不知其工作量,有可能造成多所高校的论文都送到同一评阅人手中,时间紧、任务重。解决这一问题的办法,可基于行业特点或院校联盟,建立统一的盲审送审网络平台,在平台中实时显示评阅人已审工作量、待审工作量、时间节点要求等,甚至评审人自身也可以设定自己的评审容量,避免盲目的无序送审。

(4)评阅人的素养。在对135份评审意见的调研中发现,评审人所给出的学术评语多属于套话,甚至其中有较大比例直接摘抄了被评审人学位论文摘要或结论部分的内容。相对而言,评审人所指出的不足针对性更强、对提升学位论文更

有帮助。调研还发现,不同评阅人给出的评语文字量并不相同,有的寥寥数语,有的长达半页,但总体而言,评阅人真正自己动手写下来的文字量相比被评阅人而言工作量是微乎其微的。在这种情况下,还要叫苦评审时间不足,一些评阅人的素养有待提高。而且,责任意识和个人素养对评审质量也具有重要影响。但短期内提升评审人的素养是不现实的,依据评审质量反向评价来遴选评审人,对于改进盲审具有实际意义。建立针对学位论文的高素质评审队伍,也与学位论文评审工作更加匹配。

3.1.3 过程和结果,哪个更重要

很明显,现有的盲审制带有明显的“唯结果”趋向。特别是在盲审与毕业时间问题上,由于盲审意见较多的在5月份返回,而6月份学生就要参加毕业答辩,这期间对于论文进行实质性大幅修改几乎是不可能的。由此造成负面的盲审结果成为阻碍学生毕业的屏障,而正面的评价则更像是走了一遍形式。显然,现有盲审制的“唯结果”趋向,强化了盲审意见的裁定权,却不符合人才的渐进性成长规律。从人才培养而言,培养过程要比毕业结果更重要,培养过程不佳,即便盲审结果说得过去,也有可能被反噬。近年来一些高校发生学生毕业后毕业论文被爆出学术诚信、论文质量问题,就是结果反噬的表现。改进这一问题,可基于上述的同一送审平台,动态返回盲审意见,而不是等待所有评审都审理结束后才集中返回意见。在教育实践中,学生学位论文送审过程中,一些学生处于无所事事的状态,若评审意见动态返回,则有利于学生利用好毕业前的最后时间,不断完善自己的学位论文。此外,动态返回评审意见后,不应盲从于评审意见,更应注重学生在接到评审意见后的改进、允许改进后的二次送审和申诉。由此,盲审结果服务于改进过程,改进过程有利于提升学位论文质量,才有利于提高人才培养水平。

3.1.4 是否优秀,谁说了算

现有的优秀论文评选,往往与盲审结果和校内评审密切相关,除前述的盲审结果可能存在评审不公以外,校内评审还与导师的校内地位、导师的校内话语权、投票权群体的关系网等相关,以致优秀论文评选与学位论文后续表现大相径庭。从目前的培养学制而言,我国研究生培养一般在3~4年,其中还包括部分课程学习环节,在仅仅3~4

年内,除极少数研究生能取得变革性成果外,实际上大部分研究生所谓的创新距离产业化、工业化还有一定距离。以致一些学者提出,博士毕业是研究工作的起点。135份对“创新性 & 论文价值”的评价中,有50份结论为“一般”、4份结论为“较差”,也辅证了目前对研究生学位论文创新性的要求有可能是过高了。与盲审有限的评审资源相比,学位论文公开后,将接受更大范围的审视,其选题、研究价值、创新性等方面也将接受更多拷问。操之过急的评优,不仅有可能与学位论文的后期表现不符,而且所谓的评优也并没有真正体现出学位论文的社会意义,让毕业变成了终点而不是起点。在科学界顶尖的诺贝尔奖评选中,其中不乏高龄获奖人^[17]。与成果产出的时间相比,似乎诺贝尔奖姗姗来迟,但诺贝尔奖依然足够权威。考虑学位论文和学生毕业后的后续表现,在毕业3~5年后甚至更久的时间来评选优秀论文,也会让学生把眼光放得更长远,更加追求论文的长效性,以毕业为起点开展更丰富的研究,而不是仅仅着眼于毕业。滞后性评优,也会让盲审结果更加有利于学位论文质量提升,让评优变成惊喜,让获奖的荣誉更加实至名归。

3.1.5 其他方面

(1)分项评价指标方面,本轮评价中并没有细化评价指标,以致于评审之中存在一些分歧和矛盾之处,建议后续进一步精细化评价指标,并实现对每一指标的定量评分。但应注意指标的建立应适宜,评价指标过多时评审人工作量较大,过少时不能反映评价全貌,笔者建议在百分制评价的基础上,可以进一步细分为10项左右的细分项。

(2)回归对学位论文本身内容的盲审。在盲审中,一些单位还附加了作者已发表的论文,实际上这有可能导致双盲审变为单盲审、影响评审公平。另一方面,调研中发现部分评审人的学术评语中参考了作者的已发表论文,甚至根据论文数量及期刊给出评审结论。须知,在早些年“唯论文”的不良导向之下,一些学生发表的期刊论文有可能与其学位论文关系不紧密,就导致这种评价背离了对学位论文本身内容的评价。剥离出已发表论文的影响,有可能评审人会更关注学位论文本身的内容。

(3)建立常态化、定时定量性、批注型网络评审平台。①盲审常态化,是要区分于现有评审中

将盲审结论病态化、恐慌化应用的现状,合理使用盲审意见,并将盲审送审和返回意见常态化管理,如以学院为单位进行集中送审,但返回审理意见可以在送审后1~2周分批次返给学生,以利于学生有更多的时间来完善学位论文。②定时性评价是指要求审稿人在规定时间内完成审理工作;定量性评价是指限制审稿人同期审理的总数。笔者建议,一篇论文自接收之日起限定在7~14日以内评审完毕,同一审稿人同时最多可接收3篇评审邀请;完成审稿后,在截止日期前7天可继续接收邀请,并限制评审的累积总量,避免过量评审。③批注型评审,是要搭建相应的平台,直接实现评审人在评审过程中随时增加评审意见,有利于学生在后期有针对性地修改。基于网络平台,目前这些功能都是很容易实现的,而且在精细化定量评价指标的条件下,评阅人给出分项分数后,可直接利用系统自动计算总分。既可以减少分项量化后计算工作量的增加,也有利于改善现有评价中分项评价为定性评价、总体评价为定性加定量评价不一致的问题。

(4)盲审结果的使用改进。目前,一些院校对盲审中达不到及格水平的负面评价使用有走向延期的趋势,异议申诉往往也针对这一部分学生。笔者认为,盲审中敢于给出不及格结果的,往往是评审人认为其学位论文质量差到令人难以接受的程度,在本轮盲审中,135份总评结果中有2份的分数为60分、1份的分数为61分,不排除其达到合格分数中包含有怜悯或人情成分。但从实际质量而言,59分与61分,有可能是以五十步笑百步。因此,对于盲审结果使用,笔者认为可在以下四方面加以改善:①对出现不及格结果的学位论文进行严格审核,此类学生的异议应是建立在严格审核之上,以保证敢于给出负面评价的评阅人和学生都被公平公正对待。②建立基于校本人才的动态内部管控线,即在盲审合格线的基础上,进一步提出院校管控线,如平均分在60~75分的,修改后应经过二次送审,并将二次送审结果与一次送审相对比。院校管控线应是在当年总体水平的基础上提出的,而不是固定不变的。③及格学生的异议。目前的异议机制主要面向不及格学生,而缺少对及格学生的关照。但从前述分析可以看出,不及格学生往往自身难辞其咎,而及格学生因不用延期而恍若皆大欢喜,恰恰是这种感觉,

使得及格学生中盲审不公未得到足够重视。因此,从提升公平性而言,也应允许及格学生提出异议,并进一步通过二次送审、学术委员会讨论、学术评议等途径,解决争议。④成绩计量的权重平均法。即学生最终成绩与评审人的熟悉程度相关,原则上应避免盲审全部送到熟悉程度为“一般”的评阅人手中。在此基础上,学生最终成绩的计算,相同熟悉程度时首先计算其平均值,然后按照熟悉程度分别界定为“很熟悉”“熟悉”“一般”,其相应的权重分别为50%,30%,20%,由此获得学生最终的成绩。

此外,还有几个与此相关的问题值得探讨:①盲审评阅人的责权问题。目前盲审实践中一些盲审工作属于义务性盲审,也导致盲审结果出现争议时,无法对评阅人实行反向有效的制裁,使得一些盲审人表现出素养很低却管控困难的问题。要破解这一问题,笔者认为在完善网络送审平台的基础上,可改善现有盲审的奖惩机制。例如,在一定周期内对硕导、博导分别提出基础性的评审数量要求,并按照本、硕、博的层次来提出相应数量,此部分为义务性审理工作,相应人员必须完成。超出基础评审量,实行有偿评审。评审费的发放实行滞后性发放,根据其评审质量核定评审费用,由评审费用直接对其形成鞭策作用。对于无法胜任评审工作的,暂停其未来1~2年评审工作,甚至影响其未来招生。在此奖惩模式下,评审的意见会更客观、评审人的责任感也会更强。②探索建立公开评议制度。盲审优于明审,一部分原因是盲审条件下评阅人更敢于给出真实意见,不是专业素养问题,而变成勇气问题。在此条件下,可探索公开评议制度,发挥敢于质疑、敢于发表态度的应有作用。如评审人和被评审人均自愿进行公开化审理,可首先建立评审人的公开库,送审后评审意见网络化公开,被评审人回复评审人亦公开,在全面公开的条件下,公众的监督会促使评审结果更加公正。③两证分离模式评审。在近年来的讨论中,“唯论文”被屡屡提及,但破除“唯论文”导向,绝不是不发表论文。在此基础上,有没有可能探索毕业证和学位证的分离评审模式呢?例如,利用“课程”或“课程+非学位论文”作为毕业证的基本评审条件,利用“课程+学位论文”作为学位证的基本评审条件。相对于现在延期即无法毕业相比,两证分离评审后,毕业证的评审权回归

到培养单位手中,学位证的评审权处于培养单位和外审的制衡之间,使得即便是延期毕业的学生,也有可能先拿到毕业证,同时使得外审更关注学位论文本身,而非附录中作者简历中的已发表论文,让学位论文评审回归到评审学位论文本身。

3.2 关于盲审制的改进

基于前述分析,建立如图4所示的盲审制改进模型。根据评审周期,对现有的盲审制提出以下改进:(1)盲审前的改进。具体措施包括完善导师过程式管理、扩大评阅人队伍并建立相应的

专家库、精准匹配评阅人、为评阅人提供必要的评审要求说明。(2)盲审中改进。具体措施包括精细量化指标评审,扩大评阅人规模,实行常态化、定时定量性、批注型网络平台评审。(3)盲审后的改进。具体措施包括动态定期返回审理意见、评审质量的反向评价及专家遴选、评审结果的可异议机制、二次送审、学术评议、滞后性评优等。

相关策略具体内容已在前文分析,在此不再赘述。

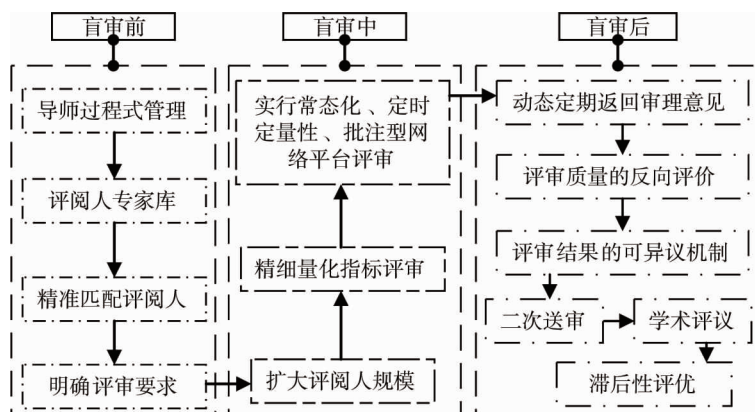


图4 盲审制改进模型

4 结语

学位论文盲审作为学术评价的一部分,对研究生毕业有重大影响。希冀于盲审给出更客观的评价,但同时不能忽略盲审意见本体的公正与否。在盲审中,普遍存在着轻微分歧,应采取措施降低中等及严重分歧。盲审条件下,自由裁量权得到淋漓尽致发挥的同时,造成可接受的盲审结果却未必是公正的结果。学生平均成绩越低,往往最大分差越大,不公正的评价更容易发生在平均成绩较差的学位论文中,权威的可谬性更容易发生在高估的场景当中。对评审内容自我感觉越熟悉,越敢于给出负面或正面评价,而对于熟悉程度仅为“一般”的评阅人,其对于分项内容评价的把握未必精准,应避免同一论文全部送到熟悉程度仅为“一般”的评阅人。从学位论文的后期表现而言,对于被延期学生而言,否定性的盲审结果是相对客观的,但这种否定对于学位论文的后期表现并没有起到显著改善作用,以致于被延期后提升论文质量有可能变得名存实亡;但对于非延期学生的评价,盲审分歧之中有可能存在不当评价。

对于现行盲审制,其往往表现为程序正当,但结果未必绝对公正。未来,有必要在建立统一的网络化评审平台的基础上,改善盲审评审前、评审中、评审后的各个环节,让盲审更大程度地回归到对学位论文本身的评价之中,并把提升学位论文质量作为重中之重。

参考文献:

- [1] 宋旭红.论我国学术评价中的程序正当和结果公正[J].清华大学教育研究,2019(2):72-82.
- [2] 王小栋,王战军,蔺跟荣.中国研究生教育70年发展历程、路径与成效[J].中国高教研究,2019(10):33-40.
- [3] 周文文,余倩,倪加施.我国研究生学制的历史演变、现实问题与改革建议[J].研究生教育研究,2019(6):42-46.
- [4] 朱鹏宇,马永红,白丽新.新中国成立70年来研究生招生制度变迁逻辑:回顾与展望[J].中国高教研究,2019(11):27-33.
- [5] 罗英姿,李雪辉.我国专业学位博士教育面临的问题与改进策略——基于“全国专业学位博士教育质量调查”的结果[J].高等教育研究,2019(11):67-78.

- [6] 王顶明,杨佳乐,黄颖.我国研究生教育结构的现状、问题与优化策略[J].研究生教育研究,2019(2):1-5.
- [7] 高耀,陈洪捷,王东芳.博士生的延期毕业率到底有多高——基于2017年全国离校调查数据的实证研究[J].研究生教育研究,2020(1):42-51.
- [8] 绳丽惠.博士生延期毕业现象:影响因素与治理策略[J].学位与研究生教育,2019(6):60-64.
- [9] 于永凤,高连兴.研究生学位论文双盲审分类及典型方式对比分析[J].高等农业教育,2018(6):87-93.
- [10] 李艳,赵世奎,马陆亭.关于博士学位论文质量评价的实证分析[J].学位与研究生教育,2014(10):50-54.
- [11] 杜庆良.学术型硕士学位论文质量评价的实证分析[J].华北水利水电大学学报(社会科学版),2016,32(6):154-157.
- [12] 李尚群.权威的可谬性:学位论文评审规则的当下隐忧[J].研究生教育研究,2019(2):26-29.
- [13] 龚向和,张颂昀.论硕士、博士学位授予的学术标准[J].学位与研究生教育,2019(3):56-64.
- [14] 安洁.硕士研究生学位论文盲审效果研究及探索性建议[J].教育教学论坛,2013(42):176-178.
- [15] 汪小会,王莉.研究生学位论文质量双盲评审机制探析[J].黑龙江教育(高教研究与评估),2015(7):5-7.
- [16] 罗泽胜.硕士学位论文双盲评审意见异议机制研究[J].研究生教育研究,2018(1):44-48.
- [17] 鲍雪莹,陈贡,刘木林.基于履历信息的国际科技人才特征分析——以近十年诺贝尔物理、化学、生理或医学奖得主为例[J].现代情报,2014(9):4-9.

Research on Differences and Improvement Measures of Concealed Evaluation of Postgraduate Thesis

DU Xueling

(Institute of Mining Engineering, Guizhou Institute of Technology, Guiyang 550003, China)

Abstract: The concealed evaluation of thesis is an important means to improve the quality of postgraduate training, and the fair concealed evaluation is an important prerequisite for the discretion of the concealed evaluation. Taking 135 dissertations of 45 doctoral students of two majors in B college as a sample and applying the statistical survey method, this paper, on the one hand, analyzes the differences in the concealed evaluation of the sub-items in the sample, the maximum difference in the concealed evaluation results, the impact of familiarity on the concealed evaluation results, and the relationship between the concealed evaluation results and the later performance of the paper; on the other hand, this paper discusses the problems in the concealed evaluation system and improvement strategies based on the above analysis. The research shows that the concealed evaluation is conducive to the exercise of discretion; there are generally slight differences in the concealed evaluation; individuals are prone to bigger differences; the acceptable results of the concealed evaluation are not necessarily fair. Unfair evaluation is more likely to occur in academic dissertations with poor average grades, and authority's fallacy is more likely to happen in overestimated results. The research also shows that the more familiar the reviewers are with the content of the review, the more they dare to give a negative or positive evaluation, but for those who are roughly familiar with the dissertations, their review results may not be accurate. In terms of the later performance of the thesis, for the postponed students, the negative concealed evaluation results are relatively objective, but for the evaluation of non-postponed students, there may be improper evaluation among the differences in the concealed evaluation. In the future, it is necessary to improve the various links before, during and after the concealed evaluation on the basis of establishing a unified networked review platform to allow the concealed evaluation to return to the evaluation of the thesis itself to a greater extent.

Keywords: thesis; concealed evaluation; postgraduate; higher education; fair review

(责任校对 王小飞)