文章编号:2095-1663(2022)05-0040-07 DOI:10.19834/j.cnki.yjsjy2011.2022.05.05

考研分数能否精准预测研究生学业表现
——基于倾向值匹配(PSM)的实证分析

刘 进,陈恺哲

(北京理工大学 人文与社会科学学院,北京 100081)

摘 要:学术界对研究生选拔方式长期存在矛盾冲突。一方面,考研分数作为研究生水平的预测指标饱受质疑,另一方面,如果采取本科学业表现作为选拔标准尚缺乏实证验证。使用研究生入学之后各个学业表现指标,通过 PSM 倾向得分匹配等多个模型从各方面验证影响因素。结果表明,考研分数只影响学生硕士阶段 GPA 成绩,而学生本科阶段学业表现影响研究生科研与竞赛表现,考研面试成绩影响研究生毕业论文成绩。研究反映出当前研究生选拔制度的弊端,并从相关角度提出政策建议。

关键词:学业表现;考研分数;本科学业表现;PSM 倾向得分匹配

中图分类号:G643 **文献标识码**:A

全国硕士研究生招生考试是国家选拔人才的重要方式,在考研选拔中,“唯分数论”的导向一直受到学术界质疑。追根溯源,主要是对研究生招生考试预测准确性的怀疑,考研分数是否能代表学生日后的学业表现? 又在学业表现的哪些方面起到作用? 除考分外是否有更好的预测指标去选拔研究生? 这些问题亟待解决,为此,本文借助某 985 高校某学院研究生数据,深入验证二者之间关系,并探究多个因素对研究生学业表现的影响,为我国研究生选拔工作提供政策建议。

一、研究生考试“高分低能”论调日益尖锐

硕士研究生入学考试作为研究生教育的起点,对研究生培养起到至关重要的作用。近年来,通过研究生选拔考试脱颖而出的人才却往往受到“高分低能”的诟病,即只有考试分数特别高,但却缺乏创新意识和科研实践能力。例如,有研究者提出,当前研究生考试没有跳出“应试教育”的窠臼,其命题原则实际上鼓励了学生对考题的死记硬背,这样的考试方式很难挑选出有才能的学生,最后往往善于考高分的人进入宝塔顶尖,而真正有潜力的学生被埋没[1]。高升(2014)认为,现行研究生入学考试基本上是对考生本科阶段学习成果的“回顾性”考试,而非对学生未来研究生阶段发展潜能和综合能力的“预测性考试”[2]。陈睿(2016)则认为研究生入学考试范围有限,只能选拔出一些“考试机器”,却很难考查学生动手能力,语言表达能力等与创新能力有关的因素[3]。从已有研究结果来看,考研分数对研究生学业表现的预测准确性不容乐观。朱原(2016)对浙江大学 546 名硕士生的硕士毕业论文成绩与入学考试初试成绩进行相关研究,发现二者之间甚至出现显著的负相关关系[4]。涂罡等(2020)使用回归模型对山东某高校研究生进行建模分析,发现研究生学术科研能力与其初试、复试成绩无关,而与学业背景以及指导教师等因素显著相关[5]。

一些学者对这一现象产生的根本原因进行了讨论,概括起来主要有以下几个方面。一是研究生选

收稿日期:2022-04-01

作者简介:刘进(1985—),男,江苏东海人,北京理工大学人文与社会科学学院副研究员、国际争端预防和解决研究院研究员、博士生导师。
陈恺哲(1994—),男,浙江杭州人,北京理工大学人文与社会科学学院硕士研究生。

拔考试命题的问题。根据2021年9月教育部出台的《2022年全国硕士研究生招生工作管理规定》，现今研究生入学考试中，全国统一考试的全部或部分考试科目由教育部考试中心负责统一命题，其他考试科目由招生单位根据大纲自行命题[6]。采取自命题笔试的高校较多，水平也参差不齐。一方面在试题质量与管理规范性上与统考存在差距；另一方面初试笔试科目各不相同，复试阶段却采取统一划线，导致初试分数的可比性变差[7]。也有研究者认为，在评卷环节，各省、各招生单位分别组织评卷，导致同一科目在主观题评卷的过程中宽严程度难以统一，造成再次的不公平[8]。

二是研究生选拔过程重笔试，轻复试，学生非认知能力因素（如考生考研动机、意志品质）较难得到考查。近年来社会“学历门槛”急剧上升，由此也导致大量考研学生只是为了拿到硕士文凭改变工作处境，对学术科研毫无兴趣，被录取后对所学专业也缺乏认同[9]。但科技创新人才的成长恰恰需要学术热情以及坚定的意志支持[10]。虽然有复试的环节可以考察学生的动机和兴趣，但生源一般的招生单位，为满足扩招需要，往往采取等额复试，或者重初试分数，淡化复试表现[11]。因此，学生的考研动机和创新能力仍然难以得到有效的甄别。

三是高校对学生本科学业表现考查不足。与国内各高校看重研究生初试考试成绩的选拔模式不同，国外大学对于申请人认知因素进行考察的方式主要是本科阶段成绩（Grade Point Average，GPA）和标准化入学考试（如GRE）。Puryear等对美国107所医学院进行调查发现，本科GPA被89所医学院列为最重要或者第二重要的研究生选拔指标[12]。同时，Truell等以179名研究生硕士阶段的GPA作为因变量进行多元回归分析，发现他们在本科阶段的GPA是最为显著的预测变量[13]。虽然我国各大高校在硕士招生复试阶段也会要求学生提交本科成绩单，然而研究生复试周期较短，招生单位没有时间核实材料真实性，也就无法发挥本科成绩的评价作用[14]。同时，国外高校在研究生选拔的过程中，往往采用本科院校竞争性水平这一概念来衡量学校之间教学质量水平的差异[15]。例如，Hall发现在学生本科GPA的变异中70%都可以归因于本科学校竞争性水平差异[16]。因此，有研究者认为如果把本科院校竞争性水平作为本科GPA的权重因素，可以进一步提升本科GPA的参考价值[17]。在我国，部分招生单位在自主招生的过程中，也会看重申请者的本科院校水平，研究者指出，有30%高校在调剂的过程中会表明优先接收本科毕业于“985工程”，“211工程”及行业特色高校生源。有些高校甚至会在复试过程中优先选择本科生源更好的考生[18]。这是由于现行研究生招考制度难以分辨学生是否具备科研潜力。然而也有人对这一行为存在质疑，岳昌君对首都高校硕士研究生创新能力进行调查发现985高校的创新精神得分高于211高校，两者均高于普通本科，但是这几种差异在统计上并不显著[19]。

那么，到底是考研分数对研究生未来学业表现预测更准确，还是本科学校层次和学业成绩对研究生综合素质反映更全面？是否又有其他因素对研究生学业表现产生影响？对这些问题的回答，将有助于洞悉人才的成长规律，有望在未来制定出更科学、有效的研究生人才选拔制度。

二、数据描述与模型构建

（一）数据介绍

本实验数据来源于北京某985高校某学院2015年到2020年所有硕士研究生数据。数据由三个部分组成，第一部分是该研究生群体在入学前提交的原始本科成绩单，总计2570份，为了进一步研究需要，通过扫描全部转换为表格形式保存，得到约128000条成绩数据。第二部分是2015年到2020年研究生入学考试数据，录取名单共有3246名学生，其中统考生1720人，推免生1526人。统考考生占比53%，推免考生占比47%。之后对统考生数据进行深度挖掘，将考研各科目成绩单独作为各个变量研究，以判断不同科目成绩对学生学业表现的影响。第三部分是研究生入学后学业表现数据，对于学业表现的界定，要从多个方面展开，不仅要考虑研究生课程学习情况，还要考虑到研究生综合素质与创新能力。对于前者，选取研究生在校期间所有课程均分代表学生对课程的学习能力，共有3136条学生在校期间评价GPA数据。后者则选取研究生毕业论文答辩分数、在校期间学术发表与竞赛获奖情况作为依据。由于2020与2019级学生尚未毕业，共收集到2270条毕业论文成绩数据。此外，在统计研究生科研竞赛时发现，有45%的研究生有过论文发表或者竞赛得奖经历，说明在该专业领域此学院研究生学术表现较好。

经过对三个数据源的数据清洗后，借助于R语言将三部分数据按照学生学号为匹配项合并入同一数据库进行分析。由于存在小部分学生延毕，退学等其他原因引起的数据缺失情况，严格剔除掉各项存在缺失的样本数据，最后得到2178名研究生数据，每名学生学号下包含其考研各科分数情况以及本硕期间各项学业表现数据。合并后的数据库按照时间维度主要包括以下信息：(1)本科学业表现信息：本科学校、各门课程成绩与平均GPA、大学英语4级成绩、本科获奖情况等条目。(2)考研分数信息：其中初试成绩包含外国语、政治、专业基础课(数学一)、专业课(物理类)、面试成绩。(3)研究生阶段学业表现信息：研究生期间平均GPA、研究生阶段学术发表与竞赛获奖情况、硕士答辩论文分数。

(二)PSM倾向得分匹配模型

上述回归模型未能完全控制相关因素，难以涵盖同时影响研究生入学考试成绩和研究生学业表现的协变量。研究生阶段的学业表现可能受到该学生本科成绩、获奖经历等因素影响，因此产生内生性问题，倾向评分法（Propensity Score Method，PSM）是1983年Rosenbaum和Rubin提出，采用反事实模拟状态下对处理效应的估计，可以有效消除混杂因素产生的选择性偏差。本文中将研究生入学成绩前30%的学生设定为高分组，也就是处理组，后30%的学生设定为低分组，即对照组，PSM研究方法的思路是在通过logit或probit模型估计得到的倾向得分基础上，从对照组中挑选出一个与实验组各种特征均相似的个体进行匹配，从而比较那些其他特征相似，只有研究生入学考试成绩高低不同的学生，在学业表现上的差异。倾向值得分可以表示为：

$$P(x_i)=\Pr(D_i=1\mid x_i)$$

其中，x_i表示影响学业表现的变量矩阵，D_i表示第i个研究生学业表现，$D_i=1$表示为高分组学生，$D_i=0$表示为低分组学生。

得出倾向得分后，本研究将同时采用核匹配、最近邻匹配以及半径匹配对结果进行估计，以提高实证结果稳健性。匹配结果使用stata16进行结果输出，最后估算出ATT(Average treatment effect for the treated)，来表示研究生入学考试高分对于学业表现的平均处理效应，具体如下面公式所示：

$$\begin{aligned}ATT&=E(Y_{1i}\mid D_i=1)-E(Y_{0i}\mid D_i=1)\\&=E(Y_{1i}\mid D_i=1)-E\{E(Y_{0i}\mid D_i=0,P(x_i))\mid D_i=1\}\\&=E\{[E(Y_{1i}\mid D_i=1,P(x_i))-E(Y_{0i}\mid D_i=0,P(x_i))]\mid D_i=1\}\end{aligned}$$

Y_{1i}和Y_{0i}分别代表第i个研究生在高分组和低分组时潜在的学业表现情况，$D_i=1$表示高分组学生，$D_i=0$代表低分组学生，$P(x_i)$为之前得到的倾向得分。同样，为使研究更有说服力，将学业表现分为平均GPA、科研成果、毕业论文答辩分进行综合分析(表1)。

表1　PSM模型中变量定义与描述

自变量	变量定义
本科平均GPA	连续变量
本科学校层次	985、211高校=1，普通高校=0
本科获奖经历	有获奖经历=1，无获奖经历=0
考研初试分数	连续变量
考研面试分数	连续变量
大学英语4级分数	连续变量
研究生平均GPA	连续变量
硕士毕业论文答辩分	连续变量
研究生阶段学术发表与竞赛获奖	有相关经历=1，无相关经历=0

三、研究生学业表现影响因素实证分析

(一)入学分数与硕士研究生日后学业及科研表现无关，推免生入学后更擅长拿“高分”而非做科研

推免生的平均GPA要高于统考生，但是在同一类型的硕士研究生中，本科学校的层次不影响其学业表现。985高校统考学生研究生阶段平均GPA为83.62，211高校统考学生平均GPA为82.24，非985、211高校统考学生平均GPA为82.26。而985推免生中平均GPA为85.63，211推免生中平均GPA为85.59。同时，取每个统考学生入学考试笔试分数与其当年学校线的差作对数值，笔试成绩与在校平均GPA的皮尔逊积差相关系数值为0.023，$p<0.05$。虽然相关显著，但是相关性非常弱。统考学生复试分数与在校平均GPA皮尔逊积差相关$p>0.1$，相关不显著。统考学生笔试与面试成绩，分别与其日后有无科研以及其毕业论文答辩成绩做点二列相关与皮尔逊积差相关，显示p值均>0.1，

说明无显著相关。同时，也可以看出虽然推免生在学业表现（GPA）上高于统考生，但是在科研表现上与统考生无显著差异（图 1）。这一现象原因可能是由于获得推免资格的学生本身就是那些在本科阶段学业优异的学生，因此在研究生阶段比经过笔试选拔上来的学生更擅长于拿高分。

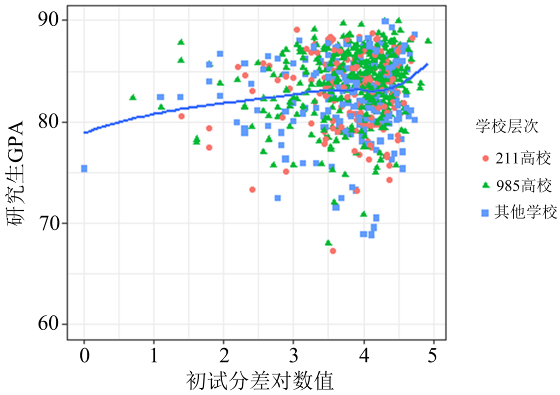

图 1　初试分数与研究生阶段平均 GPA 关系图

（二）学生英语能力与本科学业表现成为影响硕士阶段科研表现的关键因素

图 2 数据分析发现，研究生英语水平、本科学业表现与其科研能力具有相关性。将学生大学英语 4 级成绩作为变量，与其是否有科研成果进行点二列相关分析，发现二者之间相关系数为 0.255，且在 0.01 级别相关性显著。具有科研成果的研究生在本科阶段英语 4 级分数更高，其分段主要集中在 500 到 550 分，而无科研成果的研究生，本科阶段英语 4 级分数主要集中于 450 到 500 分。两类学生在本科 GPA 上也有明显差异，有科研成果的研究生在 80 到 90 分段上的人数要多于无科研成果的研究生。但是分布在 90 分以上的极端分数人数密度，无科研经历的学生人数更多。同时，本科阶段获奖经历的差异也将影响研究生科研表现。在本科阶段有获奖经历的学生，在研究生阶段具有科研成果的比例为 22%，而无获奖经历的学生，研究生阶段科研成果比例仅为 13%。一方面是由于在阅读文献和发表外文期刊上，英语能力强的学生有一定优势，另一方面，相比于一次考试制的研究生入学考试，本科四年的成绩以及获奖经历更能代表一名学生的科研潜力。

（三）211 高校推免生比 985 高校推免生科研表现更好

在平均 GPA 与毕业论文答辩分数上看，985 高校生源与 211 高校以及其他学校生源上无明显差异，但是在单独分析推免生科研表现时发现，985 统考，211 统考，非 985、211 统考学生，在研究生阶段产生科研成果的比例分别为 15.8%，19.5%，18.2%。而推免生中只有 985 及 211 学生，科研占比分别为 16.7%以及 32.1%，211 推免生是 985 推免生的两倍。可以看出，211 高校推免生有学术论文发表或竞赛得奖的比例显著高于 985 高校推免生，这与常理并不一致。继续深入分析发现，985 高校推免生中本科有获奖经历的占比 80.7%，而 211 高校推免生本科有获奖经历的占比高达 98.2%。这可能是由于在考核推免生资格的时候，985 高校学生凭借更好的本科优势，相对更容易进入推免名单，而 211 高校学生往往只有本科具有特别优秀的成绩及科研获奖经历，才可以拿到这一资格，因此在综合科研能力上基础更扎实。同时，在 985 推免生中，本校生源占据 63%。

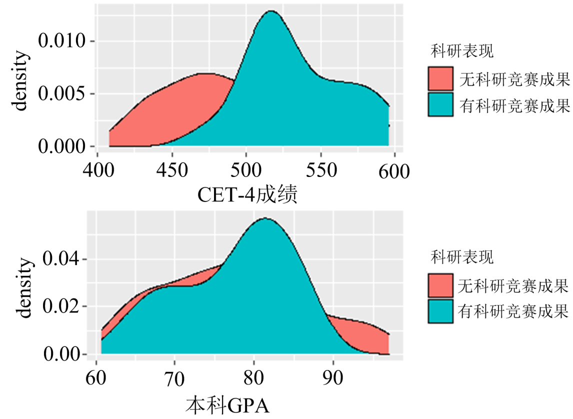

图 2　两类学生在本科阶段学业表现差异

（四）考研单科成绩对研究生各类学业表现均无显著影响

将研究生考研分数继续细分为各个单科分数建立 OLS 与 logit 回归模型，以深入考查是否某些科目对研究生学业表现有较好预测性。由表 2 可知，研究生阶段平均绩点与所有入学前考试因素均无显著关系。但在表 3 中发现，在毕业论文答辩成绩方面，研究生复试中的面试分数与其本科阶段大学英语 4 级分数的 p 值均在 0.01 水平上显著，系数为正，表示大学英语 4 级分数与复试的面试分数越高，该学生越倾向于在毕业论文答辩的过程中获得更高的分数。这说明那些统考学生的考研分数并不能保证其在入学后的学习过程中取得更好的学习成绩。同时由于毕业论文的答辩涉及到多方面因素，除了学生查阅文献与写作能力，也受到学生临场发挥和表达能力影响，因而研究生面试中取得高分可在一定程度上帮助该学生在答辩环节表现得更好。

表 2　多种因素对研究生平均绩点与毕业论文答辩成绩的 OLS 回归模型

解释变量	研究生平均绩点	毕业论文答辩成绩
外国语分数	0.059(0.121)	−0.011(0.140)
政治分数	0.069(0.130)	0.075(0.243)
专业基础课分数（数学）	0.022(0.292)	0.033(0.234)
专业课分数（物理类）	0.023(0.218)	−0.041(0.107)
面试分数	0.085(0.080)	0.179**(0.006)
大学英语 4 级	0.000(0.995)	0.135**(0.009)
本科平均绩点	0.029(0.287)	0.063(0.109)
本科获奖经历	0.734(0.133)	0.553(0.413)

注：***、**、* 分别代表在 1‰、1%、5% 的水平上显著。

此外，表 3 中 logit 回归模型显示，学生在研究生阶段在科研方面的表现依然与研究生入学考试各单科成绩无关，只与学生在本科阶段的各项表现有关。如表 3，大学英语 4 级成绩与本科获奖经历对研究生科研表现存在显著影响。二者分别在 0.01 与 0.05 水平上影响显著。这说明，学生考研的各项分数对学生日后科研能力或者创新潜力没有显著影响。而学生本科阶段的学业表现十分重要，由于本科平均绩点还受到学生本科学校竞争性水平影响而没有得到显著的关系，但在本科大学英语 4 级考试和获奖经历上表现更好的学生，在研究生阶段也更容易取得科研与竞赛成果。

（五）PSM 倾向得分法估计的考试分数对于研究生学业表现的影响效应

研究生入学考试分数对于学业表现的影响效应结果（见表 4）表明，考试分数的高低对研究生硕士学业表现的影响存在差异。在有关学业表现的三项变量中，硕士阶段平均 GPA 存在显著差异，高分组考生在硕士阶段平均 GPA 在三种匹配方法中分别高于低分组考生 3.326 分、2.536 分和 2.224 分。在最近邻匹配中 t 值为 2.84，结果在 0.01 上显著，而在核匹配与半径匹配中，t 值分别为 3.04 和 4.36，结果在 0.001 上显著。这说明那些在研究生入学考试中取得高分的统考学生，在研究生阶段也会获得更高的 GPA。而从学业表现变量中与学生考试能力无关的科研成果与毕业论文答辩分上来看，无论是最近邻匹配、核匹配还是半径匹配，实验平均处理效应均不显著，说明在科研能力或者综合学术创新能力上，高分组的研究生与低分组的研究生没有表现出明显的差异。

表 3　多种因素对研究生科研表现影响的 logit 回归模型

解释变量	α 值	Z 值
外国语分数	−0.053(0.109)	−1.60
政治分数	0.059(0.161)	1.40
专业基础课分数（数学）	0.032(0.104)	1.63
专业课分数（物理类）	−0.003(0.864)	−0.17
面试分数	0.018(0.674)	0.42
大学英语 4 级	0.017**(0.002)	3.10
本科平均绩点	−0.019(0.452)	−0.75
本科获奖经历	0.940*(0.042)	2.03

注：***、**、* 分别代表在 1‰、1%、5% 的水平上显著。

表 4　研究生学业表现倾向得分匹配结果

学业表现形式	匹配方法	高分考生（处理组）	低分考生（对照组）	实验平均处理效应（ATT）	标准误（SE）	t 值
平均 GPA	最近邻匹配	84.560	81.233	3.326**	1.173	2.84
	核匹配	84.653	82.117	2.536**	0.834	3.04
	半径匹配	84.560	82.336	2.224***	0.510	4.36
科研成果	最近邻匹配	0.196	0.130	0.065	0.144	0.45
	核匹配	0.162	0.150	0.012	0.091	0.13
	半径匹配	0.196	0.128	0.068	0.064	1.06
毕业论文答辩分	最近邻匹配	85.204	86.061	−0.857	1.246	−0.69
	核匹配	84.978	85.627	−0.649	1.143	−0.57
	半径匹配	85.204	84.893	−0.311	−0.789	0.39

注：最近邻匹配采用默认的单一匹配（$k=1$），***、**、* 分别代表在 1‰、1%、5% 的水平上显著。

PSM 倾向得分匹配模型需要在运用过程中满足条件独立假设，保证匹配前后的语文高分组与低分组不存在明显系统差异。需要在分配后进行平衡性检验，采用核匹配的方式对各变量进行平衡性检

验，结果如表 5，所有变量在匹配后均值接近，且标准偏差在匹配后大幅降低，均控制在 5%左右或小于 5%，且 t 值均不拒绝高分组与低分组无系统差异的原假设。因此，该匹配平衡性得到了较好控制。

表 5　平衡性检验结果

变量名	是否匹配	高分组	低分组	标准偏差	偏误消减	t 值
本科 GPA	匹配前	83.52	82.54	27.2%	75%	2.50
	匹配后	82.94	83.19	−6.8%		−0.46
本科学校层次	匹配前	0.67	0.70	−8.7%	40.2%	−0.80
	匹配后	0.67	0.65	5.1%		−0.50
本科获奖经历	匹配前	0.43	0.50	−27.6%	85.1%	−12.69
	匹配后	0.43	0.44	−4.1%		−1.76
考研面试分数	匹配前	79.67	73.58	55.3%	94.8%	5.06
	匹配后	79.27	78.96	2.9%		0.41
大学英语 4 级分数	匹配前	502.88	448.42	88.3%	96.7%	8.10
	匹配后	502.86	501.06	2.8%		−1.77

四、研究生选拔应更重视学生综合能力及本科学业表现

本研究对北京某 985 高校某学院研究生数据进行相关分析和多角度数学建模。模型中具体分析考研笔试中每一门科目的影响作用，并将学生本科阶段的学业表现纳入到模型，全面探究研究生选拔过程中各个影响因素，为研究生入学考试的改革提供实证性的数据支持。研究主要有以下发现，第一，当前研究生入学考试选拔的是考试能人而非创新型科研人才。因而在 PSM 倾向模型中初试高分组的学生相比低分组的学生，在入学后的考试表现更好，取得了更高的 GPA，但是在科研成果和答辩论文分数上与低分组学生无显著差异。说明当前研究生选拔机制重知识，轻实践创新现象严重，通过考试分数筛选出的学生与当前研究生培养教育的宗旨相违背。第二，本科学业表现对研究生入学后表现影响至关重要，尤其是本科获奖经历与大学英语 4 级成绩。一方面，本科就积极参与各类竞赛或者学术发表的学生，一定程度上可代表学生的研究热情与综合创新能力，而这类学生具备更高的科研潜力。另一方面，在研究生阶段需要更强的论文写作能力，而英语优秀的学生无论是在外文文献阅读还是在外文期刊上发表都有明显的优势。这也是为何在研究生复试或者调剂阶段，众多院校逐渐开始重视学生各类英语证书的原因。第三，本科学校竞争性水平在推免生群体出现“反向作用”，具备科研潜力的学生比例在 211 高校群体中更高。这主要是因为 985 高校对 211 高校推免生的要求更为严格，学生需同时拥有优异的学业成绩以及科研奖项才能通过筛选，而这些因素在模型中都对学生科研潜力具有显著的影响。反而 985 高校推免生，尤其是本校的推免生条件较为宽松，且最优秀的人才往往选择海外高校。

北京某学校某学院是在该专业领域学科排名靠前，发展速度较快的学院。对于该学院研究生数据分析得到的有关发现对于当前中国研究型人才选拔和培养具有一定的典型意义。基于本研究的结果，本文提出如下政策建议：

一是改革研究生入学考试内容，增加综合能力考察，促进初试命题统一性。当前研究生入学考试内容主要注重基础和通识的考察，导致高分考生会考试而不会做学术研究，因此在命题过程中增加专业能力和创新能力试题尤为重要。同时，初试中众多高校采取自命题形式，难以保证笔试环节的公平性与科学性，应当逐步采取初试命题国家统一，复试由各校自行组织的统分相结合形式。

二是缩小初试分数占比，采取多种形式考核策略。研究生入学考试的目的是选拔专业能力与科研潜力强的学生，而这些能力在笔试分数中无法得到全部反映，例如从模型中可以看出，面试表现更好的学生易获得更高的论文答辩分，究其原因这与学生的表达、沟通能力相关。应将高校多样性和自主性选拔权力在复试中得到体现，使用多种策略考察学生各项专业能力。如在复试中增加学生外文文献阅读，现场实验操作以及学术汇报等环节，多角度判别学生是否适合科学研究道路。

三是加大对学生本科学业表现考察，联动学生本科高校进行评价。当前各大高校在复试环节一般都会要求学生提交本科成绩单以及本科获奖经历信息。但这一制度往往流于表面，从提交到录取结果公布相隔如此短暂，本科学业表现很难得到充分考

察。相反在国外，学生本科学业成绩单及获奖经历往往在申请环节就将提交给学校进行筛选。笔者认为，应要求学生在报名研究生入学考试之时就提交本科成绩单，给目标院校留下充分的时间将其量化进行比较，甚至可以作为初试分数的加分项，给予小部分本科学业优异但不擅长考试或发挥失常的考生一次复试机会。而在复试或者推免过程中，除了学生本人提交本科成绩单，本科高校也应当提供对该学生的综合评价，包括学生具体的获奖以及处分记录，教师推荐信等内容，保证高校对学生本科学业表现评价的真实性和多样性。

参考文献：

[1] 肖立山，夏从亚. 对研究生入学考试命题原则的再认识[J]. 中国高教研究，2007(8)：23-24.
[2] 高升. 关于全国硕士研究生入学统一考试初试科目设置的思考[J]. 中国高教研究，2014(6)：31-35.
[3] 陈睿. 硕士研究生招生初试内容与形式的再思考[J]. 学位与研究生教育，2016(1)：47-51.
[4] 朱原，李华静. 公共政策的公平性与有效性：对研究生招生政策的评估与建议[J]. 中共宁波市委党校学报，2016(2)：112-117.
[5] 涂罡，王杰茹，毕秀玲. 创新人才选拔培养的特征性事实与实证研究—基于研究生学术科研能力的分析[J]. 山东财经大学学报，2020(6)：110-117.
[6] 2022年全国硕士研究生招生工作管理规定[EB/OL]. (2021-8-30)[2021-10-1]. http://www.gov.cn/zhengce/zhengceku/2021-09/04/content_5635387.htm.
[7] 胡伟力，张立迁. 我国硕士研究生初试自命题工作的现实困境及对策研究[J]. 学位与研究生教育，2021(2)：53-57.
[8] 李新华，孟宪军，王春超，等. 基于构建"国考 ＋ 校试"模式的硕士研究生统一入学考试招生制度探索[J]. 高教探索，2018(3)：87-91.
[9] 管国锋，吴松强. 基于学科交叉研究生创新能力培养机制研究[J]. 江苏高教，2017(5)：71-73.
[10] 王越，刘进，马丽娜，等. 从新中国70年看高等教育如何培养科技创新人才—王越院士专访[J]. 重庆高教研究，2019(5)：5-13.
[11] 孙友莲. 硕士研究生招生制度的问题及改进建议[J]. 江苏高教，2014(1)：95-97.
[12] Puryear J B，Lewis L A. Description of the Interview Process in Selecting Students for Admission to U. S. Medical Schools[J]. Journal of Medical Education，1981，56(11)：81-85.
[13] Truell A D，ZHAO J J，Alexander M W，et al. Predicting Final Student Performance in a Graduate Business Program：The MBA[J]. Delta Pi Epsilon Journal，2006，48(3)：144-152.
[14] 肖文英，孙秋红，孙乐平，等. 硕士研究生入学考试复试中科学性与公平性问题研究[J]. 学位与研究生教育，2015(2)：56-60.
[15] Barron E S. Profiles of American Colleges：29th ed[M]. Hauppauge，NY：Barron's Educational Series，2010.
[16] Hall F R，Bailey B A. Correlating Students' Undergraduate Science GPAs，Their MCAT Scores，and the Academic Caliber of Their Undergraduate Colleges with Their First-year Academic Performances Across Five Classes at Dartmouth Medical School[J]. Academic Medicine，1992，67(2)：21-23.
[17] 薛刚，孙晓敏. 如何有效选拔研究生－国外研究生选拔实践的证据[J]. 教育科学，2011，27(4)：67-78.
[18] 李安萍，陈若愚，胡秀英. 我国硕士研究生招考中自主与自律的误区[J]. 高校教育管理，2017，11(2)：93-99.
[19] 岳昌君，吕媛. 硕士研究生创新精神特征及影响因素分析[J]. 复旦教育论坛，2015(6)：20-26.

Can Postgraduate Academic Performance Be Accurately Predicted with the Entrance Examination Score：An Empirical Analysis based on Propensity Score Matching

LIU Jin，CHEN Kaizhe

(*School of Humanities and Social Sciences*，*Beijing Institute of Technology*，*Beijing* 100081，*China*)

Abstract：There are long-standing disputes in the academic circle on postgraduate selection method. On the one hand，postgraduate entrance examination scores used as a predictor to judge the level of postgraduates is in question. On the other hand，if it is better to use undergraduate study performance as the selection standard still lacks empirical verification. This study uses several models such as PSM tendency score matching to analyze the academic performance indicators of the postgraduates matriculated so as to verify from various aspects the influencing factors. The results show that the postgraduate entrance examination score only influences the GPA score of postgraduates during the period when they study for master's degrees，the academic performance of postgraduates in their undergraduate study period influences the research and competition performance of postgraduates，and the interview score from the postgraduate entrance examination influences the score of dissertation for graduation. This study exposes the disadvantages of the current postgraduate selection system and puts forward some policy suggestions from relevant perspectives.

Keywords：academic performance；postgraduate entrance examination score；undergraduate academic performance；PSM (propensity score matching)