

农业、食品和营养	新计划:ESSP 全球环境变化和食品系统 被提议的研究计划:健康以及可循环食物链
生物多样性	在研计划:DIVERSITAS
横向科学问题	
数据、信息和数字鸿沟	相关实践:有关数据和信息的优先领域评估 在研计划:CODATA、FACS、WDCs、INASP 新计划:参与 GEO(地球观测集团)
能力建设和基础科学投资	相关实践:优先领域评估能力建设
科学、社会和伦理	新计划:对科学和社会所扮演的角色及其责任的 战略评论

(柯春晓译)

世界范围内的大学科研评价方法(上)

本文在世界范围内考查了对大学科研产出的数量和质量进行评估工作的增长趋势。评估在高校科研绩效考核中扮演着一个相当有用的角色并对研究所之间不断增强的竞争所引起的科研产出的增加和质量的提高起到了激励作用。本文还讨论了应用于各国高校科研产出的评估方法。

一、导言

不断增加的科研产出被普遍认为是对社会大有益处的。由于有能够助长经济增长的潜力,世界上各国政府都加强了对高校科研资源的投入力度。但追加的资源投入是否就能够增加科研产出的能力取决于整个科研系统的生产力水平。而增强科研生产力的关键一步是要衡量大学科研产出的质量和数量。提高科学研究生产力的重要性也许正好解释了为什么高校科研产出的评估近 20 年来在世界范围内呈不断增强的趋势。

本文的目的就在于提供一些评估中会应用到的基本原则,并以一种规范的方式展现评估是怎样进行的。尽管本文的主要目的是考查评估系统,但本文把讨论置于一个更加宽泛的、科学家与社会之间的相互作用不断加强的环境中,我们将在导言中进行简要讨论。

传统的“科学”观认为科学是一种“建立在研究个体非凡的认知基础上”的活动。物理学家 John Ziman 将这种观点称作“传奇”,即把科学认为是冷静的人类在充分理性和无私奉献基础上产出的成果的神话观点。另一种传统认为科学家们会为了扩展研究领域的目标而人为地设定科学的发展方向。这里

当然包括科学界开放的言论,但科学本身从根本上说是一个闭合系统。这种“背离”的观点正在改变。

科学发生在社会环境中。高校、政府、企业或广义上的社会构成了社会环境,这些机构中的变化都能够推动科学组织的改变。本文没有考查科学家个体的行为,而是关注了这些机构的体制结构。

社会和科学的相互作用是双向的。不仅机构的决策会影响到科学,科学本身也从包括资源在内的各个方面更加深远地影响着社会和各个机构。科学中纯粹的信任正被侵蚀。Anthony Griddens 在 BBC 的演讲中说道:

“如今,我们(人类)同科学技术的关系的已不同于早期的特点……同过去相比,绝大多数人(包括政府当局和政治家们)都已经或将不得不与科学技术结成一种更加活跃和密切的关系。我们不能仅仅因为科学家之间很少达成共识,就简单地全盘接受科学家的研究成果……目前,并没有能让我们监测国内外技术变化的机构。如果事先建立了关于技术变革及其未知后果的公众对话机制,在英国和其他地方突发的疯牛病灾难也许就能避免。虽然与掩饰真相相比,让公众更多地介入科学与技术的方法,会使科学与技术陷入公众谣言的两难处境,但却能在某种程度上减少它们更严重的破坏性后果。”

伴随科学对社会的影响,闭合的科学系统变得无法维持。

科学与社会日益频繁的交互已经导致有研究提出了在科学与知识生产模式上存在变化的观点。表 1 把这种变化描述成为一种从模式 1 到模式 2 的运动。这种变化使得新的科学知识的定义及其可信度的建立遭到了根本性的挑战。这些新问题对于传统的科学方法提出了挑战,因为在闭合的科学世界中,传统的方法仅仅满足了科学家。科学家们现在必须说服社会,他们是在生产丰富的社会知识。

表 1 知识生产模式

模式 1	模式 2
确定问题并在学术界的研究共同体中解决	确定问题并在实际应用中解决
单一学科的	跨学科的
同质的	异质的
分级的,倾向于保留现有的组织形式	变化分层的,包括各种历史的组织形式
内部质量控制	更加能够为社会负责的质量控制

资源来源: pp. 12, Arnold et al. (2001)

对科学越发严峻的挑战也导致了科学激励制度的挑战:

1. 有效分配稀有资源

不断地改进分配机制,可以支持科学研究活动按照科学发展、大学或政府确定的优先次序来进行。这对于分配资源的决策者来说是一个核心的目标。

2. 向更广阔的社会延伸

可以向更多的公众提供证据,说明社会和经济的发展得益于对科学研究的支持。

从某种意义上来说,可以把评估的引入看作迎接这些挑战的第一步。因为评估本身所提供的信息可以提供帮助。而如果评估过程是建立在与学术界合作的基础上,则会加强其合理性,这些过程将会逐步加强学术界和社会的交流。

评估可从两方面实现上述第一个目标:(1)激励研究机构,强调其日趋重要的研究使命;(2)至少提供一些应用信息以引导更多的研究产出。

评估给与了研究机构进行学习的机会,同时也给了它们证明研究系统能够成功提高科研产出的机会。因此,这些评估应该更多地考虑项目执行责任方面的导向(以利于项目进行资源的再分配、责任的再分配并不断修正其目标)而非仅仅瞄准学术本身。

严格来说,决策者们面临的挑战在于:联邦政府需要衡量其 R&D 投入是否有效。我们可以只展现 R&D 投入的产出,而不与我们对经费进行决策的绩效信息相联系。没有这些信息,项目的决策就成了无源之水,只是想当然、依据上年经费水平和当地投资团体的政治目的。

根据上述第二个目标,评估可以影响到外界的反应。随着对国库投入需求的不断增加,高校很可能需要逐步利用法律来保障科研经费的投入。高校有责任建立其社会资产。加拿大高校联合会在 1995 年指出:

在财政紧缩之际,应强调“金钱的价值”,或者说保障经费投向了公共教育才是做到了物尽其善。随着社会各部门不断强调对经费投入去向的解释,各界对于科研绩效及其产出的度量越来越感兴趣了。高校已经意识到了必须接受这种财政情况的事实,并正在着力寻找支持决策的科研投入效率、效力和产量等方面新的信息。

对于高校来说,公布更加易于公众理解和更加有意义的产出衡量标准,将有利于其解释自身科学研究的效率和效果。而当缺少了这种保障时,OECD 认为,对于基础研究会带来可预见的长期利益的主张则会因为看不到利益的出现,而使得公众对这种期待大失所望。例如,澳大利亚科学院在 2003 年指出,在英国,皇家航空研究中心在提高科研产出质量的同时,大大提高了其优

秀的高科技企业在公众和政府中的知名度。

本文的逻辑思路如下:假设政府对于科研产出的效率和效果的提高感兴趣,本文首先考查了各国的产出是否有差异,并找出了答案。在用于解释科研绩效的国际差异的众多因素中,经费的影响是最显著的。本文重点探讨了评估系统的国际差异。很多科研绩效很高的国家已经或者即将建立针对研究产出的评估系统。

本文接下来讨论了科研评估中应用的方法及其优缺点。然后介绍了英国的经验,因为有证据证明英国的研究系统相当高产。世界各国采用了各种评估方法,文章也阐述了这些评估方法实施的发展历程,主要讨论了荷兰、丹麦和澳大利亚三国的情况。介绍评估系统实际上是一个学习的过程,而且由于其原创性,评估系统在所有的国家都有所演变。其他很多国家都在建立或发展科学研究评估系统。

文章的结论很真切地反映在一份加拿大国会的工业、科学与技术常务委员会的报告中提出的两条相关建议:

- 加拿大政府要求资助单位就自己的项目和执行情况进行更加常规的內部评议(包括同行评议)并定期考察其他加拿大和国外相关单位的决策程序,以保障科研经费得到最佳的分配。并且应该让研究界和大众能够便捷地获得这些内部评估的结果。

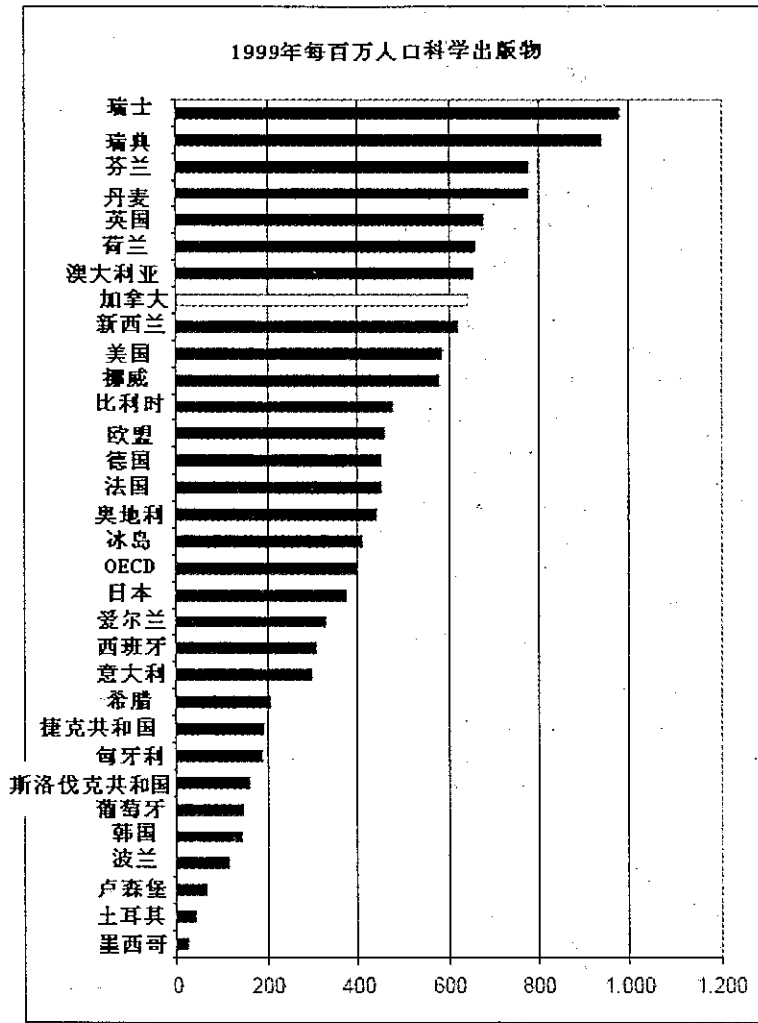
- 加拿大政府保证联邦资助单位会采取措施更好地衡量和汇报科研产出,而且只要有可能,还会包括对大众有利的研究项目的成果。

二、世界各国的研究产出差异如何

本节将给出世界范围内研究成果的一些背景。由于不存在单一的比较科研产出的方法,因此本节将考查一系列因为来源不同而有所变化的指标。因为国家的规模可能对总体研究产出有较大影响,所以本节将着眼于对“生产力”的测度:每单位资源的研究产量。图1粗浅地描述了一些标志 OECD 国家研究机构效率的固定因素。欧洲中部和北部的一些小国在单位人口的出版量方面是研究强国。列入这一个组的还有澳大利亚和英国,尽管是大国但却一直保持着高产的科研能力。加拿大也相当出色,在 OECD 国家中列第八位。

把科学论文总数作为科研产出的测度有一个问题,在于它不能控制(论文的)质量差异。一般认为高被引的论文质量好一些,引文率常被用于论文质量的控制。拥有高被引的文章比率的国家,加上美国,都生产高质量的论文。较大的欧洲国家,除英国以外表现都差强人意。这些数据都有一个特点,即无论出版物产量的变化如何,法国和德国的表现都不太好。对此有很多可能的解释:不愿用英语出版的倾向、高校在科研活动中舞台太小、法国和德国学术研

究者同工业界的接触比较多。



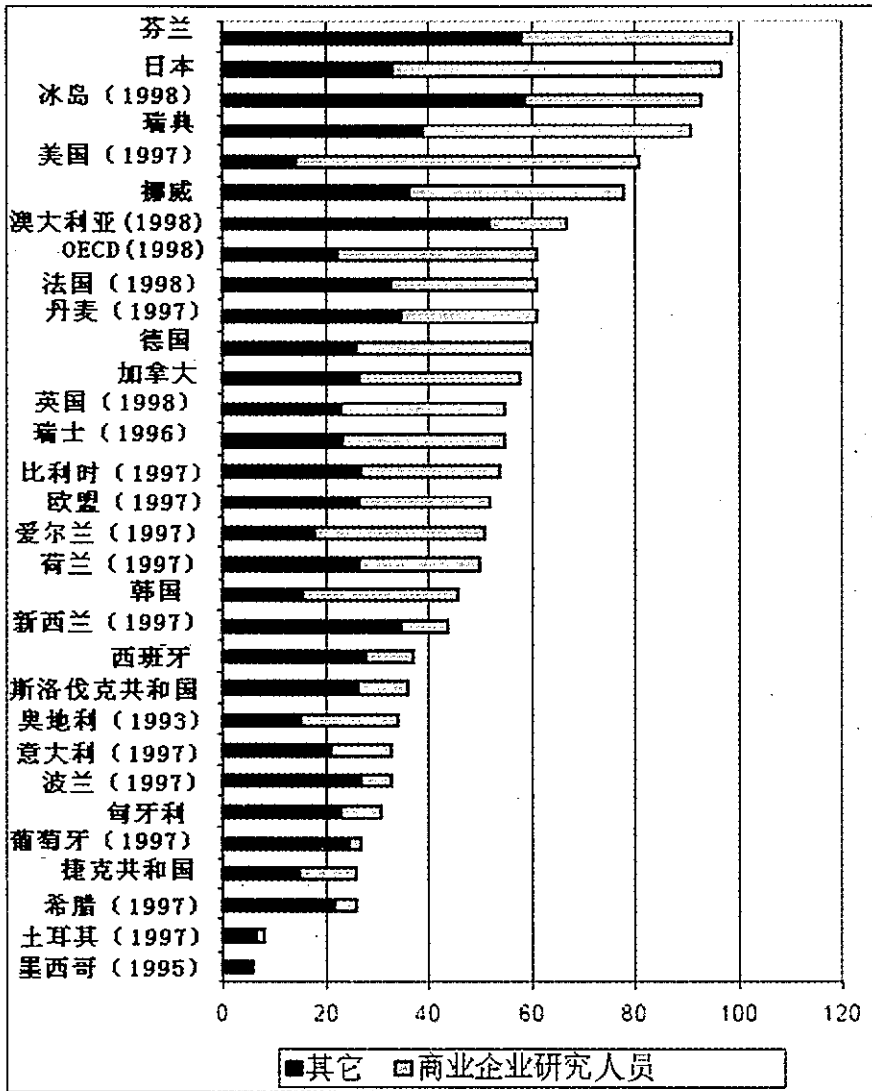
资料来源: OECD (2001a)

图1 1999年每百万人口科学出版物

另一个科研产出的测度标准见图2。图中显示了科研场所中研究人员比例。研究人员在科研过程中属于投入要素,但他们同样是研究型大学的一种“产出”,北欧国家的该指标同样有绝对优势。

简而言之,表2列出了在科研产出的不同排行上名列前10名的国家。从第一列中可以看到较大的国家相应地科研产出也较高。科研产出的排名和GDP的排名几乎是对应的。例外的国家有:英国,它的GDP排名是第四,但科研总产出是第三,而澳大利亚和荷兰1999年的GDP也分别排名11和12。说到科研的强度、生产力和质量,从表2中的第二栏可以看到很多小型的开放经

济体在进行高质量的研究时比大国要更加高效。



来源:OECD(2001b)

图2 每万名劳动力中研究人员数量

表2 按不同指标的各国研究产出的排名

	1999年OECD科学出版物的国家分布*	1999年人均科学出版物	1999年高校研究人员的人均科学出版物*	1999年研究人员人均科学出版物*	至2000年止按前1%的论文的篇均引文数的排名
1	美国	瑞士	美国	瑞士*	瑞士
2	日本	瑞典	荷兰	新西兰	美国
3	英国	芬兰	英国*	意大利	荷兰

4	德国	丹麦	瑞士*	荷兰	丹麦
5	法国	英国	丹麦	英国	瑞典
6	加拿大	荷兰	意大利	丹麦	英国
7	意大利	澳大利亚	奥地利*	加拿大	加拿大
8	澳大利亚	加拿大	捷克	瑞典	芬兰
9	西班牙	新西兰	加拿大	西班牙	比利时
10	荷兰	美国	瑞典	澳大利亚*	德国

来源:OECD(2001)Evidence(2003)。*为1998年的数据。°为2000年的数据。
 *科学出版物似乎更多地来自于高校研究人员,但一些政府和大型公司也进行科学研究。由于美国有大量私有部门的研究者,其研究人员人均的科学出版物排名低至21名。

图3考查了一段时间以来科学出版物的相关数据。初见端倪的一个趋势是,欧洲和亚洲一些地区作为科学出版物源泉的重要性不断加强。中国台湾的出版物在1986年到2000年之间增长了6倍,而韩国增长了10倍之多。美国呈现出相对下降趋势,但是其基数比其他国家大很多。加拿大同样呈明显的下降趋势,而英国自从1989年以来基本处于稳定。

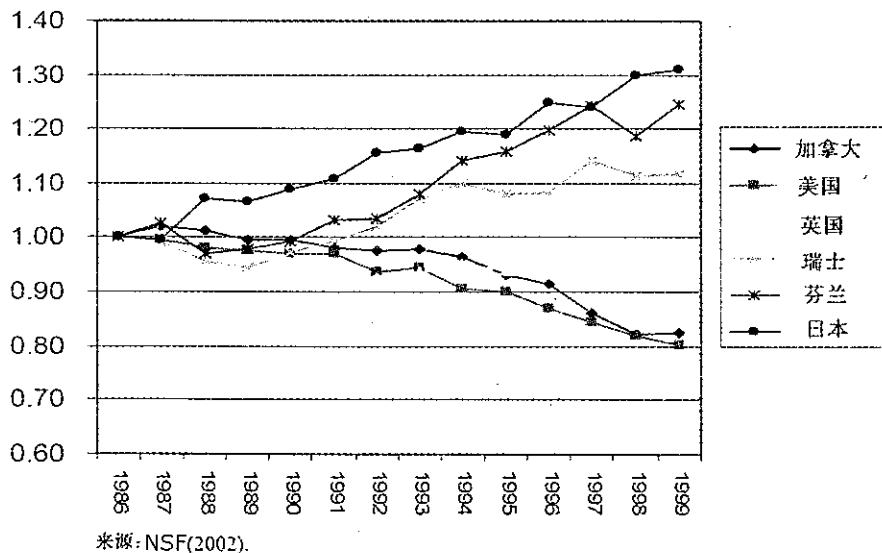
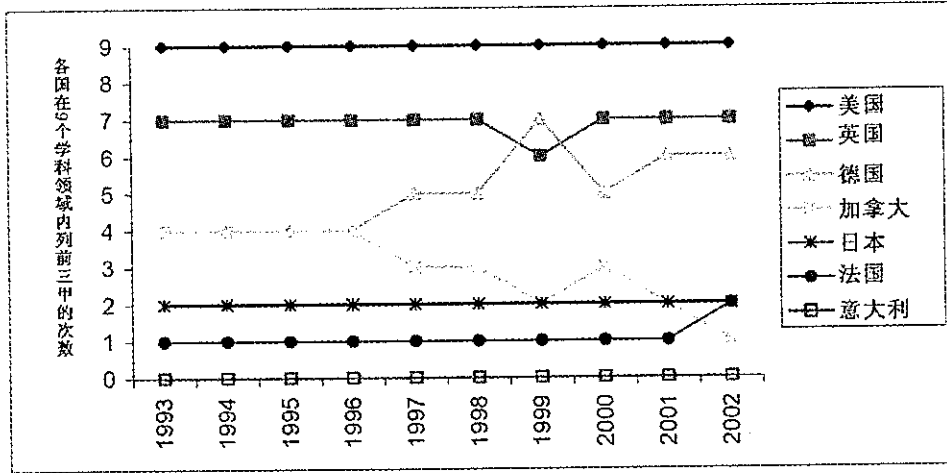


图3 1986~1999年间科学出版物分布趋势图

图4显示了1993年到2002年之间这些国家是否名列9个研究领域引文数量的前三位。通常美国在这9个领域中都能列入三甲,而英国在其中7个

领域仅有一年没有进入前三。由于一些不明原因,加拿大在4个领域内高质量的研究水平开始下降,只有2002年在其中1个领域进入了前三名。



来源:Thomson ISI, reported in Evidence(2003)

图4 9个研究领域内引文数居前列的国家

表3说明的是7个领域里每个领域中前五强的国家的排名和加拿大在其中的位置。很明显,美国和英国是前五强国家中长期以来仅有的大国。很多小国家表现得很出众也很稳定。

表3 国家影响力排名,1998~2002

	工程、计算机科学和技术	物理、化学,地球科学	农业、生物,环境科学	生命科学	临床医学	社会与行为科学	艺术与人文
1	瑞士	美国	荷兰	美国	美国	美国	荷兰
2	美国	瑞士	丹麦	瑞士	荷兰	荷兰	美国
3	以色列	丹麦	英国	荷兰	芬兰	英国	英国
4	丹麦	荷兰	美国	英国	瑞士	加拿大	芬兰
5	荷兰	德国	瑞典	芬兰	加拿大	芬兰	以色列
加拿大的排名	11	8	9	6	5	4	6

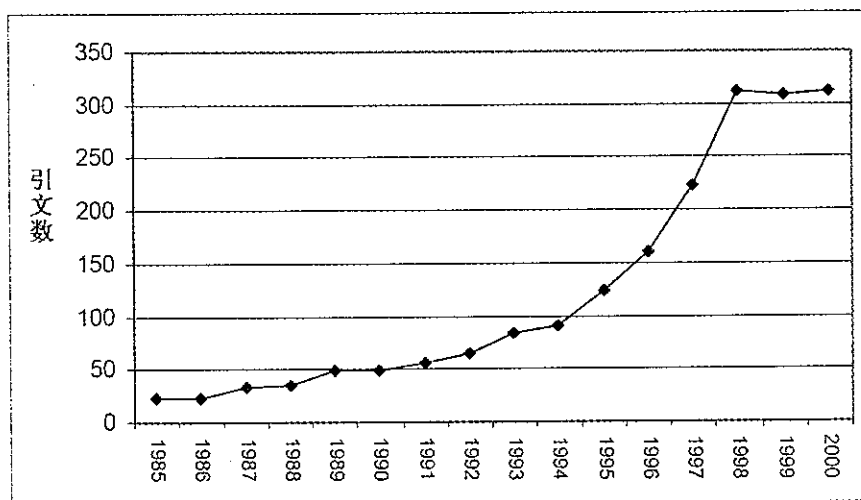
来源:科学计量学记分板,CEST,瑞士

通过前述的讨论,加拿大科研活动给我们的总体印象还是很好的。然而引文数显示加拿大在国际水平上的研究产出已大不如前了。芬兰、瑞典、瑞士和英国的科研绩效说明它们还可以获得更高的研究产出。

三、高校研究是否重要

提出这个问题的根据在于社会是否应该关注学术出版物的质量或数量,甚至于科学的通俗性。

图5展示的是以出版物形式体现的科学发展的潜在重要性,说明科学对于开发新的商品越来越重要了。图6表明科学论文的参考文献,而非专利,在20世纪90年代大量地增加了。



来源: NSF (2002)。引文包括学术论文的所有参考文献。引文数以12年为基数还有3年的尾数,例如,2000年的引文是指美国1986~1997年间发表的论文在2000年申请的专利中被引用的次数。美国专利与商标办公室改革后的程序,简化了的学术论文查询方以及加强引文引用的激励,都有助于提高论文的引文数。

图5 美国专利中的科学技术论文引文数

图6的左半部分展示的是高校作为专利的源泉变得越来越重要。然而,高校专利的重要性或意义平均而言正在降低,说明申请专利的回报率正在减少。以上粗浅的分析表明政策应该保护并加强高校在进行高质量基础研究当中的核心地位。

应该记住,广义而言,全球的2%的研究机构生产了世界69%的出版物和79%的引文。

四、政策问题:我们是否可以向他国学习

即便我们接受“科学”对于未来的发展有重要作用的观点,我们仍需要提出以下问题:

- 国家应对 R&D 活动投入多少经费?
- 从基础研究到科技创新扩散的过程中,应该把经费投入到哪一个环节最合适?
- 把钱投入到某些部门、分支机构或团体能够比投到其他部门有更好的回报吗?

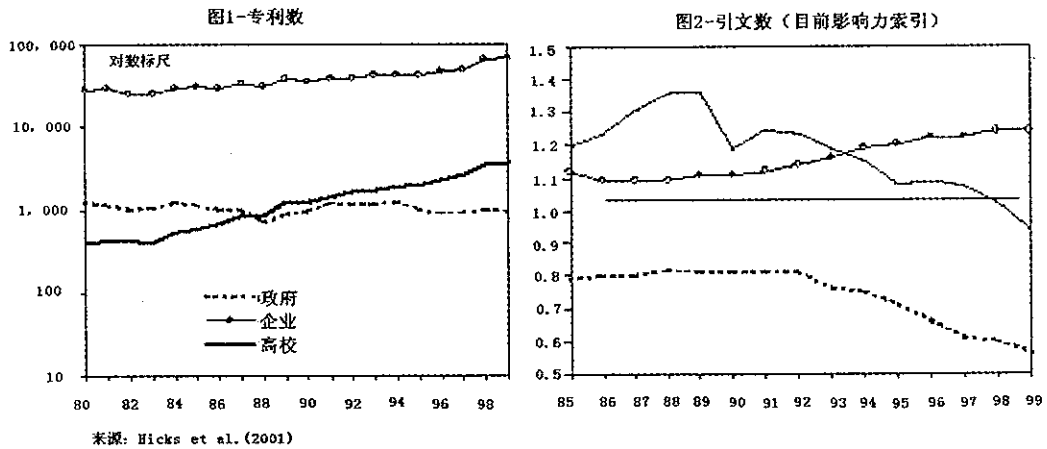
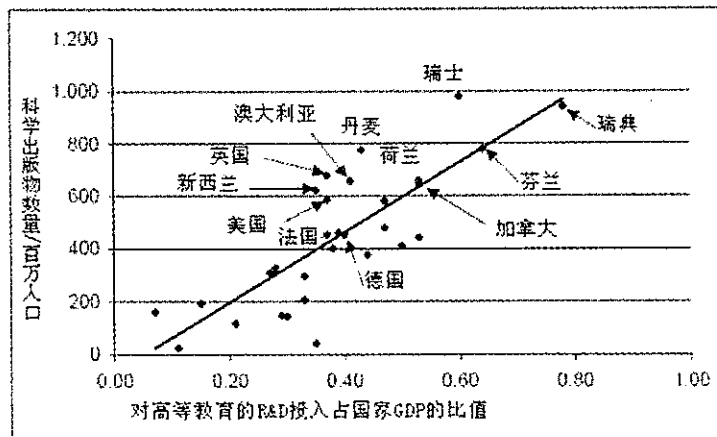


图6 高校申请专利的增长

回答这些问题的先决条件在于,能够获得可以激发科研积极性的信息。目前,这些问题还没有令人满意的答案。

图7展示了研究产出的一种指标——每百万人口的科学出版物数量与科研投入的指标——科研投入占GDP的比例之间的关系。我们可以大致地认为出版率的上升一般与资源投入的增加同步。在图7当中,线性趋势线显示了增加的资源如何使得产出也增加。



来源: OECD(2001,2003)

图7 研究产出与研究投入的关系,OECD国家1999

一些科研投入呈线性趋势的国家得到了平均水平的回报:法国、德国、加拿大、

瑞典和芬兰。然而,一些国家却比这些国家每单位资源产出了更多的出版物:瑞士、丹麦、澳大利亚、新西兰、英国以及美国。

这提醒了我们,出版物仅仅是科研产出指标之一。有证据证明,增长的资源投入将提高出版物的数量,但是其质量可能不会因此而提高。利用美国高校的数据,Abigail Payne 和 Aloysius Siow 于 2003 年发现科研投入每增加 100 万美元,将获得 10 篇新的论文和 0.2 项专利,但是对每篇论文的引文数却没有影响。这些结果暗示了增加资源投入将会提高科研产出但却未必是优秀的。看看前面的一些图表,很明显资源对于决定研究产出是很重要的。但也可以清楚地看到其他的因素也起到了一定的作用。

各国之间的差异是如何产生的呢?

科研产出是进行研究过程当中资源消耗和制度结构(或宏观经济)的产物。各国的科研产出在数量和质量上都存在差异。很明显,下一个问题就应该是:不考虑投入的多少,各国在制度上的不同特征能否解释这些产出的差别?

本节的第一部分对研究投入的全球分布进行了一个总揽。其中,提到了不同国家在选用绩效指标上的差别。然而目前各国并没有关注到这些绩效体制的具体细节。本节的下一个部分将考查其他一些可能解释这些差异的因素。

1. 组织和科研绩效

本文的一个核心内容就是高校受资助的途径以及(可能是分散的)现行的不同评价系统对人们进行科研活动时不同动机的影响。这些动机的变化可能会引起研究效率的提高。因此,我们将把资助制度或者评估改革作为改变研究产出的方法加以研究。

从某种意义上来说,所有的资助都是建立在评价的基础上的,因为高校的资助都有其相关的衡量标准。然而,这些标准可能不包括研究产出。因此,研究评估的范围将会因高校资助制度的不同在各国有所差异。

表 3 的第一栏显示了世界范围内高校所获得的资金随体制变化而波动的范围。每种资助体制都有其相关的评价系统。当然,这些系统不会是互相排斥的。传统上,高校获得一笔财政拨款时,其研究资助只是这笔财政拨款的很小部分。

其他传统的方法将在下两段中介绍。荷兰的高校资助是建立在高校在过去已获得资金多少的基础上的,而奥地利和法国的高校获得的资助则基于同中央部门的谈判。

在过去几十年中,由于高校作为科学源泉的重要性,其他的资助形式也越来越多。拨款委员会建立的直接科研资助是基于对于研究提案的同行评议这

样一个竞争性过程之上的。政府也可以对研究资助的去向进行指定。这种情况下,该过程也可能是充满竞争的,尽管这种竞争并不发生在政治领域。

同财政拨款相类似(第一种资助形式),加拿大目前基本上使用后两种方式。各省在很大程度上根据高校的规模提供直接的科研资金。

然而,世界范围内正在不断增加的一个重要的资金来源是基于对研究产出评价的资助。很多北欧的国家已经明确或间接地采用了基于研究绩效的资助体系。

受到评估的研究项目的比例及其受到评估的方式由此成为对高校获得资源过程中的一种要求。例如,英国采用的是表4中的后三种方法,因此,同德国获得财政拨款的研究项目相比,英国有更多的研究项目进行了质量评估。不断增加某一种形式的资助将会引起相应的资助评估工作的增加。

关于对制度结构影响的一个重要实例是芬兰的科研资助的发展历程。早在20世纪90年代,芬兰就把科技政策作为政策模式中的战略核心。比如,芬兰整体的经济危机造成了强制性的财政预算紧缩,却没有波及科研。这使得芬兰被看作战略科技政策史上的成功范例。

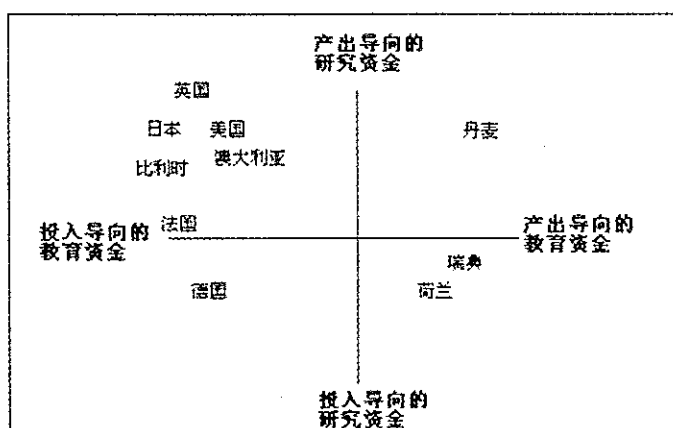
这一科技政策同样被应用于高校。在改革之前,芬兰高校就已经自成一番天地了。比如说,他们十分质疑高校和企业之间不断增强的相互接触。而在20世纪90年代,政府就增加了通过竞争过程中分配的资金来改变高校获得资金的方式。1990年到1996年之间,竞争性经费占政府调配资金的比例从42%增加到了52%。高校通过竞争和其他外部资助的分配所获得的预算比例从33%增长到了47%。可以清楚地看到,随着芬兰科研经费分配体制逐步向竞争性过程转移,受到评价的研究项目的比例也增加了。在前面已经讨论过了,这种变化似乎并没有破坏芬兰的科研绩效。同样需要强调的是,芬兰的拨款委员会理所当然,也进行了自我评价(见下文)。

表4 范例国家中的经费和评估

资源转化方式	评估形式	范例国家
核心经费和财政拨款经费	没有研究的评估	a)完全基于学生规模的财政拨款:德国、意大利
		b)研究领域的评估:挪威、瑞典
		c)一些根据绩效评估的经费分配:芬兰、丹麦
历史上的	无明确的评估	荷兰(进行评估但不与经费挂钩)
协商	可能包含质量评估	进行研究评估的国家:法国
		不进行研究评估的国家:奥地利

由财政委员会批准的直接项目资金	同行评议	大多数国家
合同	产出	大多数国家
研究绩效指标	基于研究质量的评估	数量指标:澳大利亚,波兰
		知情同行评议:英国,香港

以产量为导向的资助方式同样可以引入教学活动中。从图8中可以看到世界范围内不同体制的变化。然而,总体而言,西北半球的国家明显地有一种增加的集中趋势:对于高校科研产出的资源分配量在不断提高。但是只有很少的一些国家在对学生的教育上转向了基于产出的机制。



来源: Jongbloed and Vossensteyn (2001)

图8 高校资助机制中的相对绩效导向

表5 大学研究资助不同方法的优缺点

	优点	缺点
基于绩效的	精英阶层的	高成本
	提高个人及社会公共机构的绩效	可能导致研究的同质化
	竞争可能引起效率的提高	可能阻碍富于风险的研究
	鼓励研究以适当的方式结束并以适当的方式传播	可能导致“出版物膨胀”
	对政府经费的公共义务	可能导致“学术化”研究
	鼓励强调研究策略	把研究从教学中分离出来
	提供连接研究和政府政策的机制	奖励过去的绩效
	资源的集中使最优秀的研究部门能够与世界领袖级的研究机构竞争	可能导致“过度的”政府干预

教育规模	低成本的管理	对于研究绩效的激励十分有限
	允许高校投资新的研究领域	分配研究经费的人掌握过多的权力
	允许长期研究	较差的责任感会导致“象牙塔”式的研究
	鼓励研究多样性	巩固了慵懒学术的陈规
	允许任何高校进行研究	学生数量和研究工作之间可能并没有关系
	能够鼓励研究与教学的整合	与政府政策无关的资源分配
	保护自主权	将资源过于分散地进行分配则会丧失与国际一流研究所竞争的能力

来源:整理自 Geuna 和 Martin(2003)

2. 其它的国际差异

还有一些其它的方式可以提高研究产量。其它体制上的变化,如高校的集权化程度或者市场力量所允许的体制变化程度,都可以影响研究进行的程度。另一个可以解释瑞典的科研高产出的体制特点是科研人员不承担本科生的教学并因此可能有更多的时间用于科研。

很多国家的研究活动是在独立的研究机构中进行的,同大学的交往较少,比如法国和德国。在德国,大多研究在马普学会的各个机构中进行,而在法国,科研活动主要在国家研究中心(CNRS)和国家健康与医学研究院(IN-SERM)的实验室中进行。这种现象在图9中有所反映,显示了政府和高校R&D活动的差别。法国和德国高校外的研究机构承担了大量政府资助的研究项目,而且对欧盟有很大影响。

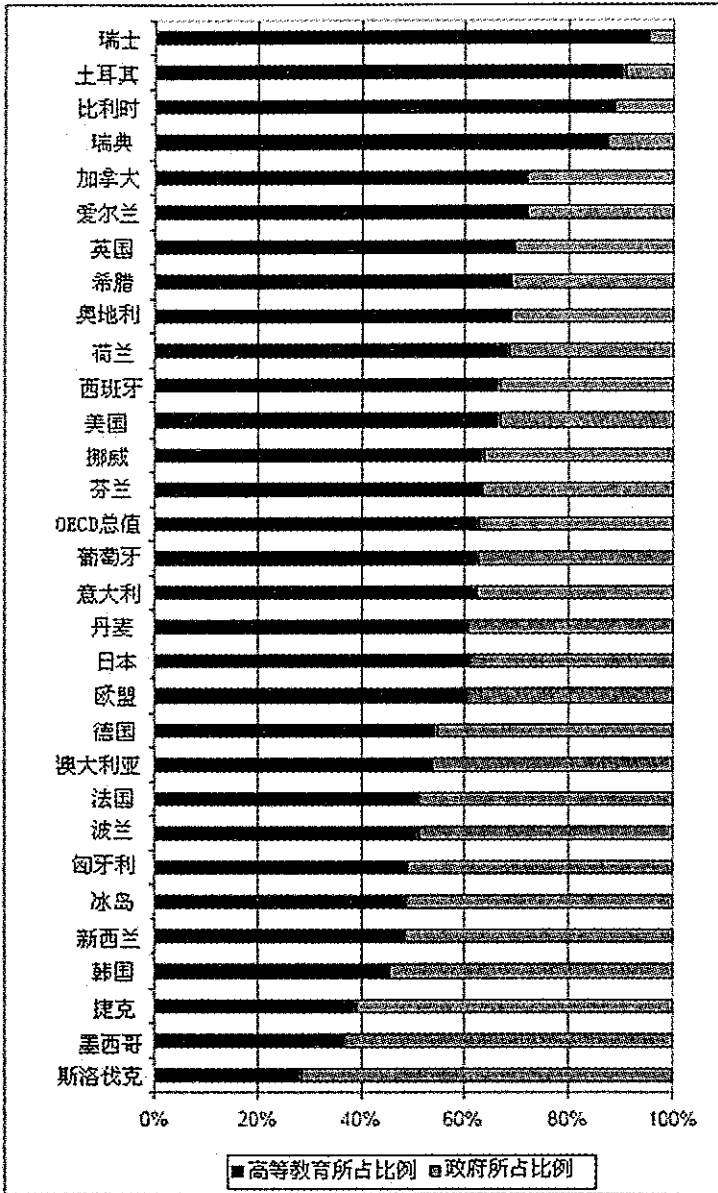
依靠科研机构进行研究活动有下述意义:

- 对科研成果的产出施加的竞争压力小一些。体制结构可能可以解释图2中德国和法国相当差的研究绩效。

- 很明显,高校在政府科研投入(图9)中的重要性和一国引文数的排名之间有一种合理的相关性。尽管出版物并非研究系统的目标所在,但还是应该把这种相关强调出来。

- 对高校科研产出的评估需求在不同国家会因高校承担科研活动的普遍程度而改变。如果研究活动不是在高校进行,就无需评价高校的产出了。

- 受政府控制的研究机构,其成果可能会受到更广泛的社会目标的影响,如环境保护或健康安全问题。一些国际性的经济危机说明加强对政府研发的投入有助于科研生产力的提高,但只有那些在高校分担了更多公共部门研究项目的国家,这种促进的效果才会更好。(未完待续)



来源:OECD(2003)

图9 高等教育部门和政府部门支出经费所占比例

(彭颖舒 译)