

“任人唯贤”的制度悖论*

——以明代科举为例

柯荣住 洪伟 郑恩营

提要:本文总结了人才选拔的三个重要制度目标,即公平性、匿名性与有效性,并以这三者间的矛盾和取舍为框架梳理有关任人唯贤的文献与历史实践。基于我国科举制巅峰时期(明代中期)的制度细节和对中国传记数据库中12752名进士信息的量化分析,本文揭示了集中式人才选拔制度在其设计与执行过程中的悖论,即难以同时协调上述三个目标之间的冲突。研究发现,任人唯贤作为一种社会选拔制度,既受制于选拔者的能力与私利,又受制于区域或个体间的起点不公平。

关键词:任人唯贤 科举 制度悖论 明代中期 定量分析

一、导言

“任人唯贤”(meritocracy,亦称贤能主义)是指以个体才能而非家庭出身为基础分配职位或资源的选拔制度(Scully,2014)。任人唯贤是无数先贤的共同追求。墨子在两千多年前就提出了“故官无常贵而民无终贱。有能则举之,无能则下之”的主张(方勇译注,2007:52)。孔子也强调贤能对为政者的重要性,认为为政之道“先有司,赦小过,举贤才”(张燕婴译注,2006:186)。更重要的是,儒家“学而优则仕”的观点为教育与政治精英的选拔提供了理论和实践的依据。

在理念层面,任人唯贤有三个优点。一是提升社会效率。这里隐含的假定

* 本研究得到北京大学人才启动项目(7101502649)、浙江省哲学社会科学规划重大课题(22QNYC03ZD)的资助。感谢张明君、赵卓、徐灿对数据工作的支持。特别感谢匿名评审人对本文的建设性修改建议,同时也感谢龚启圣、高翔、白营等学者对本文的修改建议。文责自负。通讯作者郑恩营(Email:enyngzheng@nsd.pku.edu.cn)。

是职位与才能之间具有“互补性”,按照才能高低匹配职位可以比随机配置创造更多的社会价值。这个假定是由社会系统所决定的,也是社会成员广泛接受的规范。二是产生激励与示范效应,使高能力者更努力,进而产生更高的社会价值。虽然这可能会削弱低能力者的努力效果,但对于整体而言益处更大。三是提高社会流动性。如果根据家庭等先赋因素分配职位,有才能但出身不好的个体难以获得机会,容易导致阶层固化。

然而,在实施层面,要实现任人唯贤却并非易事。以传统中国的政治精英选拔为例,从世禄制到察举制、九品中正制的实践转型,为强调人才选拔中的自致因素奠定了基础。科举考试于隋朝应运而生,统治者允许符合基本要求的男性参加由国家组织的选拔性笔试。经唐宋等朝代演变,这一选拔机制逐渐成为选拔政治精英的主要渠道,并在明代达到顶峰,于清末被废除。后科举时代的考试制度仍是现代社会人才选拔的重要形式。人才选拔制度与人才的定义、社会价值取向以及国家意志有关。我国的高考以及公务员考试均沿袭了科举中任人唯贤的相关理念和某些实施细则,成为人才选拔的核心环节。刘精明(2014)通过分析在校大学生的调查数据发现,学生主要是依靠个人能力进入高校,这符合任人唯贤的理念。

与任人唯贤制度相伴的是对以考试作为资源分配方式的批判。如李约瑟(Joseph Needham)、何莫邪(Christoph Harbsmeier)认为,科举阻碍了工业革命在中国的发生(Needham & Harbsmeier, 1998)。近年对高考的反思先后推动了三个方面的改革尝试(郭振有, 2005; 李春玲, 2014; 梁晨等, 2017; 吴晓刚, 2016; 谢爱磊等, 2018):考核内容更多样、考核形式更灵活,并且高校招生权限更大。然而,与之相伴的招生腐败又引发了新的问题。对任人唯贤的探究并非中国独有,英国社会学家迈克尔·杨(Michael Young)最早在英语中提及贤能主义一词,是为了批评单一的考试模式会对个体及社会产生负面影响(Young, 1958)。有鉴于此,当代西方社会推行素质教育,强调如领导力等主观性较强的才能定义(卡拉贝尔, 2014),但这也导致精英大学招生丑闻频发,引发人们对贤能主义的反思(Taylor, 2021)。

那么,实现任人唯贤的核心矛盾是什么? 什么样的制度设计能在更大程度上实现贤能主义? 本文旨在剖析任人唯贤制度目标之间及其实施过程中的悖论,即“不可能定理”。贤能主义旨在实现三个目标:一是公平性,指在进入选拔前的起点公平。起点公平并非消除先赋因素的差异,而是强调在政策层面为个体提供更公平的教育和发展机会。二是匿名性,指在选拔过程中不参照任何个人特征,平等地对待所有人。三是有效性(狭义),指制度目标的实现程度,即选

拔结果与制度目标的一致程度。从理念来看,这三个制度目标互相支撑:若不在选拔过程中强调匿名性,则公平性也难以保证;若没有公平性与匿名性的保证,任何选拔目标都有可能被扭曲;若有效性得不到保证,则不能持续获得社会认可,人才选拔也就无从谈起。

然而,社会是个复杂系统,任人唯贤制度只是其中一个组成部分,而且其三个制度目标在实施过程中存在着悖论:提高选拔的有效性需要对人才进行全面考核,但在严格遵循匿名性下是难以实现的;若因此降低对匿名性的要求,则需赋予选拔者更大的自由裁量权,这就难以保证公平性和有效性;此外,全面考核还会增加个体的投入,进而放大起点不公平,不利于来自弱势家庭的被选拔者。

为厘清这三个目标间的悖论和取舍,本文以明代中期的科举为例,利用历史档案进行量化分析。一方面挖掘微观数据折射出的制度实施细节,另一方面也尝试跳出特定的历史场景,将贤能主义放到更为普遍的背景下进行综合考察。从实证的角度看,科举是体现贤能主义制度悖论的典型案例。一方面,它经由国家严密设计,全方位实施,深刻影响到社会各阶层,是一项资源投入巨大、颇具有有效性的人才选拔制度;另一方面,在实施过程中它又面临种种掣肘,是一个典型的难以兼顾公平性、匿名性和有效性的案例。

本文期望在以下方面有所贡献。第一,从组织社会学的视角对明代中期的科举进行实证研究,剖析贤能主义制度如何嵌入于复杂的社会系统之中。第二,通过检验科举考试中社会关系和才能的相对作用,明晰任人唯贤制度的有效性程度,以及匿名性、公平性对人才选拔的影响。第三,以史为鉴,探讨对人才选拔制度设计的政策建议。本研究试图说明:任何具体的选拔制度都有其局限性。在类似科举的集中选拔模式下,公平性、匿名性和有效性三者存在内在矛盾,政策制定者需要根据主要矛盾做出适当的取舍。

二、任人唯贤:制度运行的微观基础及其悖论

鉴于以往文献缺乏将任人唯贤作为制度的研究,在此,我们汇总有关教育、求职、升迁等相关领域的实证研究,提炼其背后的制度逻辑及实施细节。因西方学术界对此的实证研究积累较多,本文以英文文献为主总结了相关的机制和机理,同时结合国内现状,梳理出更具普遍性的理论框架。

自20世纪70年代以来,新制度主义理论(Meyer & Rowan, 1977; DiMaggio &

Powell, 1983) 成为制度研究的重要理论, 它强调制度环境的重要性, 探讨由文化、规范和法律等不确定性所引发的组织趋同行为。后续研究侧重于制度运行的微观基础及其影响 (DiMaggio, 1988; Hallett & Ventresca, 2006; Powell & Colyvas, 2008), 关注社会行动者在一定时空权力背景下如何有选择地制定和执行制度, 为抽象的制度与微观运行机制提供了衔接。

“meritocracy”作为一个独立的英语词汇, 最早出现在杨 1958 年出版的《贤能主义的兴起: 1870—2033》(*The Rise of Meritocracy: 1870 - 2033*) 一书中。在书中, 杨批判了英国的教育改革, 认为以单一的分数评价学生会导致异化。与其初衷相悖的是, 贤能主义一词迅速成为一个制度神话, 2001 年杨还在《卫报》上针对当时英国首相布莱尔的发言再次阐述过度追求单一评价标准的危害 (Young, 2001)。近年来西方社会对不平等问题的关注促进了对贤能主义的实证研究和哲学思辨 (Sandel, 2020; Markovits, 2019)。从理念上看, 贤能主义极富吸引力, 但就实践层面的难度来看, 丹尼尔斯 (Daniels, 1978) 和斯库利 (Scully, 2014) 总结出三个难点 (简称 DS 框架): (1) 对于社会认可的才能应该有清晰、明确的界定, 其考核标准应该相对客观透明。(2) 社会应该为个体学习、展现才能提供相对公平的机会。(3) 对才能的奖励应该与其贡献相称。这三点与上文所述的“不可能定理”密切相关。限于 DS 框架背景的局限性, 本文将结合他们的论述, 以公平性、匿名性和有效性为要点梳理相关的文献。

(一) 起点公平

针对起点公平, DS 框架下的定义之争和机会平等是两块基石。所谓才能的定义之争, 也即让社会成员普遍接受的选拔标准是什么。才能的具象化是社会建构的, 具有垄断性和排他性, 对才能的定义反映了特定时空背景下的权力关系。不同的定义对先赋因素差异有着加剧或弥合的作用, 对促进社会流动的作用不同, 并最终影响对贤能主义理念的认可程度 (卡拉贝尔, 2014)。

以科举的变迁为例, 从唐代的二十多个科目到明代对儒家经典的强调, 考试内容变得更单一、范围更狭窄、形式更固定、答案更标准化, 导致人们对科举能否选拔出治理国家的人才持怀疑态度 (Ho, 1964)。但若考试内容过于广泛, 资源匮乏的考生缺乏了解和准备的机会, 会加剧既有的不平等; 同时, 若在评阅阶段评判标准的主观性大, 对考官的能力和品性要求过高, 操作性则不足。无独有偶, 近年来我国对高考制度改革的各种尝试背后也隐含了对才能定义的合法性之争: 应试教育与素质教育所考察的内容在多大程度上加剧了既有的不平等?

对此,不同群体因立场不同,所倡导的改革方向也有较大差异。

西方社会尽管没有科举传统,但对如何定义才能也极富争议。以卡拉贝尔(2014)的经典研究为例,他剖析了哈佛、耶鲁和普林斯顿三所大学录取标准的百年变迁史,发现才能定义的变更是传统精英审时度势下的妥协。更主观的才能定义(如强调领导力)有助于提高学校录取委员会的自由裁量权,在社会更加开放的前提下继续倾向于录取来自传统精英家庭的学生。该标准虽然不再强调考试成绩,但其具有的灵活多样性让人无所适从,这在实际运行中反而削弱了人尽其才的可能性,^①实际上成为阶层再生产的隐秘武器。

毋庸置疑,不同的才能定义下机会平等的实现程度存在差异。对于主观性较强的才能定义,来自资源匮乏家庭的学生既无法理解模棱两可的才能标准,也缺乏机会去发展、展示社会认可的才能(Golden,2007;Hoxby & Avery,2013),进而影响到他们进入精英大学的可能性(可汗,2016)。即便对精英大学的学生而言,能否展现此类才能也会影响其求职结果(Rivera,2015)。

(二)匿名性基础

匿名性不仅是一种技术性手段,也反映了对公平的认可程度。然而在很多情况下,非匿名几乎是必然的选项。由于存在信息不对称,选拔者很难真正了解候选人。解决该问题的一个途径是寻求第三方的私人信息,那么伴随而来的往往是非匿名性。如在欧美社会盛行的推荐信制度,其有效传递私人信息的前提是评估人需阅读推荐信,与其他材料相印证并形成对候选人的全面评估。如果无法保证评估人的能力和品性,那么贤能主义的实施可能流于形式。

因数据及伦理要求,已有研究很难验证匿名性是否削弱了教育领域的系统性歧视,但在劳动力市场的实证研究领域,相关发现不胜枚举(Aslund & Skans,2012)。以戈尔丁(Claudia Goldin)和罗斯(Cecilia Rouse)对某交响乐团雇用音乐家时的盲评为例,当隐去简历信息且在考场中拉上一个帘子隔开求职者与打分者时,女性求职者被录取的可能性更大(Goldin & Rouse,2000)。匿名性规则在优惠政策(affirmative action)下遇到了新挑战。优惠政策多以群体身份为基础展开(Zwart,2005),在具体执行时需要将候选人所属的群体身份进行归类。这

^① 马克维茨(Daniel Markovits)的新书《贤能制度的陷阱:美国的奠基性神话是如何促进不平等、消解中产阶级和侵蚀精英的?》(*The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite*)以从幼儿园到法学院一直接受精英教育的律师为例,说明过度竞争使得所有人都筋疲力尽,但却在个体与社会层面没有显著收益(Markovits,2019)。

种做法虽然对他们的个人信息进行了匿名化处理,但强化了群体身份的信息,导致了另一维度上的不公平。

(三)有效性

本文取有效性的狭义定义,即目标与结果之间的一致性程度。除了实施过程不尽如人意外,令目标与结果之间出现偏差的一个主要原因在于目标过于理想化,缺乏可操作性。制度涉及多元行动者之间的博弈,尤其是激励问题(Holmstrom,1979)。决策制定者未必有信息优势,在制定目标时需预判社会行动者的反应;过于理想化的目标可能扭曲行为,适得其反。制度的有效性往往是追求次优而无法做到最优时实现的;若不切实际地把最优作为目标,那么很可能连次优都达不到。

次优的目标并不是自动实现的,而是需要在实施过程中对行动者加以激励,使其利益与制度目标一致。这也意味着需要关注制度实施的微观基础。以学校招生为例,如果考官腐败,全方位考察的申请考核制可能更低效。美国大学对才能的定义主观性较强,自由裁量权更大,徇私舞弊的手段更隐秘。申请制是否比高考更有效?答案目前尚不确定。但近年来有关职位招聘的实证研究发现,简单、清晰、客观的量化标准更有效(Estrada,2019;Hoffman et al.,2018)。

综上所述,贤能主义的三个制度目标之间存在悖论。要做到起点公平和有效,在考核过程中则无法做到完全匿名。而要做到匿名和有效,就难以保证起点公平。要追求匿名和起点公平,只能让候选人自我提名或选择一个同质性目标,这等于放弃了选拔,也就因此失去了有效性。当然,对这三个目标无需在所有节点都等量齐观,基于整体考量,在人才选拔的不同节点可以有所侧重。例如,在高考阶段可能更需要侧重匿名性和起点公平,而研究生招生阶段相对更需要侧重有效性;在地方和部委公务员的招聘过程中也有类似的差别化设计。目标错位及侧重点不当容易引发对贤能主义的质疑。

三、明代中期科举的制度设计

本节回到明代中期科举的制度设计,主要关注以下四个方面。本文不对科举制度的演进溯源,感兴趣的读者可参见相关的断代史和专题史研究(Kracke,1953;何炳棣,1964;Hartwell,1982;Chaffee,1985;康文林,2020)。

(一)更具可操作性的才能定义

科举制度纵贯中国历史一千三百多年,到明代中期已经高度制度化,尤其是在考察内容方面。明代科举侧重考查考生对儒家经典四书五经的掌握程度,并且对四书五经的理解以朱熹的注释为标准,不鼓励个人发挥;对考生应答的文体和格式也有具体的要求。^①从选拔考试的有效性看,这些变化为考生提供了更清晰、明确的标准,有助于弥合家庭资源上的差距;同时,在评阅阶段,因评判标准更加透明、统一,限制了考官的自由裁量权,也更容易维护公平。从这个意义上讲,才能定义的合法性需要在全面考察与公平性之间折中,而明代科举的制度设计在这个维度上做出更倾向于后者的选择。

(二)对起点公平的重视

明代科举为才能培养提供了更公平的机会。除家庭和社会积极助学外,官学的发展也达到了一个新高度。1369年,明朝政府规定每个县及以上行政单位要建一所官学,以笔试成绩选拔当地学生,且报名时严格审核生源地资格。只有官学学生才能报名参加三年一度的科举考试。据郭培贵(2014)统计,明初有1435个行政单位,同期有1318所官学。官学学生享有学费减免、奖学金支持、赋税减免、免除徭役等福利,有助于弥合家庭资源差异带来的起点不公平。

明代科举在程序上更为规范、严密。科举由三场独立的考试组成,第一场是农历八月在14个省的省会举行的乡试。^②明代实行严格的户籍制度,考生必须在本省参加乡试,竞争有限的通过名额。通过乡试的考生赴京参加翌年二月的会试,这个时间差为偏远地区的考生提供了公平竞争的机会。会试通过后则需留在北京参加三月的殿试,殿试不淘汰考生,但会重新排名,名次会影响进入官场的速度和起点。此外,考生远赴京城备考所需的费用也有官方和民间的资助,如各地集资在北京建会馆,为考生提供食宿(何炳棣,1966)。

① 一种观点认为,尽管四书五经的内容与真实的治国才能之间没有必然关系,但是通过测试对四书五经的掌握程度,未必不能选出有真实水平的人才。根据艾尔曼(Benjamin A. Elman)的描述,四书五经共约40万字,假定一个人从7岁发蒙,到20岁参加科举(1400—1580年间年龄最小的进士仅13岁),那么大约每天要背诵85个字(Elman,2000)。这要求有相当强的意志力与记忆力才可以做得到。而记忆力与智商和学习能力往往呈正相关,意志力也的工作态度以及决策能力相关。所以只要能选出高智商与高意志力的人,即便考试内容与能力没有直接关系,选拔方式仍具有合理性。

② 明代有15个省。由于云南和贵州两省人口较少,乡试时两省联合举行,直到1537年两省才单独组织乡试。本文使用两省的合并信息。此外,由于篇幅和数据所限,本文没有把讨论的重点放在乡试上,因为通过乡试但在随后的会试中失败的考生可以重复参加会试,而不必再经历整个过程。但现有数据没有标注考生通过乡试的时间,故而无法对考生的乡试成绩进行客观、系统的比较。

(三) 地区配额制下的匿名性

明代科举延续了宋代以来实施的弥封制,但在匿名性上更加严格。以会试为例,组织者以考号替代考生名字,隐去所有个人信息;为防止考官通过笔迹识别考生,会雇佣 200~300 名专业抄写员将所有考卷誊写一遍。艾尔曼(Elman, 2013:227)对此总结道:考场内的过程暂时抹去了考生的姓名、家庭和社会地位。每一个人都被假定为是未知的,因而在考官眼中是平等的。

如此严格的匿名措施却因地区发展问题而有所妥协。为避免地区间在科举考试中的优劣过于悬殊,明代开始在会试阶段设置地区配额制。^① 配额制起源于 1397 年的南北卷一案,于 1427 年正式实施,是当代人才选拔优惠政策先驱。相关研究可参见田方萌对配额制的综述,以及莱斯利等对优惠政策后果的总结(田方萌,2017;Leslie et al.,2014)。虽然其初衷是矫正地区间的不平等问题,但在执行过程中难免要标注考生的地区来源,以确保录取结果符合配额规定。^② 这一地域信息的披露为考官在阅卷过程中庇护同省考生制造了机会,这也是本文定量分析的重点。

(四) 科举中的社会关系

对任人唯贤最大的挑战莫过于社会关系。本文重点分析与科举有关的两类社会关系:一是来自同省的考生与考官之间的同乡关系,二是考官与考生父亲间的同年关系(同一年中进士)。

基于地缘的同乡关系对中国社会有着深刻的影响(Opper et al.,2015)。同乡关系因地区配额制在明代科举中得到进一步强化。严格的户籍制度限制了自由迁移,即使进士遵循回避原则去家乡外做官,退休之际也需告老还乡;他们与家乡的联系既是心理上的,也是利益导向的。由于配额制使会试评卷时最多只

① 明代华南地区(占总名额的 55%)包括浙江、江西、福建、南直隶(首府为应天)、广东、湖广等省;华北地区(占总名额的 35%)包括北直隶(首府为顺天)、山东、山西、河南、陕西等省,华中地区(占总名额的 10%)包括广西、四川、云贵(云南、贵州)等省。基于进士籍贯的数据可判断,这一政策的确得到了严格执行。

② 现有史料对此细节记载不详,无法确定在卷面上标注的是三个地区之一还是 14 个省之一。研究明代科举的专家郭培贵(2014)认为至少地区信息是标注的,但也不能否定在此基础上能够进一步得知省份信息,毕竟每个地区只有 3~6 个省份。艾尔曼(Elman,2013:27)则提出考官可能知道省份信息,他使用“官僚体制漏洞”来描述配额现象。但不管是区域信息披露还是省份信息披露,对主要结论都没有实质性影响。为了行文方便,这里用同省关系来测量定量分析中的同乡关系。在稳健性检验中,也用同一地区作为关系的指标,结论没有改变。为使行文顺畅,以下提到地区配额制时通称省份信息、同省等。

能看到考生来源省份信息,无法精确到更细的方面,本文将同省作为老乡关系的边界。1400—1580年的46次科举平均取士288人,即10年(约3次科举)每个省平均约有60名进士活跃于官场。对进士们来说,同省老乡是有效的身份认同,在此基础上提携彼此,形成了官场的非正式圈子。

科举制度下的另一个重要概念是同年关系,也称年谊。明代有1318所官学,平均每所官学通过科举的考生数量极少,现代意义上的校友关系在此并不适用。考生需要背井离乡到北京参加会试和殿试。史学家卜正民(Brook, 2010: 150)将共同参考的过程概括为高度社会化的经历:“如果你通过了,那些和你一起通过的就成了你的同道中人,彼此交往,呼风唤雨,成为一生一世的朋友。”因此当考官与考生父亲是同年进士时,可能会偏袒该考生,将其排名前提。但由于会试严格的匿名程序,考官无从得知考生的个人信息,无法以同年关系为基础来徇私。而殿试阅卷并不匿名(Elman, 2013: 56),考官因此有机会获知考生个人信息,为基于同年关系的庇护提供了机会。

四、数据与方法

(一)数据和关键变量

本文数据的主要来源是哈佛大学维护的中国传记数据库(CBDB)。原始数据包括51次科举考试通过的14116名考生信息,但其中有5次科举关于考生的信息缺失较多。本文将重点放在记录相对完整的1400—1580年间的46场科举考试,其间共产生了12752名进士。此外,本文额外匹配了史料记载的考官信息(郭培贵, 2014),生成同省关系和同年关系这两个本文的关键解释变量。

对考生才能的测量是一个难题。现代社会一般用标准化考试或者智商测试来测量认知能力,但明代没有类似的标准化测试。对此本文采用折中的办法,对考生中进士时的年龄进行标准化处理,并对之反向排序,以此来衡量才能。这样做的理由有两点:首先,考生决定何时参加科举考试在某种程度上反映了他事前对自我才能的判定;其次,考生能否在某一年龄考中进士在事后也反映了此前试错的次数,侧面反映其才能。明代科举不限年龄,只要符合资格即可重复参加三年一度的科举,直到考中进士为止。如果条件允许,所有具有资格的考生都希望早一点参加科举考试,因为越早考中对职业生涯越有利。但越早参试,准备的时间越短,考中的概率也就越低;且一旦中了进士,则不再允许参加以后的科举,所

以每位考生只有一次中进士的机会,对应的年龄也是唯一、固定的。

因变量是考生的会试和殿试排名。原则上,我们需要分析中榜和落榜考生,由于缺乏落榜考生的资料,本文使用中榜考生的相对排名作为因变量。从理论上说,科举是对所有考生排名,落榜考生的排名一定低于最后一名中榜考生,只是为了简便起见,对落榜考生不再进行实际的排序。中榜考生的排名逻辑在适当的理论框架下可以推论到中榜与落榜的区别。当然,这其中涉及的数据删截问题无法完全消除,本文的发现应慎重推论到其他场景。

此外,本节就可获取的历史资料做了细致的分析,在参考文献以及专家意见的基础上梳理每个变量的理论和实际意义。表1汇总了所有变量及背景信息。

表1 变量和描述统计

变量	定义	观测值	平均值	标准差	最小值	最大值
乡试排名	乡试排名	12552	43.284	34.590	1	295
会试排名	会试排名	12642	153.011	96.566	1	400
殿试排名	殿试排名	12752	152.385	96.090	1	403
才能(倒序的标准化进士年龄) ^a	对考生中进士时的实际年龄做处理	12687	0.567	0.120	0	1
父亲的官职级别(1~5)	1~3品=5,4~7品=4,8~9品=3,不入流=2,无官职=1	12752	1.834	1.245	1	5
会试中的同乡关系 ^b	来自同一省份的主(副)考官数量	12752	0.223	0.444	0	2
会试中的同年关系	与考生父亲同年通过科举考试的主考官人数	12752	0.011	0.11	0	2
殿试中的同乡关系 ^c	来自同一省份的阅卷官考官的数量	12752	1.357	1.443	0	7
殿试中的同年关系	与考生父亲在同年通过科举考试的阅卷官数量	12752	0.070	0.394	0	5
教育资源(1~5)	国子监=5,府学=4,州学=3,县学=2,无信息=1	12668	3.968	1.336	1	5
家庭财富 ^d	妻妾的数量	12752	1.168	0.425	0	5
长子	考生是否家中的长子(是=1,否=0)	12752	0.367	0.482	0	1
科目						
《春秋》 ^e	五种考试科目之一	12752	0.087	0.281	0	1
《礼记》	五种考试科目之一	12752	0.077	0.266	0	1
《诗经》	五种考试科目之一	12752	0.345	0.475	0	1
《尚书》	五种考试科目之一	12752	0.237	0.425	0	1
《周易》	五种考试科目之一	12752	0.249	0.433	0	1

续表 1

变量	定义	观测值	平均值	标准差	最小值	最大值
身份						
官籍	四种户籍类型之一	12752	0.048	0.212	0	1
匠籍	四种户籍类型之一	12752	0.050	0.217	0	1
军籍	四种户籍类型之一	12752	0.291	0.452	0	1
民籍	四种户籍类型之一	12752	0.601	0.490	0	1

注：^a 才能 = $[(59 - \text{中进士时年龄}) / (59 - 13)]$ 。59 是本研究数据中年龄最大值, 13 是最小值。^b 以会试主考官的信息为基础。^c 以殿试阅卷官的信息为基础。^d 考虑到明代女性并不从事直接的生产经营活动, 无法为家庭带来直接的财富, 考生拥有的妻妾数量在很大程度上说明其家庭的富裕程度, 至少能够养得起额外的人口。其他历史学家 (Sommer, 2015) 也对此有相关说明。^e 明代科举规定会试考察五经, 即《易经》《尚书》《诗经》《春秋》和《礼记》, 并要求“士子各占一经”。考生不必精通五种经书, 只需专习一经即可。之所以可以只专一经, 是因为科举发展到明代, 对五经的注释和阐述已经颇为繁杂, 若不加区分要求掌握五经, 会增加考生学习和财务上的负担, 不利于公平、有效地选拔人才。科举在原则上对经的选择没有限制, 但在实际中, 因地区传承、历史积淀等因素影响, 某一地区的考生可能偏好某一经。因此考生对五经的选择也是考官在匿名评审中可以推论考生地区来源的一个佐证信息。

(二) 方法

本文使用为离散排序偏好设计的次序 Logit (rank-ordered logit, ROL) 方法 (Beggs et al., 1981; Hausman & Ruud, 1987) 估计各种因素对科举成绩的影响。假设在第 i 个人在第 j 次考试中的得分为:

$$Y_{i,j} = X'_{i,j}\beta + \varepsilon_{i,j}/q_j \quad (\text{式 } 1)$$

其中, $X_{i,j}$ 是一组观察到的变量, β 是系数, $\varepsilon_{i,j}$ 是误差项。假设误差项服从极值分布, q_j 代表了第 j 次考试的方差差异性, 体现了不同考试在排名准确度上的差异。设 $r_j = (r_{1,j}, \dots, r_{k_j,j})'$ 为第 j 次考试观察到的排名, 其中 K_j 为获得排名的科举考试考生总数。如果考生总数与排名数相同, 那么可以应用标准 ROL 模型得到观察到的排行榜 r_j 的似然函数。即使这两个数字在某次考试中不同, 或者在不同的考试中考生总数或排名不同, 但最大似然估计仍然适用, 因为似然函数不要求每次中榜考生人数相同。ROL 模型的另一优势是“无关选择独立性”, 即当其中某个排名缺失时不影响其余的排名 (详见 Fok et al., 2012)。基于这些优点, 用上述 ROL 模型来估计系数 β 是适用的。同时, 由于不同考试年度的考官不同, 本文用 q_j 来控制这些差异。因数据所限, 明代史料缺乏落榜考生的信息, 这会带来数据删截问题。虽然在 ROL 方法下基于考中的样本做出的估计在一定程度上可以外推, 但不可否认, 这可能会对估计产生一定的影响。本文只能尽量减小删截问题所带来的在录取边界上的断崖式差异, 这也是 ROL 方法的一个潜在优势。

在使用 ROL 方法进行的典型研究中,经验数据的形式是个体对提供的所有可用选择进行排名。科举考试的排名与 ROL 文献中数据的结构有相似之处,^①但也存在一个明显的差异:大多数使用 ROL 方法的研究都只涉及对少数几个选项的排序,在科举考试的场景下,排名的数量很大(最大值为 403),这可能给模型估计带来严重的计算负担,而 ROL 方法的优势是因为其似然函数有显式形式(explicit form),不需要通过模拟来逼近。此外,针对如此长的排名是否能准确区分考生间差异的顾虑,本文使用类似于福克等(Fok et al.,2012)的方法进行稳健性检验,没有发现系统性的问题。这也说明在巨大的资源投入下,科举的排名准确性还是较高的。为了说明本文的结论并不依赖于 ROL 模型下随机扰动分布的特殊假定,我们参考蒋勤和龚启圣(Jiang & Kung,2021)的做法,本文也以标准化考试排名为因变量,运行线性最小二乘(OLS)估计。线性模型因为把序数变量变成基数变量,可能会产生偏差,但它的优点是计算简单并且允许更为灵活的误差分布,与 ROL 模型的估算形成互补。

(三)实证结果

实证研究的结果说明了贤能主义在明代科举的实现程度。才能、社会关系和阅卷过程的匿名性都会对会试考试成员产生影响(详见表 2)。

表 2 社会关系及匿名评审对会试考试成绩的影响 N = 12351

方法(变量名称)	ROL(会试排名)			OLS(标准化成绩)		
	1	2	3	4	5	6
模型						
社会关系						
会试同乡关系	0.155 *** (0.021)	0.065 *** (0.023)	0.117 * (0.062)	3.997 *** (0.583)	1.274 ** (0.626)	1.389 ** (0.634)
会试同年关系	0.052 (0.084)	0.059 (0.084)	0.142 (0.169)	0.881 (2.330)	1.353 (2.300)	1.368 (2.303)

① ROL 的经典应用场景是若干消费者对几个产品进行排序(如式 1 所示),所有消费者 I 可能都知道所有的产品 J,每个人在排序时信息相同而只是偏好不同。这一经典的应用场景可以推广到更为现实的场景:消费者只看到自己选择的商品(并排序),但对别人的选择只有概率意义上的认知。在本文中,考官对考生的评判比较类似于这种场景。在每一次的科考中,考官是选拔者,他未必能看到其他年份的考生,但是他会有一些对考生特质和品性的偏好,然后根据考卷中或者其他信息(如是否老乡)推断出考生的特质并对考生进行评分,再根据评分进行排序。评分背后的决定因素如式 1 所刻画。当然,读者可能会关心另外一个问题,为什么这 46 次考试的考官会有某些稳定的偏好?这正是本文的贡献所在。通过结合 12752 名考生的信息和 46 次考官的信息,本文解码出历时 180 年的科考背后隐藏的决定因素(即使考官自己都没有意识到有这样的标准)。

续表 2

方法(变量名称)	ROL(会试排名)			OLS(标准化成绩)		
模型	1	2	3	4	5	6
才能	0.379*** (0.082)	0.468*** (0.083)	0.902*** (0.252)	12.93*** (2.261)	15.25*** (2.242)	14.176*** (2.296)
控制变量						
教育资源	0.019** (0.008)	0.032*** (0.008)	0.070*** (0.020)	0.511** (0.201)	0.801*** (0.201)	0.818*** (0.210)
标准化乡试排名	0.236*** (0.041)	0.359*** (0.044)	0.846*** (0.154)	8.782*** (1.118)	12.10*** (1.172)	13.152*** (1.205)
父亲官职级别 (1~5)	-0.004 (0.008)	0.011 (0.008)	0.011 (0.018)	0.107 (0.212)	0.582*** (0.212)	0.423** (0.214)
长子	0.005 (0.019)	0.036* (0.019)	0.050 (0.045)	-0.0291 (0.533)	1.116** (0.534)	1.013* (0.535)
家庭财富	-0.034 (0.022)	-0.025 (0.022)	-0.053 (0.049)	-0.512 (0.609)	-0.0510 (0.602)	-0.435 (0.606)
科目(《春秋》为参照组)						
《礼记》	-0.003 (0.045)	-0.000 (0.045)	0.032 (0.106)	-0.368 (1.262)	-0.359 (1.246)	-0.387 (1.246)
《诗经》	-0.004 (0.034)	-0.006 (0.034)	0.010 (0.080)	-0.334 (0.965)	-0.116 (0.953)	-0.424 (0.954)
《尚书》	0.069* (0.036)	0.071** (0.036)	0.168* (0.087)	1.028 (1.006)	1.182 (0.994)	1.153 (0.994)
《周易》	-0.013 (0.036)	-0.024 (0.036)	-0.026 (0.082)	-0.841 (1.001)	-0.939 (0.988)	-1.423 (0.992)
户籍(民籍为参照组)						
官籍	-0.024 (0.044)	0.032 (0.045)	0.058 (0.104)	-1.061 (1.233)	0.557 (1.238)	0.824 (1.240)
匠籍	0.079* (0.042)	0.074* (0.043)	0.119 (0.102)	1.956 (1.193)	2.082* (1.183)	2.063* (1.185)
军籍	-0.037* (0.021)	0.007 (0.021)	-0.014 (0.050)	-1.016* (0.581)	0.460 (0.590)	0.545 (0.592)
省份固定效应	否	是	是	否	是	是
年份固定效应	否	否	是	否	否	是
常数项				35.20*** (2.274)	31.97*** (2.351)	39.316*** (3.025)
组别数量	46	46	46			
对数似然率	-57850.550	-57837.132	-57244.420			

注:(1)括号内为标准差。(2)* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$ (双尾检验)。

首先,才能对会试排名起到正向作用。教育资源、乡试成绩也对会试成绩有显著的正向作用(二者在某种程度上与才能正相关)。整体上来看,会试选拔的结果符合任人唯贤的理念。

其次,社会关系依然影响了考试结果。表2中模型3显示,会试中同乡关系的系数为0.117($P < 0.1$),意味着来自与考官同一省份的考生比其他考生获得好名次的概率更高。尽管ROL模型的系数不像线性模型中的回归系数那样可以直接解读为对因变量的作用,但是通过式1可知,考试成绩仍然是这些因素的线性函数,从而可以直观地解读,也可以比较社会关系与才能等因素对评分的相对影响。在控制省份等因素后,一个同乡考官的作用可以使分数提高0.117,相当于一个标准年龄作用的13.0%,也即5.97年。对此直接的解释是,会试中一个老乡考官的作用大致相当于经历两次科举积累的才能效果。OLS模型回归的结果与ROL模型高度相似(表2中模型6)。此外,控制考试年份固定效应虽然改变了估计的系数大小,但并不影响同乡关系与才能效应的系数比值。比较模型2和模型3,同乡效应都为年龄效应的13%左右。

最后,匿名性能抑制社会关系的作用。如前所述,会试匿名而殿试不匿名,因此在控制其他变量后,同年关系和家庭背景变量不会影响会试排名。^①表3的结果表明(模型2),同年关系对殿试排名有正向关系,且效应(系数=0.097, $P < 0.05$)是同乡关系(系数=0.018, $P < 0.01$)的6倍多。

这两种关系的差异不仅体现在相对效应上,还有助于进一步厘清这两种关系对科举成绩的因果效应。会试和殿试相隔一个月,对于平均年龄为32岁且受教育时间超过20年的进士来说,其才能不太可能在短时间内发生显著变化。而这期间的一个关键因素是会试匿名而殿试不匿名,并且会试因地区配额制而披露考生的地区来源信息,这使得会试中的同乡关系可能产生系统性的偏袒,而同年关系却不起作用。作为对比,因殿试能够定位个人信息,在同乡关系效应也存在的前提下,同年关系的重要性更加突出。

① 同年的一个先决条件是父亲必须是进士,因此人们可能会担心同年效应是父亲教育程度的一个副产品。对此问题可以作如下回应:首先,虽然回归中没有纳入父亲是否进士这一变量,但是考虑到明代的官员大部分由进士担任,尤其是中后期,考取进士成为进入官僚体系的唯一途径,所以控制父亲的官职也大致上相当于控制了父亲的科举功名。在回归中同时加入父亲官职和进士这两个条件后,父亲是否进士不管在会试还是在殿试中作用都不显著,限于篇幅,在正文中没有报告。其次,即便父亲是进士,他与考官存在同年关系也是小概率事件。所以同年的作用更多是社会关系而非父亲本身的教育程度在起作用,尤其是在控制了父亲的官职之后。感谢匿名审稿人指出这个同年的先决性问题。

表 3 社会关系及匿名评审对殿试考试成绩的影响 N = 12422

方法(变量)	ROL(殿试排名)			OLS(标准化成绩)		
模型	1	2	3	4	5	6
社会关系						
殿试同乡关系	0.018 ** (0.009)	0.018 ** (0.009)		0.663 *** (0.207)	0.597 *** (0.211)	
殿试同年关系	0.089 ** (0.040)	0.097 ** (0.046)		1.850 *** (0.680)	1.598 ** (0.739)	
会试同乡关系		0.022 (0.024)	0.034 (0.025)		0.993 (0.649)	1.397 ** (0.636)
会试同年关系		-0.028 (0.092)	0.119 (0.082)		2.045 (2.503)	4.177 * (2.304)
才能	0.327 ** (0.136)	0.353 ** (0.144)	0.346 ** (0.140)	15.557 *** (2.301)	15.596 *** (2.303)	15.627 *** (2.304)
控制变量						
教育资源	0.021 ** (0.009)	0.022 ** (0.009)	0.020 ** (0.009)	0.973 *** (0.211)	0.979 *** (0.211)	0.967 *** (0.211)
标准化乡试排名	0.285 *** (0.105)	0.301 *** (0.108)	0.303 *** (0.107)	11.351 *** (1.208)	11.410 *** (1.209)	11.487 *** (1.209)
父亲官职级别 (1~5)	0.038 ** (0.015)	0.040 ** (0.016)	0.045 *** (0.017)	1.637 *** (0.223)	1.631 *** (0.223)	1.764 *** (0.215)
长子	0.036 * (0.020)	0.039 * (0.021)	0.039 * (0.021)	1.294 ** (0.536)	1.290 ** (0.536)	1.300 ** (0.537)
家庭财富	-0.051 ** (0.025)	-0.053 ** (0.026)	-0.051 ** (0.025)	-1.688 *** (0.608)	-1.681 *** (0.608)	-1.671 *** (0.608)
科目(《春秋》为参照组)						
《礼记》	-0.076 (0.046)	-0.079 (0.048)	-0.079 * (0.047)	-0.977 (1.250)	-0.955 (1.250)	-0.988 (1.250)
《诗经》	-0.023 (0.030)	-0.023 (0.031)	-0.026 (0.030)	-0.131 (0.958)	-0.129 (0.958)	-0.167 (0.958)
《尚书》	-0.004 (0.030)	-0.003 (0.032)	-0.002 (0.031)	1.573 (0.997)	1.595 (0.997)	1.557 (0.997)
《周易》	-0.033 (0.031)	-0.032 (0.033)	-0.034 (0.032)	-0.142 (0.995)	-0.127 (0.995)	-0.172 (0.995)
户籍(民籍为参照组)						
官籍	0.030 (0.039)	0.034 (0.042)	0.030 (0.040)	1.400 (1.246)	1.399 (1.246)	1.334 (1.246)
匠籍	0.067 (0.042)	0.069 (0.043)	0.072 * (0.043)	0.708 (1.190)	0.697 (1.190)	0.707 (1.191)
军籍	0.018 (0.018)	0.019 (0.019)	0.016 (0.018)	0.766 (0.593)	0.769 (0.593)	0.760 (0.593)

续表 3

方法(变量)	ROL(殿试排名)			OLS(标准化成绩)		
	1	2	3	4	5	6
省份固定效应	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
常数项				31.387*** (3.041)	30.931*** (3.059)	31.668*** (3.040)
组别数量	46	46	46			
对数似然率	-57800.883	-57837.132	-57809.660			

注:(1)括号内为标准差。(2)* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$ (双尾检验)。

匿名性对公平的保障还可以从一些控制变量的效应中窥见一斑。例如,父亲的官职级别对会试排名没有显著影响,^①但对殿试排名有显著的正向作用。虽然殿试不采取淘汰制,但其排名决定了谁能进入官僚体系的快速通道(如进入翰林院),这个效应足以颠覆一个考生通过多年努力而带来的才能上的优势,从长远来看不利于社会流动。虽然父亲的官职级别既可解释为社会资本,也可被理解为某种默会知识或文化资本,但回归结果表明默会知识的传递可能不是父亲官职级别发生作用的机制,而更可能的原因是社会资本,即殿试考官认识考生的父亲。^②类似地,以妻妾数量衡量的家庭财富指标不影响会试排名,但对殿试排名有负向影响。^③其余控制变量对会试及殿试排名的影响趋势一致。从省份来看,表现最好的是浙江,其次是江西。如果不控制省份的固定效应,同乡关系的系数要高出一倍(参见表2中模型1和模型2的比较)。另外,各模型中乡试、会试排名对后续考试排名都有正面影响,说明科举能在一定程度上逐层选拔人才。

综上,明代科举体现了本文提出的贤能主义的制度悖论。由于采取相对统

- ① 我们根据史实,也尝试对父亲的任职进行其他分类。一种分类是以是否任官为哑变量作简单分类。另一种分类是建构一个连续变量,取值1~20(明代官职品秩为九,每一品秩又分正职和副职,同时还有不在品秩上的公职人员,以及从未有官职的;我们给正一品赋值为20,副九品为3,不入流为2,从未在系统内任职的为1),考察不同官职品秩的影响。不同的分类并未改变分析结果。
- ② 蒋勤和龔启圣(Jiang & Kung, 2021)用清代的科举数据剖析了父亲官职及科举成就对儿子考试结果的影响。他们未提及匿名性问题,但强调了殿试考察内容有所不同,因此默会知识更重要。默会知识固然有一定的作用,但一般而言,平均年龄32岁的进士在漫长的备考过程中对殿试考察内容、预期的风格以及历届考题及范文都有相当深刻的理解。相对而言,通过父辈个人经历所传递的默会知识不是唯一解锁殿试备考的途径。
- ③ 一个可能的解释是与中国历史上的抑商传统有关。但限于资料的可获取性,对此我们无法展开更严谨的讨论。

一的选拔标准且在会试阶段实施了严格的匿名性,科举的确能在一定程度上实现任人唯贤的选拔作用。但基于地区平衡考虑的配额制以及在殿试环节的非匿名程序(类似于面试)也为不同的社会关系发挥影响开启了方便之门。这样复杂的制度设计或许折射了更深层次的国家治理逻辑,即通过会试广纳贤才,但通过殿试进一步选拔符合皇权利益的考生,以反哺政治忠诚,系统性地提高来自官僚家庭考生的排名,为他们提供进入官场的快速通道。

五、结论与启示

若以促进社会流动、以考试呈现的才能作为人才选拔标准等方面来评价,科举是相当成功的举措。但如前所述,在制度层面,科举在公平性、匿名性与有效性之间存在悖论。在明代,为了保证有效和公平,决策者在匿名性上做了很多努力,但这突显了既有的地区间不平等;为此实施的地区配额导致地区信息披露,也意味着在匿名性上让步,继而为老乡关系发挥作用提供了空间,最终降低了有效性和公平性。类似的悖论在大范围选拔制度的顶层设计中难以避免。明代科举提供了一个真实而丰富的制度案例,它的初衷与实际结果仿佛“百密”与“一疏”,在某种程度上反映了人才选拔制度在复杂社会系统中所面临的约束。

以史为鉴,对科举制度的剖析也对当今的人才选拔制度具有启示意义。以高考为例,一个普遍观点是应试教育过于强调重复做题,忽视了对创造性的培养。但如果在选拔中加入过多素质教育的内容,则来自弱势家庭的孩子机会更少。目前素质教育的考核方式尚不明确、透明,如何在选拔过程中保证公平也是一个难题(杨东平,2004)。2004年,《新闻调查》的一期节目《命运的琴弦》就揭露了艺术招生中的腐败问题。如果赋予学校更大的自由裁量权,实施和监督成本可能会更高,实际选拔结果不一定更好(吴晓刚、李忠路,2017)。

要评价任人唯贤制度的实现程度,就需要将其放置到社会系统中来审视。任人唯贤仍然是值得追求的制度,但对于在何种场合需要选拔制度、以什么样的方式选拔、对竞争程度的干预需要一个综合评估。选拔的成本不仅包括实施过程中所需要的有形资源,还包括因选拔而产生的行为扭曲及其社会代价。如费西金(2015)所言,如果一个社会过于依赖某次大考来衡量个人的才能并分配机会,那完全的公平是不可能实现的。一个良性运行的社会应该为其成员充分发挥个体潜能创造机会,即“三百六十行,行行出状元”。在一个多元社会,如果选

拔制度存在内在悖论,就应有事后补救的措施。反之,如果过于重视一次选拔结果,最终会使其成为不能承受之重,进退失据,宽严皆误。

本文剖析了人才选拔制度本身的悖论,提出三个目标最多只能取其二。针对此困境,本文无法给出一个标准答案,解决问题的关键在于建立全局系统观,审时度势,根据主要矛盾在三个制度目标之间做出权衡。如在选拔制度的不同节点,可以选择性地侧重于某个制度目标,进而使得整个选拔体系能相互呼应,尽可能降低三个制度目标无法全部实现所带来的社会代价。

参考文献:

- 方勇译注,2007,《墨子》,北京:中华书局。
- 费西金,约瑟夫,2015,《瓶颈:新的机会平等理论》,徐曦白译,北京:社会科学文献出版社。
- 郭培贵,2014,《明代学校科举与任官制度研究》,北京:中国大百科全书出版社。
- 郭振有,2005,《应试教育之弊端:反科学发展观》,《中国教育学刊》第9期。
- 何炳棣,1966,《中国会馆史论》,台北:台湾学生书局有限公司。
- 卡拉贝尔,杰罗姆,2014,《被选中的:哈佛、耶鲁和普林斯顿的入学标准秘史》,谢爱磊、周晟、柳琳译,北京:中国人民大学出版社。
- 可汗,西莫斯,2016,《圣保罗中学精英教育的幕后》,蔡寒韞译,上海:华东师范大学出版社。
- 康文林,2020,《清末科举停废对士人文官群体的影响——基于微观大数据的宏观新视角》,《社会科学辑刊》第4期。
- 李春玲,2014,《“80后”的教育经历与机会不平等——兼评〈无声的革命〉》,《中国社会科学》第4期。
- 梁晨、董浩、任韵竹、李中清,2017,《江山代有才人出——中国教育精英的来源与转变(1865-2014)》,《社会学研究》第3期。
- 刘精明,2014,《能力与出身:高等教育入学机会分配的机制分析》,《中国社会科学》第8期。
- 田方萌,2017,《配额制的功能与代价——以科举及其他制度为例》,王天夫主编《清华社会学评论》第8辑,北京:社会科学文献出版社。
- 吴晓刚,2016,《中国当代的高等教育、精英形成与社会分层——来自“首都大学生成长追踪调查”的初步发现》,《社会》第3期。
- 吴晓刚、李忠路,2017,《中国高等教育中的自主招生与人才选拔:来自北大、清华和人大的发现》,《社会》第5期。
- 谢爱磊、洪岩壁、匡欢、白杰瑞,2018,《“寒门贵子”:文化资本匮乏与精英场域适应——基于“985”高校农村籍大学生的追踪研究》,《北京大学教育评论》第4期。
- 杨东平,2004,《教育公平是一个独立的发展目标——辨析教育的公平与效率》,《教育研究》第7期。
- 张燕婴译注,2006,《论语》,北京:中华书局。
- Aslund, O. & O. N. Skans 2012, “Do Anonymous Job Application Procedures Level the Playing Field?” *ILR Review* 65(1).
- Beggs, S., S. Cardell & J. Hausman 1981, “Assessing the Potential Demand for Electric Cars.” *Journal of*

Econometrics 17(1).

- Brook, T. 2010, *The Troubled Empire: China in the Yuan and Ming Dynasties* (Vol. 5). Cambridge: Harvard University Press.
- Chaffee, J. W. 1985, *The Thorny Gates of Learning in Sung China: A Social History of Examinations*. Cambridge: Cambridge University Press.
- Crosby, F. J., A. Iyer & S. Sincharoen 2006, "Understanding Affirmative Action." *Annual Review of Psychology* 57(1).
- Daniels, N. 1978, "Merit and Meritocracy." *Philosophy & Public Affairs* 7(3).
- DiMaggio, P. 1988, "Interest and Agency in Institutional Theory." In L. G. Zucker (ed.), *Research on Institutional Patterns: Environment and Culture*. Pensacola: Ballinger Publishing Co.
- DiMaggio, P. J. & W. W. Powell 1983, "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality." *American Sociological Review* 48(2).
- Elman, B. A. 2000, *A Cultural History of Civil Examinations in Late Imperial China*. Berkeley & Los Angeles: University of California Press.
- 2013, *Civil Examinations and Meritocracy in Late Imperial China*. Cambridge: Harvard University Press.
- Estrada, R. 2019, "Rules versus Discretion in Public Service: Teacher Hiring in Mexico." *Journal of Labor Economics* 37(2).
- Fok, D., R. Paap & B. Van Dijk 2012, "A Rank-ordered Logit Model with Unobserved Heterogeneity in Ranking Capability." *Journal Applied Econometrics* 27(5).
- Golden, D. 2007, *The Price of Admission: How America's Ruling Class Buys Its Way into Elite Colleges—and Who Gets Left Outside the Gates*. New York: Penguin Random House LLC.
- Goldin, C. & C. Rouse 2000, "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90(4).
- Hallett, T. & M. J. Ventresca 2006, "Inhabited Institutions: Social Interactions and Organizational Forms in Gouldner's Patterns of Industrial Bureaucracy." *Theory and Society* 35(2).
- Hartwell, R. M. 1982, "Demographic, Political, and Social Transformations of China, 750 – 1550." *Harvard Journal of Asiatic Studies* 42(2).
- Harvard University, Academia Sinica & Peking University 2018, *China Biographical Database (CBDB)* (<https://projects.iq.harvard.edu/cbdb>).
- Hausman, J. A. & P. A. Ruud 1987, "Specifying and Testing Econometric Model for Rank-ordered Data." *Journal of Econometrics* 34(1–2).
- Ho, P. T. 1964, *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368 – 1911*. New York: John Wiley & Sons, Inc.
- Hoffman, M., L. B. Kahn & D. Li 2018, "Discretion in Hiring." *The Quarterly Journal of Economics* 133(2).
- Holmstrom, B. 1979, "Moral Hazard and Observability." *The Bell Journal of Economics* 10(1).
- Hoxby, C. M. & C. Avery 2013, "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low-Income Students." *Brookings Papers on Economic Activity* (Spring).
- Jiang, Q. & J. K. S. Kung 2021, "Social Mobility in Late Imperial China: Reconsidering the 'Ladder of

- Success' Hypothesis." *Modern China* 47(5).
- Kracke, E. A. 1953, *Civil Service in Early Sung China, 960 - 1067 : With Particular Emphasis on the Development of Controlled Sponsorship to Foster Administrative Responsibility*. Cambridge: Harvard University Press.
- Leslie, L. M., D. M. Mayer & D. A. Kravitz 2014, "The Stigma of Affirmative Action: A Stereotyping-Based Theory and Meta-Analytic Test of the Consequences for Performance." *Academy of Management Journal* 57 (4).
- Markovits, D. 2019, *The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite*. London: Penguin Books Limited.
- Meyer, J. W. & B. Rowan 1977, "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *The American Journal of Sociology* 83(2).
- Needham, J. & C. Harbsmeier 1998, *Science and Civilisation in China : Volume 7, The Social Background Part 1, Language and Logic in Traditional China*. Cambridge: Cambridge University Press.
- Opper, S., N. Victor & B. Stefan 2015, "Homophily in the Career Mobility of China's Political Elite." *Social Science Research* 54.
- Powell, W. W. & J. A. Colyvas 2008, "Microfoundations of Institutional Theory." In R. Greenwood, C. Oliver, R. Suddaby & K. Sahlin (eds.), *The Sage Handbook of Organizational Institutionalism*. London: SAGE.
- Rivera, L. A. 2015, *Pedigree: How Elite Students Get Elite Jobs*. Princeton & Oxford: Princeton University Press.
- Sandel, M. J. 2020, *The Tyranny of Merit : What's Become of the Common Good?* London: Penguin UK.
- Sommer, M. 2015, *Polyandry and Wife-Selling in Qing Dynasty China : Survival Strategies and Judicial Interventions*. Berkeley: University of California Press.
- Scully, M. 2014, *Meritocracy*. Hoboken, New Jersey: Wiley Encyclopedia of Management.
- Taylor, K. 2021, "Parents Paid to Open College Doors. Now They're Spending to Limit Prison Time." *New York Times* (<https://www.nytimes.com/2019/10/03/us/college-admissions-scandal-consultants.html>).
- Young, M. 1958, *The Rise of the Meritocracy, 1870 - 2033 : An Essay on Education and Equality*. Harmondsworth: Penguin Books.
- 2001, "Down with Meritocracy: The Man Who Coined the Word Four Decades Ago Wishes Tony Blair Would Step Using It." (<http://www.theguardian.com/politics/2001/jun/29/comment>)
- Zwart, F. D. 2005, "The Dilemma of Recognition: Administrative Categories and Cultural Diversity." *Theory and Society* 34(2).

作者单位:浙江大学民营经济研究中心、浙江大学经济学院(柯荣住)
 清华大学社会学系、清华大学科技与社会研究中心(洪伟)
 北京大学国家发展研究院、剑桥大学商业研究中心(郑恩营)
 责任编辑:张志敏