

内部资料
免费交流

京内资准字0609-L0057号
北京交大印刷厂印刷

中国教育科研参考

2014年 第 20 期
总第(342)期

中国高等教育学会编

2014年10月30日

目 录

大学排名与教学评估·····	邬大光(02)
德国高等教育发展中心大学学科“卓越排名”的特点及其启示 ——以2010、2011年“卓越排名”为例·····	崔艳丽(03)
系统多样性理念下的大学排名与分类:实践与借鉴·····	吴 伟 邹晓东 范惠明(08)
《美国新闻与世界报道》本科院校排行榜评估指标体系研究·····	李函颖 刘宝存(16)
QS大学排名体系剖析·····	刘 强 丁瑞常(22)
我国重点大学建设世界一流大学的奋斗方向研究 ——基于两个著名世界一流大学排行榜的视角·····	田 锋(28)
高等教育机构排名的柏林原则·····	(31)

编者的话:建立高校分类体系、实行分类管理,引领高校办出特色、争创一流是《国家中长期教育改革和发展规划纲要(2010-2020年)》提出的一项重要任务,也是教育行政部门和学界的共同追求。高等教育大众化阶段对高等教育系统多样性的期待与日俱增,高校分类从宏观上解决发展战略定位问题,通过科学归类促进高等教育体系多样化、高效率;大学排名的评估标准多样化、受众主体多元化,积极发挥社会引导、公众服务的作用。本刊以“大学分类与排行”为选题,集中选编若干文章,供读者参阅。

主 编:王小梅 本期执行主编:范笑仙 责任编辑:聂文静
地 址:北京市海淀区文慧园北路10号中教仪楼中国高等教育学会《中国高教研究》编辑部
邮 编:100082 电 话:(010)59893297
电子信箱:gaoyanbianjibu@163.com

大学排名与教学评估

邬大光

大学排名与教学评估是两个完全不相干的话题，可是如果把这两个不相干的话题放在一起作对比，看看它们在我国遭遇，就会发现其中隐含着许多有意思的现象。最为显见的一个事实是：国内外大学对大学排名与教学评估，在许多方面都有明显的反差。

第一个反差：国内外在大学排名主体上的反差。在国外，开展大学排名工作的不是大学，美国是《美国新闻与世界报道》和《福布斯》财经杂志，英国是《泰晤士报》，日本是《朝日新闻》和《钻石》周刊，德国是《明星》周刊，加拿大是《麦克林》新闻周刊。可以看出，在国外做大学排名的，几乎都是新闻媒体，我至今还没有看到一所国外的大学在做大学排名。再来看我国，以前只有中国管理科学院广东分院和深圳的网大做大学排名。这两个机构在搞大学排名之前，国内恐怕没有多少人知道它们的存在以及机构的性质。

在国外，不仅仅是大学排名，凡是搞各种排名的几乎都是媒体。道理很简单，媒体从事大学排名的目的，就是推销杂志。在国外，大学排名不知挽救了多少发行量下降的刊物。最为典型的案例就是《美国新闻与世界报道》，通过做大学排名，走出了发行量的低谷。最近几年，我国陆续有一些大学和研究机构开始做大学排名。真不知道在我国，还有多少大学想介入此事。而我国大学做排名，挽救的是什麼，我就不太清楚了。

第二个反差：国内的大学对大学排名与教学评估在底气上的反差。对待大学排名，尤其是国内的大学排名，国内的大学似乎都很有底气，几乎所有的大学都认为自己在排名榜上被低估了，普遍有一种不服气的感觉，看起来对自己的质量都很有信心。可是，面对国外做的各种大学排名，国内的大学又都“集体失语”。更为有意思的是，几年前进行本科评估时，好像国内所有的大学都没了底气，越是好的大学越显底气不足，尽可能地把评估时间往后移。在迎评前，大学里出现的各式各样的

应付手段和“高招”，以及各高校之间的相互“取经”，无一不暴露出大学的心虚。

第三个反差：国内外大学对大学排名与教学评估在态度上的反差。看一下今天的国外媒体，抨击大学排名乃至对此不屑一顾的声音和现象比比皆是，而对教学评估的批评声则显然没有那么响。而我国的教学评估几乎是在贬大于褒的“声讨”中完成的。为什么每年一度的大学排名，总是能够吸引公众和媒体的极大兴趣，而原本应该扎扎实实、长期进行的教学评估却渐渐退出了人们的视野。面对大学排名的日渐升温 and 教学评估的前途未卜，真不知道这对我国高等教育质量而言，究竟是福气还是不幸。

面对这些反差，我们要问，究竟是大学排名对大学发展有好处，还是教学评估对大学发展有好处？大学排名的不断升温与教学评估的不断降温之间的反差，真实地反映了现阶段人们对大学理解的水平和深度，折射的是对质量的短视与急功近利，以及商业文化对大学精神的“蚕食”。

国内有人说：“排名是西方人的嗜好，在盎格鲁-撒克逊（Anglo-Saxon）国家最为流行。”我们恐怕暂时还不需要这种嗜好，千万别让它误导了我们的大学。因为说到底，大学排名是一种市场行为，目的就是为了吸引人们的眼球，并不值得大学过于认真对待。国外知名大学很少标榜自己在某排名榜上的位次，也很少把排名榜上的位次作为宣传自身形象的重要战略。

今天讲建设高等教育强国和提高人才培养质量，不是靠大学排名来完成的，主要还是靠扎实、有效的质量监控。对教学评估的压倒式谴责，恰恰反映了对大学价值认识的缺失。教学评估的方式和内容也许需要改进，但其固有的功能和价值不可抹杀。教育管理部门和大学校长，对大学排名应该多一点儿淡定，对大学评估应该多一点理直气壮，这才有助于办好大学，才会改变在西方大学排名榜前的“集体失语”。

（邬大光，厦门大学副校长，福建 厦门 361005）

（原文刊载于《中国教育报》2010年9月6日第5版）

德国高等教育发展中心

大学学科“卓越排名”的特点及其启示

——以2010、2011年“卓越排名”为例

崔艳丽

学科排名是大学排名的重要组成部分，但它又不仅仅是大学排名的一部分。随着我国建设世界一流大学目标的确立，创建世界一流学科成为世界一流大学建设的先行军。在这种形势下，对学科评价和学科排名进行研究的需要日益凸显。从世界范围看，目前学科排名中比较有特色的是德国高等教育发展中心（The Center for Higher Education Development，下文简称CHE）“卓越排名”。笔者拟以2010和2011年CHE卓越排名为例，对卓越排名的评价方法、指标体系和特点进行分析，并从中发掘对我国大学学科评价的可资借鉴之处。

一、CHE大学排名的缘起和排名方法论

CHE由德国贝塔斯曼基金会和德国校长会议在1994年发起建立，是一个非营利性有限公司。CHE将自己定义为高等教育的改革智囊团、政府和高等教育机构项目合作伙伴，它的创办目标之一是发展基于科学方法的大学排名。首份大学排名于1998年发布。

CHE的大学排名分为四种，分别是大学排名、研究排名、卓越排名和就业能力评级。研究排名和就业能力评级是对大学单个方面的评价，而大学排名和卓越排名的共同点是它们都是针对学科的排名，但两者也存在一定区别：大学排名主要评价对象是德语国家的高等教育机构，主要为满足大学新生的要求，涵盖九个维度多达几十个指标；卓越排名的评价范围则涵盖欧洲所有国家，兼具辨识欧洲杰出的学术部门及为未来的研究生提供各高等教育机构所开设学科、专业及其质量的信息等综合性目的，运用四个维度数十个指标集中评价大学的教学和学术质量。二者相

比，卓越排名更集中于对高等教育机构学术水平的评价。

卓越排名的目标群体是各国想攻读硕士或博士学位的本科生、欧洲高等教育机构、德国社会组织及全体社会。就本科生来说，主要是为他们提供基于学科的和指标的更为深入的大学信息；对高等教育机构来说，主要是使他们能够吸引更多的杰出学生申请，并使他们以宽广的视角与其他高等教育机构的相似长处进行比较以使自己进步；对社会组织和全社会来说，如基金会和其他资助机构能够有更多的信息来挑选受资助者，公众也能更深入考察在研究和教学方面能力最强的高等教育机构，还有益于推动知识社会和相关学科领域为终身学习做好准备。同时，它还作为CHE拓展民族视角战略的一部分，力图在展示欧洲大学竞争力的同时推动欧洲高等教育区的发展。

CHE排名虽然出现时间不算早，但已产生很大影响，美国未来高等教育委员会大学排行、加拿大《大学信息导航报告》都受到了CHE排名的影响，欧盟委员会拟在2013年推出的“欧标”世界大学排行榜“大学多维排名”在评价方式和指标上与其多有相似之处。这得益于它独特的排名理念和方法论。

CHE排名通过四个关键领域的方法论标准，区别于其他排名：第一，严格说来，排名只是学科相关的排名。CHE坚持这样的理念：没有“最好的高等教育机构”，每个高等教育机构不同学科的强势和弱势勾勒了其独特的轮廓，因此，对高等教育机构的整体比较并没有超越学科水平。第二，多维排名。这意味着，对给定一个学科的评价不是赋予每个指标不同权重而进行总的排

名。即使一个具体学科内部，比如课程的开设、导师指导和设备，高等教育机构也都有其特有的长处和弱项。并且，对设计每个指标的权重也并没有经验或理论方面的基础。第三，从不同视角描绘高等教育机构。除关于系和学习项目的事实之外，讲师、学生和校友的观点在制作排名时也予以考虑。这与其他建立在特定群体视角上的大学排名区别开来，允许在主观评价和客观指标之间进行对比。第四，不提供具体的位置，而是将大学分成三组，即高组、中组和低组。这种做法的目的是避免票面价值指标的微小变动被曲解为质量和表现的差别。

卓越排名也坚持了同样的方法论。CHE认为卓越是多样化的，高等教育机构是异质的而且没有最好一说，高等教育机构中各专业、学科和系的表现存在诸多不同，在它们内部也是如此。而且它认为，当分配位次时，存在偶然的波动导致的表面差别被错误解读为真正不同的风险，如果高等教育机构在排名中相互接近，也有将最小的不同解释为许多排名位次的风险，比如，排名10与排名20的高等教育机构实际上差别非常小。因此，它选择对高等教育机构进行深入的分析，要提供一个多维的排名，而不是不谨慎地将研究结果加在一起给出一个总的排名和加冕一个总的胜利者。卓越排名采取了以下做法体现上述方法论：一是以“星”代替排名位置。不像其他排行榜那样给每个高等教育机构确定一个“排名位次”，而是为各个指标中表现特别好的系分派相应的“星”，作为分组的基础。二是将排名与学科关联。对高等教育机构整体的合计排名不能够为将来修习一门学科的潜在学生的决定过程提供充分有用的信息，且会模糊特定高等教育机构学科之间的区别，因此卓越排名只针对学科，是与特定的学科联系在一起的编排，是基于各个维度而不是将其综合之后的排名。三是方法的多样化。卓越排名意味着提供不同视角来看待高等教育机构及其学科，除了学科的事实和对其提供的项目进行研究，学生的观点也是排名的组成部分。

最早的卓越排名是2007年发布的生物、化学、数学和物理学科排名，其后的2009、2010、

2011年度也开展了排名活动。2009年排名增加了政治科学、心理学和经济学学科；2010年对自然科学重新调查后进行了排名；2011年新增了英美研究、日耳曼语、社会工作、机械工程、教育、计算机科学等学科的排名。下面以2010和2011年卓越排名为例，分析说明其预选过程与指标体系。

二、CHE卓越排名的预选过程与指标体系

排名始于预选。预选主要是对数据进行甄别、筛选和分析。在数据甄别和筛选阶段，高等教育机构的学习环境、各种项目和其他一些标准会进入考虑范围。CHE和外部合作者将进行问卷调查并检视调查结果，去除一些不能被大部分高等教育机构回答的问题而加进新的内容。接下来，根据数据的有效性和完整性，来自高等教育机构的数据和学生调查问卷搜集的信息将被分析和排名。预选包括多个指标，只有许多院系在某一指标上提供了完整和有效的数据，该指标才能被赋予“星”，也即，各指标在满足以下三个条件后才会被赋予“星”，才能在预选中得到使用：有充足比例（至少超过一半）的高等教育机构能够提供可信的和独立的数据；在合理范围内它们提供的数据在国家间具有相当的可比性；各个指标上的学生评价只有当学生一般群组规模达到至少15名才会得到使用。只有达到上述三个标准，各指标才能真正在预选过程中发挥作用。

在2010年，预选中实际运用的指标有：①硕博水平学生判断，包含整体状况、训练和课程、辅导和建议、计算机设施、实验室与图书馆。对于博士生还有研究方向的判断，比如通过发表论文、参与工作组和参加会议而加入科学社区的机会。②博士学位教师中国际化比例。③博硕士学生中国际化比例。④教师、博硕士学生中性别比例。⑤可获取的图书馆订阅的纸质和电子专业学术期刊数量。⑥每10个科学家中在主要学术期刊编委会中任职人数。⑦师均获得的知名学术荣誉的数量（政治学独有）。⑧近5年举办的国际会议数（政治学独有）。⑨在暑期学校中任讲的教师比例（政治学独有）。衡量表现的“星”被分配到这些指标，各学科在这些指标上的表现决定了它们是否能在预选阶段脱颖而出入围正式的排名。预选结果的分析呈现比较说明了各国

在各学科及各大学在学科各方面的实力。以2010年卓越排名为例,从预选阶段总体情况来看,入选系较多的国家及其数量是英国120个,德国79个,荷兰41个,法国35个,意大利29个;从学科来看,在生物、化学、经济学、数学、物理、政治科学和心理学英国入选系的数量分别是21、13、23、9、11、21、22,而德国是15、12、7、11、16、8、10,荷兰则分别是7、7、7、3、6、2、8。对入选的大学也进行了分析,首先是每个大学中包括的卓越学科数量占入选大学的比例,以呈现各大学的卓越学科;其次在各个学科内部对上述8个预选指标进行了深入的分析,展示该学科的优劣之处;最后是国际教师和学生的比例、不同资格证书层次的性别平衡以及学生对整体学习状况的判断等内容。

预选过程的结束是正式排名的开始,而排名

表 1 排名维度

指标分类	排名维度	表征内容
事实指标	研究	研究水平
	应用	教学和研究的实践联系
	国际化	教学和研究的国际化水平
主观指标	学生定向	学生对学习情况的判断和观点

指标体系是任何一个排名的关键特征,要了解一个排名的特点,必须首先考察其指标体系。CHE学科卓越排名包含了事实指标和主观指标两大类指标,主要考察高等教育机构的各学科在研究、应用、国际化和学生定向四个方面的表现。表1提供了排名维度,表2展示了指标体系及其内涵,表3则分析了各维度赋值与分组的有关方法和标准。

具体的赋值与分组要根据不同的学科和指标进行,但大体来看,赋值与分组也遵照一定的规则。整体上看,事实指标和主观指标两大类的赋值与分组方法存在较大差异。

三、CHE卓越排名的特点及其对我国大学学科评价的启示

尽管CHE卓越排名不是国际性的,但它在欧盟范围内对大学进行排名,超越了国家和民族界限,是大学间的国际比较,它对于什么是卓越,什么是一流,如何评价,有自己独到的见解和方法。总的来看,CHE卓越排名具有五个显著的特征。

第一,排名建立在卓越多样化理念基础上。这不仅体现在排名指标的丰富性和全面性上,还体现在不对大学进行总的排名,不判定一个总的优胜者,而是基于学科进行评价;即使是对学科

表 2 排名指标体系及其内涵

维度	指标	内涵
研究 维度	出版物	基于 WisoNet 等数据库
	国际出版物	基于 ISI web of science 等数据库调查
	每出版物引用	基于 ISI web of science 等数据库调查,主要是与均值对比
	外部资金	3年平均,CHE 调查数据;主要包括德意志研究联合会、联邦基于与研究部、各种基金、德意志学术交流中心、私营部门资金、非德国欧盟推动者及其他机构的资助
	推广	6个学期,CHE 调查数据
	专利	衡量研究的转让和应用维度;考察3年内每10个科学家产生的发明;主要是在工程领域
应用 维度	学生判断	专业方面主要是学生对学校为他们提供的课程中专业和市场定向的程度的判断。包括与事业和劳动力相关的信息研讨,教授与工作相关的专门技能的特殊服务和课程,对找工作给予的支持,毕业论文题目与专门知识和有助于毕业后找工作之间的关联的评价。实践方面主要是学生对学校提供的实践阶段的支持,实践阶段事前和事后组织与理论阶段是否密切配合,工程习明纳和由实践者开设的讲座的范围和质量的评价
	实践检验	对外部与专业和工业相关的技能介入课程的程度,课程中讲授方法技能、社会技能和实践技能状况的评价
	来自私营部门资金	3年平均,CHE 调查数据
	专利	衡量研究的转让和应用维度;考察3年内每10个科学家产生的发明;主要是在工程领域
国际化 维度	学生对海外学习支持的判断	学生对交换项目和学校的吸引力、是否提供足够场地、申请阶段的支持和建议、财政支持(补助金、免学费),有没有浪费时间、学习效果、融入海外学习等情况的评价
	国际取向指数	比照四个标准进行评价:有强制海外分析访学或联合学位项目;有海外学生;教学力量的国际化;开设外语课程
	来自欧盟的资金	3年平均,CHE 调查数据
学生 定向 维度	整体学习形势	对该领域整体学习形势的判断,不是多个问题的复合指标
	支持	对接触教师的机会、教师办公室工作时间、非正式商讨、对学术论文和报告的评论、检视等的评价
	课程	对课程的国际化幅度,教学中的多学科联系等的评价
	学习能力	课程规则,课程选修,课程与考试规则、审计组织、质量体系透明度的协调性的评价

表3 指标赋值与学科分组方法

指标类型	赋值与分组
事实指标	<ul style="list-style-type: none"> · 采用四分位法,第一四分位是高组,第二和第三四分位是中组,第四四分位和0值被划分到低组。因此,有时会发生最后一组显然包含超过系的总数1/4这样的情况。有时组边界会进行调整,比如,中间组的第一个系的数据非常接近高组的最后一个系的峰值,它就会被提升至高组。 · 研究维度中高等教育机构只有一半指标都进入高组才称得上研究密集型。出版物按照文章长度、作者数量、期刊种类等赋予不同的权重,不到5页1分,5至9页2分,10至19页3分,20-39页4分,40至100页5分,100页200页8分,200页以上10分;2个作者半分,3个作者1/3分,4个作者1/4分,如没有长度信息,按照该类型文献平均长度计算;分国际国内,根据学科选取数据库;经费、出版物、推广、专利都为绝对值和相对值;分析经费数量的同时还分析了经费结构;还对教授进行了声誉调查,并与经费、出版物排名的关联度进行比对。 · 应用维度中高等教育机构只有一半指标都进入高组才算在该维度实力强。专业和实践两方面的学生判断都是赋1-6分;实践检验属于应用相关产出范围的指标,分成学士和硕士学位项目,15分以下低组,15-29.9分中间组,30是高组;资金、专利都为绝对值和相对值两个方面。 · 国际维度中高等教育机构只有一半指标都进入高组才算在该维度实力强。学生对海外学习支持的判断根据情况赋1-6分;国际取向指数最多有14分,达到6分为高组,达到2分为中组,2分以下为低组。
主观指标	<ul style="list-style-type: none"> · 学生定向维度中高等教育机构只有一半指标都进入高组才算在该维度实力强 · 使用平均值和方差概念,方差小于0.05 · 超过15人才能进入排名 · 有至少10%的问卷返回搜集到的数据才有效

的排名也不是整体性的,不对各个指标进行合计,而是建立在以维度和指标为基础之上的,每个学科排名表实际上包含了数个表格,这些表格是各学科在各维度、各指标上表现的具体呈现。

第二,既重视客观(事实)指标,又强调主观判断。这表现在四个维度中包含一个基于调查的主观指标“学生定向维度”,2011年排名对每个学科的评价中都包含这个维度;主观指标在客观指标维度中也有所体现。如国际化维度包含学生对海外学习支持的判断,这种做法综合了主客观指标在表征高等教育机构质量方面的长处,目的在于使两者优势能够得到充分发挥。

第三,尊重高等教育机构和学科特点。CHE将高等教育机构分成大学和应用技术大学两种类型,对大学和应用技术大学同一个学科分开排名,以体现两种机构不同的特色;还运用不同的维度和指标考察各学科,有的学科的排名包括了四个维度,而有的学科则包含三个维度或两个维度,就2011年排名来说,完全一样的学科排名是找不出来的;同时,在出版物指标上,区分国内

出版物和国际出版物,连同学科特色作为选用数据库的依据,在适当的数据库中进行数据筛查、搜集和分析,为排名提供科学合理的保障。

第四,数据处理手法细腻。如,在外部资金指标上,CHE对“其他”这一数据的种类进行了规定,以使数据搜集较少发生错误;在出版物指标上,对作者数量、作者调动和署名问题、文献类型、文章长度都进行了细致的规定,以使数据能够真实地反映客观状况;在资金、专利、推广等指标上,还区分了教学科研人员中的教授、科学家等不同群体,以便更清楚地反映高等教育机构及其成员的行为;CHE还对各学科进行了声誉调查,将其作为额外信息并对资金、出版物排名与声誉排名之间的关联进行了分析,2011年排名报告显示,它们之间有高度关联的关系。

第五,独特的分组法。从上述各学科排名表中可以看出,CHE卓越排名中的排名只意味着对学科进行分组,即高组、中组和低组,它主要基于四分位的划分,同时,该学科在某一维度上是否很强,要看其在该维度的各指标中有没有一半达到高组,如果达到则该学科在该维度实力就强。这种排名策略与目前较为流行的排名中给予高等教育机构位次的排名方式是不同的,它增强了高等教育机构位次的“弹性”,能够较好地避免历次排名中因数据的波动放大为质量差异,而给高等教育机构造成的不利影响。

毋庸置疑,CHE卓越排名是极富特色的,它的独特之处吸引着一些排名机构的目光,产生了积极影响。CHE坚持的理念被欧盟委员会大学排名所借鉴,欧盟委员会拟在2013年推出“欧标”世界大学排行榜,并将这一排行榜直接命名为“大学多维排名”,它采用“多维”大学排名体系的评判标准,从大学的功能、类型和学科多维视角开展大学排名研究,将会使各项指标取得更好的平衡,更为全面地评价大学。在2010年多维大学排名的设计和测试可行性阶段发布的文献看,“大学多维排名”主要包括下列维度和指标:教学和学习维度、学生维度、研究维度、知识转移维度、国际取向和地区衔接,与CHE卓越排名有很多共同之处。近年来,我国正在创建一批世界一流大学,一些高水平大学正加快建设步伐,积极

为迈向世界一流进行有效的准备，建设世界一流学科即是其中重要途径之一，开发和利用学科评价和排名工具，将会对该项实践起到积极的促进作用。CHE“卓越排名”最主要的特点是针对学科进行的“定制”的和个性化的大学排名，它的多元卓越理念、多维指标体系，高中低分组方法，对大学独特性的重视，给我国世界一流学科和优势学科的建设评价带来新鲜有益的启发：

第一，明确目标群体，设计多样化指标，使用丰富细致的数据。对于大学排名来说，首要工作是明确目标群体，有针对性地设计较为全面的评价指标，搜集和使用丰富和细致的数据。CHE卓越排名的目标群体有大学、教师、大学应聘者、学生等。对于大学来说，需要的信息全面一些，对于其他目标群体来说，他们从自己的利益出发，在获取大学整体表现的同时，还要求从大学排名中获取最关注的信息。CHE卓越排名以学科、维度和指标为基础，恰恰满足了目标群体的独特需要。明确排名的目的，增强排名的针对性，是一流学科和优势学科评价的前提基础，也有利于更好地发挥学科评价的效果。

第二，关注学校类型、学科特点。我国大学体系中明显地存在着向研究型大学的趋同，但并不是所有的大学所有的专业有适用的一套发展思路。CHE卓越排名将德国高等教育机构分成大学和应用科学大学两类，目的是准确评价两类不同大学中相同学科之间的差别。在不同大学中，同一学科可能受大学历史和学术特质的影响而形成不同特色，如何在学科评价中凸显不同大学的学科特色，将是对学科评价的考验。这就要求将学

科的评价和大学的评价结合起来，真正评价出学科的优势和独特性。

第三，学科位次宜使用弹性表示方法。虽然大学及学科排名已经成为大学和学科质量监督和保障的一个重要手段，但排名表评价的是大学和学科的表现而不是大学和学科质量本身，即质量与质量的表征之间有一定的张力，其确定的位次不一定是准确的，也因此排名受到诸多批评和质疑。然而，大学和学科排名以横扫之势风靡世界，重要的原因之一是大学和学科的质量并非完全不可评价，CHE采用分组而不是一一对号入座的方式，采用分组方式分配学科的位次，使学科位次的弹性有所增加，也增强了排名位次对质量表征的适用性，分组方法独辟蹊径，值得借鉴。

第四，学科排名要引导和鼓励大学办出特点，发展优势。CHE卓越排名之所以不进行总体排名，因为它认为没有“最好的高等教育机构”。一所大学在此学科突出，在彼学科可能欠缺，一个学科在此维度优异，在彼维度可能较弱，不可能所有学科在所有维度上都是卓越的，没有完美的大学。因此，对大学的评价分学科，对学科的排名分维度和指标，是符合上述判断的。这种排名方法能够突出各高校的优势学科和学科的优势维度，明确了大学各方面的实力和差距，肯定了它们的特色，能够产生和发挥引导大学发展优势、办出特色，促进高等教育体系整体丰富与繁荣。

（崔艳丽，南京师范大学教育科学学院博士研究生、江苏科技大学助理研究员，江苏 南京 210097）

（原文刊载于《中国高教研究》2013年第1期）

系统多样性理念下的大学排名与分类：实践与借鉴

吴 伟 邹晓东 范惠明

面对来自美国、澳洲甚至部分发展中国家和地区高等教育的激烈竞争，欧盟近年来实施了一系列提振大学发展水平的措施，即许多政策文本反复强调的“现代化”（modernisation），主要包括：在欧洲高等教育区（EHEA）框架下提升产学合作水平，强化科技成果转化，合作建立研发中心；提高大学创业能力并凸显研究、教育、创新之间的“知识三角”（knowledge triangle）关系，创建知识联盟（Knowledge Alliances）；培养拔尖创新人才，提高学生的就业能力、创业能力、创造性、适应性、创新精神，推动校企合作培养；继续扩大资金投入。这些趋势在里斯本战略（Lisbon Strategy）之后的2010年6月17日欧洲理事会会议通过的《欧洲2020》（Europe 2020）战略中得到全面反映。在此背景下，近年来欧洲也不断反思单纯以研究（Research）为旨趣的大学评价（排名）的错误导向，视其相悖于大学系统多样性的内在逻辑，忽视社会各主体对高等教育需求的多元化。基于此，以系统多样性为旨趣的大学排名与分类实践逐渐表现出来。

一、系统多样性与大学分类排名理念

多样性（diversity）是一种西方社会发展和组织管理理念，也是一种重要的社会思潮，其基本目标是建立促进个人发挥最大效能的组织环境，使组织成员不因其某些特性（如性别、地域、出身、种族、宗教、性取向、社会地位、身体能力）或仅仅因为不代表社会主流而被排斥在组织之外。虽然上述定义对象偏重于个人，但从理念诉求看，系统多样性排名与分类思想很大程度上即源于这种多样性理念。从实际来看，欧盟在语言、民族、经济、科技、教育体系等方面的差异或多样化的程度超过了其他任何一个多民族国家和区域联合体。在此背景下的欧盟大学排名或分类，必须考虑各国在高等教育上的多样性及其带来的特殊要求。

（一）单一旨趣排名的宏观失效

大学排名其本意旨在客观上提供有关大学表现

的有用的、可比较的信息，以为师生的学术选择和政府的政策制定提供便利。但是，鉴于国别差异巨大，国际可比较和有意义的连续数据比较缺乏，因而比较不同国家的院校整体特别是通过量化的手段衡量大学质量非常困难。具体来看，各种指标对研究和知识创造的衡量仅仅是“貌似可信”，因为这些指标只强调了研究的狭窄定义，并未体现出知识对社会和经济的贡献的广泛性，特别是没有给予所谓“知识三角”以应有重视。正是从这个意义上说，传统大学排名误导了人们对质量概念的认知，以为排名在前的大学必然比靠后的好。此外，对大学排名的普遍批评之一还在于这种给出单一位序的排名系统在识别院校不同使命和类型上的失败，其忽略了大学系统的整体功能和院校使命的多样性，从世界范围来看，“对顶级大学排名的痴迷正在让我们拒绝‘世界一流’的全球高等教育系统”。Boulton指出，“既然大学在回应社会需求时的功能和使命上具有重大差异，那么，仅用精确的单一指数衡量显然不合适……现有影响力较大的全球排名中，鼓励院校以单一的研究导向模式占据主导，刺激了大学单枪匹马式的‘学术军备竞赛’，从而削弱了大学系统多样性和腐蚀多数大学以其他方式发挥贡献于社会的潜力”。而从政策导向上来看，“痴迷于全球顶尖的1%的一流大学排名可能颠覆各国政府的政策目标，而迎合了其他利益主体所设定的指标，追逐若干精英大学或旗舰院校大可不必”。

（二）系统多样性思想的来源

高等教育功能的系统性与社会需求的整体性应该是吻合的，大学职能发挥的差异性与社会需求的多样性也应该有其对应，分层分类的政策框架是引导大学发展的理论基础，就像卡内基分类和加州高等教育总体规划对高等教育系统所起到的作用一样。系统多样性排名与分类思想源于这种理念：大学的质量不应该用统一的标准去衡量，大学的面貌不应该是同一种状态，大学系统应该是多样和谐的整体，即“多元统一”，特别

体现在横向的功能分化上。所以，创建若干所世界一流大学的理念必须让位于创建一流的大学系统。世界一流大学系统的特征是：开放和竞争，能够为最大多数学生提供最广泛机会的开放的和竞争性的教育；横向分化且融合的院校组合，提供宽广的教育、研究和学生体验，在各自层次领域都有上佳表现；开发公民需要的知识和技能，贯穿终身并为社会所要求，同时吸引海外精英；毕业生能够在劳动力市场取得成功，推动和保持个人、社会和经济的发展，并巩固公民社会的基础；在全球市场中成功运作，具有国际化视野，并积极响应变革。可以看出，欧盟所谓的多样性是处于追求“多点”之间的互动从而提升整体素质的考虑，这与“没有创新网络中多主体互动的单一创新模式不可能建立起良好的竞争优势”的创新管理思想颇为吻合，至少在“研究”这一点上是这样。系统多样性的理念更加注重高等教育系统包括经济价值在内的公共价值以及系统内部的共生共存，并且前者是后者实现的基础，后者是前者的显著表现。

（三）系统多样性排名与分类的意义

大学排名与大学分类都是对大学进行评价，或事实上影响着评价；大学分类可能成为排名的基础或条件，如卡内基分类对美国新闻与世界报道排名（USNWR）的影响那样。从本质意义上看，排名与分类是系统多样性理念下更为客观的评价。单一的评价体制使得所有大学“狭窄于科研”，对大学系统是一种潜在的损害，“对全球顶尖大学的痴迷是冒着高等教育系统被‘别有用心’的指标所颠覆的危险……基准应该用来提高整个（大学）系统的能力和质量，而不仅仅是奖励精英院校或旗舰大学的成就”。因此，以系统的观点去评价大学及其活动就成为必要，而大学分类就是能够建立这种评价的基础。针对现有科研单一分类或排名的批评大多归结为其“糊里糊涂”的笼统，持此观点的学者认为，无论国际还是国内，评价应该充分承认不同院校类型的使命差别。大学系统与分类思想与高等教育机构评价（排名）之间紧密相连，以系统思维建立起合理的分类，并在分类的基础上进行评价（排名），就能够得到更有价值、更为客观的结果。在欧洲，最新的排名与分类行动具有类似的动机、相

似的数据和指标，分类预示着一一种排名，良好的排名实践原则可以用于分类系统，甚至分类可能是一种更好排名的手段。大学分类的必要性尤其表现在大学职能的多样性上，不同大学培养人才的类型、规格、面向是多元的，不同大学从事科研活动的领域、内容、方式是多样化的，不同大学从事社会服务的规模、对象、方式也是多元的，再加上各个职能之间存在着有机关联，使大学的功能生态更加呈现多样性与复杂性。在支持国家发展这一点上来看，有序分工、健康协调的大学体系是必然的，也是必需的。高等学校分类是高等教育规模扩张后的必然要求，分类本身不是目的，通过分类所要达到的是实现高等教育的多样化发展。高等教育建设重点应由一流的学校转向一流的（高校）体系，以体系与社会的匹配性及体系内每所学校的卓越发展为衡量标准。

二、欧洲系统多样性理念的排名与分类实践

从整体上看，欧洲高等教育的国际竞争力在二战后落后于美国，而且还受到亚洲部分国家和地区的挑战。据全球大学学术排名（ARWU）2011年结果可知，前十名大学中欧洲仅有2所，前20名中仅有3所，前50名大学中欧洲大学仅有11所。从20世纪90年代开始，欧洲高等教育领域改革力度逐渐加大，特别是在20世纪后期至今，高等教育方面的改革逐渐从政策边缘走向政策中心。欧盟高等教育改革措施主旨在于加强大学系统的吸引力、竞争力及其支持国家创新系统的能力，在此背景下推动的高等教育机构系统多样性排名与分类更加注重院校间的横向分化与整体竞争力提升。换句话说，系统多样性理念的典型体现就是建立“世界一流（大学）系统”的排名理念，以及在此理念指导下欧盟国家正在开发的多元化评价模式，即给不同的院校图景和使命愿景施以不同的价值导向。鉴于当前缺乏跨国的可比较数据和合适的指标与度量方法，针对单一维度的“科研评价”排名的流行，欧盟成立了“基于大学基础研究的评价专家组”（Expert Group on the Assessment of University-based Research），由其负责开发多维度的欧洲排名系统——“多维度排名项目”（multi-dimensional ranking projects），以此作为对上海交通大学ARWU系统

大受欢迎的应对，重点是检验、提出和讨论评价特殊学科群、教学与学习、第三使命活动，特别是附带全球拓展（global outreach）的最合适指标。在此主要以三种与系统多样性理念相关的欧洲大学系统排名和大学分类实践为例来说明，即里斯本理事会的大学系统排名（University Systems Ranking, USR）、英国QS公司全国系统实力排名（National System Strength Rankings, NSSR）和欧盟高等教育机构分类。

（一）里斯本理事会大学系统排名（USR）

受欧盟委员会文教署（European Commission's Education, Audiovisual and Culture Executive Agency）资助和OECD指导，里斯本理事会（The Lisbon Council）于2008年推出了以国家为单位的大学系统排名（University Systems Ranking, USR），成果集中体现在报告《大学系统排名：知识时代的公民和社会》（University Systems Ranking: Citizens and Society in the Age of the Knowledge）中。USR的隐含假设是“大学系统有比批量生产诺奖得主、教授职位或‘专利教授’更广泛的任务，研究的卓越不应该成为履行教育使命乏力的借口”。USR特别看重第三级教育系统在帮助公民和社会应对21世纪知识经济挑战的能力，并没有过多关注大学或大学系统在直接促进经济社会发展中的作用。报告起草者之一Ederer在与笔者的邮件沟通中提到，“至少在快速变化的信息时代，大学不能直接对经济发展有所裨益，其基本任务应该是提升人力资本水平，而不是经济发展……学术活动至少在欧洲还没有表现出对公司创新能力及其所处环境助益良多的现象”。

USR建立了六个标准：①包容性（Inclusiveness），第三级教育系统相对其人口总量培养大量毕业生的能力；②可获得性（Access），第三级教育系统接受和帮助第二级教育中低水平学术性向的进步学生的能力，而不仅仅是只接受那些无需过多费力就可显见培养效果的学生；③效力（Effectiveness），教育系统培养适应国家劳动力市场技能需要的毕业生的能力；④吸引力（Attractiveness），国家教育系统吸引来源广泛的外国学生的能力；⑤年龄域（Age-Range），第三级教育系统扮演终身学习教育机构角色的能力，关注30-39岁“老人”进入第三级教育机构的比

例；⑥反应能力（Responsiveness），教育系统的重组和变革能力，通过审视教育系统调整以适应博洛尼亚宣言所设定标准的速度和效率来实现。这六项标准中至少有终身学习、可获得性、包容性这三个标准在较有影响的大学排名中都没有得到充分反映。以此六条标准对17个OECD国家进行排名的结果与我们印象中的所谓高等教育水平的高低大相径庭。前3名分别是澳大利亚、英国和丹麦，美国仅排在第5位，德国、奥地利和西班牙的表现最糟，分列15-17名。迄今为止，USR十分独特，它不提供单个大学如何培育拔尖毕业生的更多信息，其设计初衷是对国家第三级教育系统整体在应对21世纪知识社会的经济和社会挑战的表现做一个全球比较。当然，它主要关注的对象是30个OECD国家中能够获得分项指标数据的17个。其中，除美国和澳大利亚外，其他15个国家都是参与博洛尼亚进程的欧洲国家。USR的六项指标体现出来的一国高等教育系统的整体竞争力，是一种竞争力评价。我们注意到，在世界经济论坛（WEF）《全球竞争力报告》和瑞士洛桑国际管理发展研究院（IMD）发布的《国际竞争力年度报告》中，教育相关指标与USR“不谋而合”。如“25至34岁受过高等教育的人数比例”、“教育体制满足竞争性经济需求的程度”、“大学教育满足竞争性经济需求的程度”，等等。

在全球化时代，开放的、具有竞争力的高等教育系统应该能为广泛群体提供接受高等教育的机会，而不应成为固化社会和经济不平等的途径；能否为经济社会提供必需的劳动力并积极参与到地区经济发展中也将是重要标准；使命的多样化和稳定性的水平分化也应该是大学系统的重要特征。USR是一种真正以“创建一流（大学）系统”为基本出发点的整体性排名。如奥地利和德国之所以排名靠后，源于其“教育系统的限制性”，拒绝了大量学生，结果只有相对少数人能够接受高等教育；德国毕业生遭遇低薪资，这表明教育系统不能培养拥有适应当地劳动力市场技能要求的足够数量的毕业生。USR认为，大学系统最高的、最根本的、最重要的使命应该是培养尽可能多的公民，而不论其年龄、社会地位或以前的学业成绩，即大学系统理应是开放的、竞争性的，它们应该为最大多

数的学生提供最广泛的机会，仅仅培养少量精英是不够的。它特别强调了大学教学使命的重要性，“世界一流的研究是一个重要方面，因为它能使得部分大学培养出一流的学生，但对系统整体而言，教育使命才是压倒一切的”，“如果大学系统还想在知识社会中扮演充分的社会和经济角色的话，它们（系统）就不能在关注世界一流研究的同时忽视教育使命”。里斯本理事会认为，开发USR的目的是尽力扩展院校单一使命到其能够助益社会进步的更多方面。报告题目中“知识时代”“公民和社会”两个关键词充分说明USR的理念是关注大学系统对时代需求的反映，以及对高等教育的公共价值的关切。正是因为USR不同的排名理念和排名依据，其排名结果与我们所想象的大相径庭。如美国、德国、法国这类传统高等教育强国在诸多指标上都表现欠佳，反而是一些以原有评价标准来说并不有名的国家在部分指标上表现优异，如波兰（Access，第一）、葡萄牙（Effectiveness，第二）、瑞典（Age-Range，第三）、爱尔兰和丹麦（Responsiveness，第一和第二）。

当然，这个排名的价值取向并非无懈可击。如，它强调大学系统在回应社会迫切需求上应该积极主动，“它不应该仅仅向富人或特权阶层提供机会，而应该变成社会变革的真正的发动机，并且敞开大门帮助和教育最广泛的人们”。但不可否认，那些具有坚实的科研实力并在成果孵化中做出突出贡献的大学，正可以在推动社会发展中起到不可替代的作用。所以，我们可以把USR看成排名理念的革新或排名的替代性选择，是一种整体性的思路。如，它认为“如果大学系统培养出来的是不适合劳动力市场需要的、过度教育（over-educated）的毕业生，那就是系统没有充分履行其经济和社会职责”，这一点无论是微观上还是宏观上都是成立的。

（二）英国QS公司全国系统实力排名（NSSR）

英国泰晤士报和QS（Quacquarelli Symonds）公司曾在2004-2009年间联合发布世界大学排名（THE-QS），并成为公认的三大全球大学排名（上海交通大学、美国新闻报道、泰晤士报）之一，产生了深远影响。从2010年开始，泰晤士报和QS公司分道扬镳，分别发布自己的排名。

泰晤士报与汤森路透公司合作发布全球大学排名（THE-WUR），QS公司发布自己的世界大学排名。全国系统实力排名（National System Strength Rankings, NSSR）是由QS公司下属机构QS情报部（QS Intelligence Unit, QSIU）推出的排名中的一种，旨在开发更多的支撑性分析，以突出迎合对排名结果的不同诉求并凸显更大范围院校的独特性。QS全国系统实力评价（QS SAFE National System Strength, SAFE是System、Access、Flagship、Economic的首字母组合）是运用排名结果并结合其他指标来整体评价国家高等教育系统实力而非单个院校的相对实力的第一次尝试。目前NSSR针对40个国家进行排名。NSSR数据来源广泛而充足，基于数千份针对雇主和学者的问卷调查，得出不同国家的大学系统的强度，有助于那些期望评估其教育系统的国家参考。

QS识别出四个关键指标，运用标准化统计方法辅以相同权重（同样的标准化方法也被用到QS世界大学主排名中）。四个指标如下：①系统（System）。代表系统整体实力评估，基于某国所有院校在满足特定资格标准上的绩效。更确切地说，给定国家中排名前700或更靠前的院校数量除以那些院校的平均排位。②接纳性（Access）。广泛参与如今是一个受到热捧的高等教育议题，其首要条件就是给定国家是否能够提供充足的得到国际认可水平的大学入学机会。这个指标基于选定国家（subject country）在排名前500的大学中的位序数（特别是进入QS全球大学排行榜前500的国家的大学中的全时学生总数）除以人口规模（特别是总人口的平方根）。③佼佼者（Flagship）。在一个经费并不充裕的国家，究竟是把有限的经费投到高等教育系统许多院校身上，还是集中资助从而创建至少一所“世界一流大学”上，这在全球都是有争论的。不论是通过直接投资，还是借助于国内竞争，一国的拔尖大学的表现对其所属的高等教育系统而言都是一种光荣。这项指标表现为标准化的得分，基于所讨论国家对象中的顶尖大学的全球表现。④经济性（Economic）。并非所有国家或国内的所有高等教育机构都能平等获得资助。国家经济实力是一个主要因素，但文化影响也可能是一项影响因素，如产业界资助或校友捐赠所扮演的角

色。经济性指标识别两个关键因素，第一是给定国家政府在高等教育上的相对的财政重视程度，第二是这种投资的效果或效率——最根本的是识别出相对投资的绩效。这项指标表现为索引得分（indexed score）和所讨论国家的相对人均GDP因素。索引得分分为，有一所进入前100名为7分，往后依次为：101-200名6分，201-300名5分，301-400名4分，401-500名3分，501-600名2分，601-700名1分。按照这种评分方法和标准，2008年的评估结果中最强实力大学系统前10名的国家是：美国、英国、澳大利亚、德国、加拿大、日本、法国、荷兰、韩国和瑞典，它估计了国家大学系统之间的相对强度和弱势；同样，NSSR也招致部分批评，格里菲斯大学（Griffith University）高级研究专员Tony Sheil就说，“我不确定澳大利亚能排第三，也不认为我们的高等教育部门比排第10的瑞典和排第21的丹麦强”。

应该说，USR和NSSR都与欧盟所提出的《调动欧洲智库：让高等教育在里斯本战略中充分发挥作用》报告所标识的战略框架直接相关，如通过卓越质量增强吸引力、通过完善治理加强系统的公共责任、通过增加投入和完善投资渠道提高效益等。USR和NSSR这两个系统排名与欧盟所倡导的构建高质量的、有效的、可获取的第三级教育系统的理念高度一致。不论它们使用的指标是否具备足够的解释力，也不论指标是否最好地衡量了那些想要衡量的对象，这两种排名以及其他以“系统”为对象的排名系统都对原有的排名系统带来了冲击，推动它们不得不重新思考排名在提供指南信息中所能与所作，以满足国家和国际范围内的多样化需求。

（三）欧盟高等教育机构分类与排名

鉴于传统上以排名为代表的大学评价方式的弊端，特别是排名的方法论基础的缺陷以及对院校使命多元化的忽视，欧盟开始关注高等教育机构分类，即充分考虑多样性（Diversity）、使命（Mission）、概况（Profile）的不同来进行评价，并提出不同类型的院校评价应该基于评价的目的而使用不同方法。同时，力图使评价为公众提供更具选择性、可信的、透明的完全信息。为此，欧盟启动了多个项目：①大学绘图（U-Map）项目。开发针对具有丰富多样性高等教育机构的分类模型，使

用这个在线模型可以使得类似院校按照教育、科研、创新、地区参与、国际化等方面的比较成为可能，通过绘制大学活动地图描绘大学多样性，从而达到分类的目的。②多元化排名（U-Multirank）项目。这个可行性项目试图找到测量欧盟内外高校和研究机构绩效的透明性的办法，开发多元的、独立于政府和大学的排名系统，维度与U-Map类同。项目的目的是评估大学扮演不同角色的好坏，而非让所有大学都坚持单一的学术导向标准。当然，也是基于各种使命，如教学、科研、创新、地区参与和国际化。③大学科研评价（Assessment of university research）项目。欧委会邀请学者、非政府组织和教育专家组成专家组，探究如何在充分考虑不同大学、学科和使用者（大学、资助机构、研究人员和私人部门）的立场的基础上创建多维度评价。④欧洲大学数据收集（EC European University Data Collection）项目。该项目旨在建立完整的欧洲大学的统计普查，支持欧盟统计局（Eurostat）建立高等教育统计数据库（statistical database on Higher Education），已经启动了试行的数据收集，包括所有欧洲大学的基本数据和科研活跃的多所样本大学的补充性数据，探索在院校层面建立一个连续性的、透明性的欧洲统计基础设施的可能性，以为分类提供充分依据。⑤高等教育学习成果评价（Assessment of Higher Education Learning Outcomes, AHELO）。该项目测试高校学生知道什么以及毕业后能做什么，用来评价学生学习成果，类似OECD推出的国际学生评估项目（Programme for International Student Assessment, PISA），工程和经济领域中的通用技能和学科技能。

这五个项目之间具有紧密联系，都涵盖在欧洲高等教育机构分类（Classifying European Institutions for Higher Education, CEIHE）之下。其分类目的是：便利高等教育方面的科研，为利益相关者提供透明性信息，为政策制定者提供重要依据（特别是避免“一刀切”式的方法），为院校塑造自我（使命、框架、战略）提供帮助，用于排名。从根本上说，CEIHE项目与博洛尼亚进程和里斯本战略（Lisbon strategy）直接相关，是为了支持扩大学生流动、增强院校间和产学研间的合作、提高欧洲高等教育的竞争力、鼓励高等

教育发展的一体化和多样性。分类的目标是“提高复杂系统的透明性，并抓住系统的多样性，从而改善我们对高等教育现象和大学系统的认知，进而促进利益相关者之间的有效沟通”。CEIHE的最终目的是制定一套多维度全球大学排名（multi-dimensional global university ranking），即U-Multirank，排名提供有关教学和研究质量多样性的信息，但是质量的界定必须考虑多样化高等教育机构之间的差异。多样性是新的分类系统的主要特点，也是欧洲高校分类的灵魂。多样性之所以受到如此重视，是因为高等教育多样性能够满足学生入学需求、促进社会流动、迎合劳动力市场多样性要求、提高院校运作效率、扩大创新成功的可能性、保障欧洲的区域竞争力等等。

CEIHE始于2004年，分三个阶段实施。CEIHE I（2004-2006），主要在于通过与学生、高校、工商界和政府的试探性的互动讨论，确立作为达成一个分类体系的基础的一套基本的设计原则和第一组主题和指标（indicator），最终于2005年发布了第一阶段报告《院校概况：开展欧洲高校分类》（Institutional Profiles: Towards a Typology of Higher Education Institutions in Europe）。CEIHE II项目是CEIHE项目的后续行动，始于2006年9月，2008年5月结束，最终报告为《描述多样性：欧洲高校分类》（Mapping Diversity: Developing a European Classification of Higher Education Institutions）。主要任务是：现有欧洲数据源分析（测试第一阶段的主题和指标），不同类型院校的深度案例分析，与利益相关者的咨询、会议、研讨沟通，调查以评估相关性、有效性、可靠性、可行性，提高分类的合理性，起草分类类型（typology），提供运行实施的建议。改编后，分类框架包括五个主题（scheme）、14项指标（indicator）或维度（dimension）。①教育（Education）：最高学位授予，即每个学历层次上的学位/文凭数；科目组合（Subject mix）；专业导向（Orientation of programmes）；终身学习（LLL）参与，即30岁以上学生占总注册学生人数的比例。②研究与创新（Research & innovation）：研究密集度，包括每个学术职员同行评议出版物、官方统计的

科学计量学“拔尖”指标；创新密集度，包括私人资助研究经费数量占总财政规模的比例、初创公司数、申报专利数、专利收入。③国际化导向（International orientation）：教学和职员，包括授予国际学位比例、吸纳的海外/欧洲交流学生比例、送出的海外/欧洲交流学生比例、所提供的国际联合培养计划占总计划的比例、在海外提供的计划数、全时国际学术职员比例；科研，欧盟提供科研计划经费收入占科研经费总额的比例。④规模与情境（size & settings）：规模，学生总数（每个学位层次）、全时学术职员总数、每年可运作的财政总额；教学模式（Mode of delivery），远程学习计划比、非全日制学习计划比；公共/私人特征，政府来源收入占总收入的比例；法律地位（Legal status，公共的82%，私人的18%）。⑤社区参与（Community engagement）：文化参与，音乐会数、展览会数；地区参与（Regional engagement），地方毕业生、欧盟结构化资金收入数、为地方开设的课外课程、地方收入的重要性。CEIHE III的项目预期成果是“大学地图”（U-map），评估和微调维度及其指标，并使其与其他相关指标设定联系起来；最终确定一个正常运行的在线分类工具；使其与卡内基基金会运行的分类工具衔接起来；继续开展作为该项目标志的利益相关者的有关咨询和讨论。U-map必须是具有国际应用价值的，多元维度的，包容性强，描述性和可靠的在线工具：能够使欧洲高等教育的多样性透明可见；能够为利益相关者提供相关信息；有助于提升欧洲高等教育在知识生产和知识利用中的国际竞争力。目前，U-map项目已经初始运行（网址：<http://www.u-map.eu/>）。欧盟希望在2013年年中可以实现U-map更广泛的公开应用，并进一步扩大项目覆盖的院校数量。

相对卡内基分类的成功，欧洲至今没有可操作性的高等教育机构分类体系，正在开发的多维度欧洲分类（multi-dimensional European classification）系统框架可望弥补这一缺憾。在线方式是其重要创新之一。在线方式首先是吸纳了利益相关者对分类或排名的诉求，体现了排名中互动式的趋向，承认了目标受众之间偏好的异

质性，体现了“多样性”的另一种内涵。同时，欧洲高等教育机构分类也是迄今为止世界范围内较为完善的分类，考虑到欧洲各国教育传统、地域文化、经济发展方面的巨大差异，分类系统对多样性特征的侧重尤其值得我们关注。从分类结果来看，客观指标涵盖了教育教学（学位类型、学科专业领域）、科学研究（论文及其引用）、创新与转化（专利、经费收入、孵化企业）、终身学习、国际化程度（教师、学生、国际项目、国际化专业）、高校属性（公/私立、法律地位）、社会服务（对本地的贡献、对文化传播的贡献）、院校规模（师生数量）等等。值得注意的是，该分类系统不但没有把科学研究的地位放得很重，而且很多客观指标都体现了欧洲特色，如参与欧洲交换项目的新生占总的新生的比例和参与欧洲交换项目的学生总数；特别是与卡内基分类相比，欧洲分类更加注重高校的社会服务功能。维度选择上，U-Map本意是想涵盖其所关注到的排名及其院校分类的所有维度，也曾考虑方便用户使用而力图减少到5-8个维度，即略微多于卡内基分类的维度数量，但后来考虑到欧洲内外高等教育和研究的天然复杂性会带来比美国高等教育本身更程度的“必要变异”（requisite variety），从而保留了14个维度的办法。还有一点需要注意的是，U-Multirank由作为独立第三方的荷兰特文特大学（University of Twente）高等教育政策中心（CHEPS）和德国高等教育发展中心（CHE）联合实施，保障了分类的相对公正。

三、评价与借鉴意义

从以上两个系统排名和一个机构分类实践来看，系统多样性理念下的欧洲大学排名与分类理念和传统的大学排名与分类之间具有明显差异：前者注重整体评价，注重系统竞争力与公共价值，后者侧重单体位序，关注大学自身发展。表1归纳了传统大学排名和分类与系统多样性视角下的大学排名与分类的主要区别。可以看出，传统做法难以适应知识经济时代对高等教育系统多样性的要求，也产生了诸多值得怀疑的集中型政策，如德国卓越计划、中国“985工程”。系统多样性理念下的欧洲大学排名与分类实践为我们规划高等教育发展战略和改进大学排名

与分类实践提供了新思路，即以系统提升为宗旨，而非创建“科研评价导向”下的数所一流大学。正如OECD所指出的那样，“一个整体上强大的大学系统并不能仅由研究卓越所实现，而根本上应符合社会的需求（demands）和需要（needs）”。这对于扭转汲汲于研究一流的所谓“研究型大学”建设而言是一种醍醐灌顶般的警示。其实，从实践来看，较早的美国加州高等教育分类和卡内基分类及其多次改版已经体现出了系统多样性的旨趣，虽然前者属于规定性分类，旨在管理，后者属于描述性分类，旨在研究，但都没有没落入“研究导向”的俗套。卡内基分类2000年版开始，逐步引入多样化分类系统，扭转了过分注重所谓科研的倾向（如联邦资助经费指标、博士学位授予数指标）。2005年版分类更加基于多维度，教什么（what is taught）、为谁而设（to whom）、在何背景（in what setting）成为新分类的考虑对象，而且不单有综合性分类体系，还引入了多个选择性分类体系，相互补充。

在排名成为一种社会认知习惯的时候，取消排名是不现实的，但应当适度引导排名，使得排名成为引导高等教育多样化发展和高等教育机构特色化发展的推动力。建立多元化、多维度排名，而不是简单院校综合排名成为不二选择。值得关注的是，部分排名系统现在已经在朝上述所要求的方向转变，德国CHE排名已经开发出了学科排名（subject-based rankings），上海交通大学除了有学科排名，还有领域（field-based rankings）排名，可以视为是对原有单一排名的修正。我国大学分类、排名、评价对多样性的关注目前还基本上停留在单一化阶段，表现在政策导向上就是言必称哈佛、耶鲁、MIT，不关注系统本身，更不关注经济社会背景。殊不知，即使同是所谓“世界一流大学”，其类型、使命都有所侧重，MIT、斯坦福与哈佛、耶鲁就是俨然不同的两类大学，前者创业特征明显，后者学术能力凸出。在推动我国高等教育机构评价中，欧洲有关系统多样性排名与分类的理念与实践非常值得借鉴。

还需要注意的是，在多样性排名与分类理念中，与社会相联系的“多样性要素”，即大学或大学系统与社会相联系的表现，得到了足够重视，如卡内基分类2010年推出的“社区参与的

表1 传统排名和系统多样性排名与分类的主要区别

	传统排名与分类	系统多样性排名与分类
自趣导向	体现大学单体竞争，关注大学或高等教育自身发展	系统和谐，整体竞争力提升，关注社会效益与经济价值，特别是人力资本价值
测度依据	研究导向，数量至上	教育、研究和社会参与的全谱(full spectrum)特征
指标设定	排名主体主观确定	吸纳各利益相关者诉求
测度方法	来自数据库的客观性定量方法，辅以少量定性方法	定量与定性相结合，强化结果的问责和透明
呈现形式	单一数值，固定结果，绝对绩效	多元维度，关注用户个性化需求，注重相对表现
主要特点	“后台操作”，结果导向，难以比较，尤其不能解决信息不对称的问题	透明性（指标、过程与结果公开）、描述性（提供丰富信息）、便于比较
政策理念	资源集中化战略，重点打造若干“世界一流大学”	强调系统功能，推动形成“世界一流（大学）系统”
客观效果	刺激规模发展、“研究型”及所谓精英教育与精英大学，体现纵向分化	有助于实现横向多元分化，鼓励特色发展，特别是认可、激励和奖励多元“研究”、多种活动

备选类（别）分类”（Elective Classification on Community Engagement, ECCE）设置了三大类别，即课程服务（Curricular Engagement）、拓展合作（Outreach and Partnerships）、课程服务与拓展合作。USR的基本理念就是大学系统要适应经济社会发展的需要。再如泰晤士报的最新排名（THE-WUR 2010-2011）指标体系中，“师均工业来源研究收入”也被提了出来，占比2.5%。上文所述的欧盟的几项实践中也出现了“社区参与”（Community engagement）这样的维度和指

标，这充分说明这种“创业特征”的重要性。但值得期待和进一步讨论的是，更多的这种多样性要素，与论文引用这种确定型指标相比，如何进行描述和量化，似乎是一个比较难解决的问题。

（吴伟，浙江大学科教发展战略研究中心博士研究生，浙江杭州 310027；邹晓东，浙江大学科教发展战略研究中心研究员，浙江杭州 310027；范惠明，浙江大学科教发展战略研究中心博士研究生，浙江杭州 310027）

（原文刊载于《中国高教研究》2012年第7期）

《美国新闻与世界报道》

本科院校排行榜评估指标体系研究

李函颖 刘宝存

大学排行榜的核心组成部分是它的排名方法，科学的排名方法是保证排名结果真实、可信的重要前提。《美国新闻与世界报道》（U.S. News & World Report，以下简称《美新》）杂志自1983年起对美国大学进行两年一次的排行，即“最佳本科院校排行榜”（Best Colleges），1987年起改为每年一次。《美新》对本科教育和研究生教育进行排行时使用的是两套不同的排名方法。研究生院的排名基本依赖声誉调查，比较简单；而本科院校的排名方法则已经发展成为一套完整的排名体系，注重主观指标与客观指标的结合，更具研究意义。

评估指标体系是大学排行榜的核心内容，它不仅是直接影响最终排名结果的计分方法，而且也是排行榜本身是否具备公信力的衡量标准之一。评估指标体系包括评估指标和权重系数两个部分，其中评估指标是指排行机构认为能够体现一所学校质量的关键因素，它直接反映着排行机构的排行理念。而权重系数则是排行机构依据相关教育研究成果和高等教育学者意见，以其对教育质量的重要程度为标准，赋予每一项评估指标的比例。一般而言，对大学质量影响越大的评估指标的权重也越高，反之亦然。所有评估指标的权重系数的总和为100%。评估指标体系任何细微的变化都会导致排名结果的巨大变化。例如，布林莫尔学院（Bryn Mawr College）在1989年的排行榜中名列第5，却在1990年的排行榜中骤降至第23位。

一、《美新》评估指标体系的历史变革

《美新》本科院校排行榜在诞生之初并不具备如此完整的评估指标体系，其最初三期的排行榜是以大学校长的主观调查为依据，并不包括

客观评估指标。直至1988年，《美新》才正式推出兼具主观评价和客观指标的评估指标体系。此后，《美新》评估指标体系的多次调整均是在1988年的大框架下进行的。

1988年《美新》本科院校排行榜第1版评估指标体系中的一级指标只有“学术声誉”、“新生录取”、“教师资源”、“经费资源”、“毕业率”5项，而且没有对不同类型院校使用的评估指标体系加以区分。1991-1992年，“新生保持率”和“毕业率”都被列为二级评估指标，1993年《美新》将“新生保持率”从评估指标体系中删去，单纯强调学生毕业状况，将“毕业率”重新列为一级指标，且权重占到10%。同时，其还将“教师资源”与“经费资源”两项指标的权重各减少5%。同年，《美新》还在评估指标体系中增添了“校友满意度”（alumni satisfaction）一项评估指标，其主要是指在世的校友在近两年时间内为母校捐款的比例，该项评估指标的权重是5%。虽然这一评估指标代表校友满意度的有效性受到了诸方质疑，但此举至少促使高等学校进一步提升了校友联络的通畅度，也在一定程度上增加了校友捐赠的金额。

由于许多大学校长经常抱怨《美新》对输入性指标如“新生录取”等评估指标赋予过多的权重，他们认为输入性指标仅仅能够反映大学所录取的学生有多聪明，而无法反映出他们所接受的高等教育的质量。鉴于此，《美新》1995年对其排名方法进行了重大调整。在缺乏足够论证的前提下，《美新》决定将“毕业率”作为“新生保持率”（freshmen retention rates）的二级指标，并将“新生保持率”这一项一级指标的比重增至25%。同时，其也将“新生录取”的权重从原来

的25%降至15%。

1996年,《美新》杂志将“增值”评估指标(value-added)作为一级评估指标引入评估体系,并赋予其5%的权重。在这一年的本科院校排名中,《美新》建构了一套数学模型,根据每所大学的输入性指标,用这套模型来预测这所学校“应有的毕业率”。预测模型的基本理念是:接受捐赠越多的本科院校能够吸引更多优质生源(SAT或ACT分数很高),生均教育经费更高,因而其学生毕业率也更高。“增值”指标衡量的就是学校“应有的毕业率”与现实毕业率之间的差距。一所学校的现实毕业率越高于预测毕业率,这所学校在排行榜上的名次就会越高。此后,尽管数据统计方法有所改变,但“增值”指标的基本概念仍保持不变,其后被更名为“毕业率绩效”(graduation rate performance)。此外,1996年的评估指标体系中还增加了一项二级评估指标,即“班级规模超过50人的班级比例”(class size, 50+students)。

从1998年开始,《美新》本科院校排行榜的评估指标体系的一级指标相对固定,权重保持不变,只对二级指标或更细微的评估指标稍作调整,以适应高等教育的发展。例如,2003年《美新》杂志开始停止使用“被录取学生的注册率”(yield)这项评估指标,这项指标此前在整套评估指标体系中所占权重为1.5%。长期以来,“被录取学生的注册率”评估指标一直是公众和教育家们争论的焦点之一。一部分人认为由于“被录取学生的注册率”能够在一定程度上影响大学的排名,直接导致许多高校开始加大提前录取学生的比例。此举虽然有助于提高高等学校的排名,却过早地剥夺了学生择校的权力。为了平息外界争论,加之“被录取学生的注册率”评估指标所占权重很低不足以影响整套评估体系,《美新》杂志决定停止使用这项评估指标。2010年起将“高中升学咨询教师调查”纳入评估指标体系。由于高中升学咨询教师不仅对本地区及全国的高等学校都十分了解,而且他们也会在高中生选择高等院校的过程中起到关键性的辅助作用。因此,2010年《美新》本科院校排行榜中的“全国性

大学”以及“全国性文理学院”两项排行榜都将“高中升学咨询教师调查”纳入评估指标体系。在这两项排行榜中,“本科学术声誉”的权重由25%降至22.5%,其中“大学同行评议调查”的权重由25%降至15%，“高中升学咨询教师调查”的权重则占7.5%。

二、《美新》现行评估指标体系

2012年9月,《美新》杂志如期公布了最新一期的“2013年最佳本科院校排行榜”(Best College 2013)。最新一期排行榜的评估指标体系由7项一级指标和16项二级指标组成。其中,一级指标包括“本科学术声誉”、“毕业率及新生保持率”、“教师资源”、“新生录取标准”、“经费资源”、“校友捐赠”、“毕业率改进绩效”(仅限全国性大学和全国性文理学院两项排行榜)。单就全国性大学和全国性文理学院两项排行榜而言,其包括“毕业率改进绩效”、“高中升学咨询教师调查”两项特有的评估指标(详见表1)。

《美新》对其各项评估指标都进行了详细说明。

1. “本科学术声誉”指标。这项评估指标由“大学同行评议调查”和“高中升学咨询教师调查”两项二级指标组成。其中,“大学同行评议调查”是以同一类型大学的校长、教务长以及招生主任为调查对象。同行评议得分越高,学校在排行榜中的得分也会越高。“高中升学咨询教师调查”是以公立高中升学咨询教师为调查对象,每位受访者依据自己对大学的学术成就、课程、师资、专业、毕业生等方面的认知给大学进行评分。如果一所大学获得较高的高中升学咨询教师得分,那么它在排行榜中的得分也会较高。

2. “新生录取的竞争性”指标。一所学校的学术氛围在一定程度上受到学校学生的能力和理想的影响。这项评估指标包括“录取率”、“高中成绩班级排名”、“SAT/ACT分数”这3项二级指标。“录取率”是指2011年秋季实际录取学生总数与申请入学学生总数之比。一所大学的录取率越低,其在排行系统中的得分就越高,反之亦然。“高中成绩的班级排名”是指2011年秋季入学的大学新生中高中成绩优秀的学生所占的比例。对于“优秀”的定义各项排行榜有所不同。在全国性大学和全国性文理学院两项排行榜中,

优秀新生是指其曾是所在高中班级的前10%，而在地区性大学和地区性学院排行榜中，优秀新生则是指其曾是所在高中班级的前25%。优秀新生越多的大学其得分也越高。“SAT/ACT分数”是指2011年秋季入学的大一新生在SAT和ACT考试中所获得的平均分，其中SAT考试只统计批判性阅读和数学两门科目的分数。学生在SAT和ACT考试中的得分越高，他们所进入的大学在排行榜中所获得的分数也就越高。

3. “教师资源”指标。研究发现，学生越满意与教师的互动，他们就越愿意主动学习，也有可能顺利毕业。2011-2012学年，《美新》杂志用了6项指标来评价一所学校的教师资源。其中，“教师薪酬”是指2010-2011学年以及2011-2012学年各校教师所获得的薪酬与福利的平均值，调

表1： 2012年《美新》本科院校排行榜评估指标体系

一级指标	一级指标权重		二级指标	二级指标权重	
	全国性大学排行榜、全国性文理学院排行榜	地区性大学排行榜、地区性文理学院排行榜		全国性大学排行榜、全国性文理学院排行榜	地区性大学排行榜、地区性文理学院排行榜
本科学术声誉 (undergraduate academic reputation)	22.5%	25%	同行评议调查 (peer assessment survey)	66.7%	100%
			高中升学咨询教师调查 (high school counselor's rating)	33.3%	—
2011年秋季新生录取的竞争性 (student selectivity for fall 2011 entering class)	15%	15%	录取率 (acceptance rate)	10%	10%
			高中成绩班级排名前10% (high school class standing in top 10%)	40%	—
			高中成绩班级排名前25% (high school class standing in top 25%)	—	40%
			SAT和ACT考试中的批判性阅读和数学成绩 (critical reading and math portions of the SAT and the composite ACT scores)	50%	50%

2011-2012学年教师资源 (faculty resources for 2011-2012 academic year)	20%	20%	教师薪酬 (faculty compensation)	35%	35%
			获得本专业最终学位的教师比例 (percent faculty with top terminal degree in their field)	15%	15%
			全职教师比例 (percent faculty that is full time)	5%	5%
			师生比例 (student/faculty ratio)	5%	5%
			班级规模, 1-19人 (class size, 1-19 students)	30%	30%
			班级规模, 50人以上 (class size, 50+ students)	10%	10%
毕业率及保持率 (graduation and retention rates)	20%	25%	平均毕业率 (average graduation rate)	80%	80%
			新生平均保持率 (average freshman retention rate)	20%	20%
经费资源 (financial resources)	10%	10%	生均经费资源 (financial resources per student)	100%	100%
校友捐赠	5%	5%	平均校友捐赠率 (average)	100%	100%
毕业率改进绩效 (graduation rate performance)	7.5%	—	毕业率改进绩效 (graduation rate performance)	100%	—
总计	100%	100%	—	100%	100%

查对象包括全职的助理教授、副教授和教授。在进行比较前，《美新》根据各地区生活成本差异指数对各校的教师薪酬进行了相应调整，增强了数据的可比性。学校发放的教师补偿越高，其在排行榜中的得分也就越多。“获得本专业最终学位的教师比例”是指2011-2012学年，学校全职教师中获得本专业最终学位的人数比例（多数专业的最终学位是博士，也有少数专业的最终学位是硕士）。获得高学历的教师比例越高，学校所能获得的得分也就越高。“全职教师的比例”是指2011-2012学年在校教师中全职教师的比例。此处的“教师”不包括教授临床医学的专职医生、

教导主任、图书馆管理员、注册主管、教练等管理人员、学生助教、停薪留职的教师以及代课教师。全职教师比例越高，学校的得分越高。“生师比”是指2011年秋季在校学生总数（全日制学生）与在校教师总数（全职教师）的比例。法学院、医学院、商学院以及其它单独排行的研究生专业或专业课程的教师和学生人数不包括在内，因为在这些机构中教师基本只教授研究生，很少给本科生上课。此外，教师人数还不包括助教。生师比越低的学校得分越高。“班级规模（1-19人）”是指2011年秋季注册时，班级人数不超过20人的本科班级数目与本科班级总数之比。一所学校拥有的小型班级越多，其得分越高。“班级规模（50人以上）”是指2011年秋季注册时，班级人数超过50人的本科班级数目与本科班级总数之比。一所学校拥有的巨型班级越多，其得分越低。

4. “毕业率及新生保持率”指标。这项指标是由“平均毕业率”和“平均新生保持率”两项二级指标组成。其中，“平均新生保持率”是指2007年秋季至2010年秋季的3年时间里大一新生在第二年秋季回到同一所大学继续就读的比例的平均值。一所大学的平均新生保持率高，其在评估体系中就能获得更高的分数。“平均毕业率”是指2002年秋季至2005年秋季期间入学、能如期在6年以内顺利毕业的学生比例的平均值（转学生不计算在内）。平均毕业率越高的大学能获得更高的得分。

5. “经费资源”指标。“生均经费资源”是指2010年、2011年学校用于所有全日制在校学生的平均费用，涵盖教学、科研、公共服务、学术支持、学生服务、制度的支持、使用与维护（此项仅限于公立学校）等领域的支出。需要注意的是，这里所统计的“学生总数”既包括研究生也包括本科生。学校的生均经费越高其所获得的分数也越高。

6. “校友捐赠”指标。“平均校友捐赠率”是指在2009-2010学年以及2010-2011学年中向母校捐赠的校友人数与同时段学校记录在册的校友总人数之比的平均值。这里所统计的“校友”是指在学校完成本科教育的毕业生，不包括仅在校接受研究生教育的毕业生。校友捐赠率一般被用来衡量学生对学校的满意度，校友捐赠率越高的学校在排行榜中就能获得更多的得分。

7. “毕业率改进绩效”指标。“毕业率改进绩效”是指学生的预计毕业率与实际毕业率之间的差距。这项评估指标仅出现在全国性大学和全国性文理学院两项排行榜中，其所占权重为7.5%。这项评估指标是用来测评学校的专业教育和相关政策对于学生毕业率的影响。《美新》杂志将2005年入学、经过6年大学学习于2011年毕业的学生作为研究对象，将6年前根据学生及学校质量预测的毕业率与现在真实的毕业率进行比较，如果真实毕业率更高，则表示这所学校的教育质量正在提升，其最终得分也就越高。

三、《美新》评估指标体系的特点

《美新》本科院校排行榜的评估指标体系经过近30年的发展，已日渐成熟。纵观其20多年来的变化（详见表2），我们可以发现《美新》评估指标体系在发展中所形成的特点。

表2：1991-2012年《美新》本科院校排行榜的一级评估指标及其权重

一级 指标 年份	学术声誉	新生 录取 竞争 性	教 师 资 源	经 费 资 源	校友捐赠 (校友满 意度)	毕 业 率	新 生 保 持 率	毕业率改 进绩效
1991	25%	25%	25%	20%	5%	—	—	—
1992	25%	25%	25%	18%	7%	—	—	—
1993	25%	25%	20%	15%	5%	10%	—	—
1994	25%	25%	20%	10%	5%	15%	—	—
1995	25%	15%	20%	10%	5%	—	25%	—
1996	25%	15%	20%	10%	5%	—	20%	5%
1997	25%	15%	20%	10%	5%	—	20%	5%
1998-20 09	25%	15%	20%	10%	5%	20% (全国) 25% (地方)	—	5% (全 国); 0% (地方)
2010-20 12	22.5% (全 国) 25% (地 方)	15%	20%	10%	5%	20% (全国) 25% (地方)	—	7.5% (全 国); 0% (地方)

1. 评估指标逐渐增多。《美新》最佳本科院校排行榜的评估指标体系会依据高等教育评估领域的最新研究成果和《美新》杂志排行榜的研究人员的价值判断而做出相应的调整。在1992年以前，《美新》评估指标体系的一级指标数是5项。1993-1995年间，一级评估指标的具体项目并不一致，但至少都达到6项。1996年引入的“毕业率改进绩效”使《美新》评估指标体系基本达到了7项一级指标框架，并一直沿用至今。

2. “学术声誉”指标的权重颇高。在《美新》评估指标体系20年的发展过程中，“学术声誉”指标的地位十分独特。它的权重系数不仅一直保持稳定（25%），而且还在所有一级评估指标中高居首位。虽然《美新》杂志在2010年对它进行了调整，即两项全国性高校排行榜的“学术声誉”指标的比重下调为22.5%，其他8项地方性高校排行榜的“学术声誉”比重仍为25%，但仍然保证了其在评估指标体系中的重要地位。

3. 逐渐提高输出性指标的权重。表2中清晰地显示了20年来《美新》的输入性指标的比重在下降。如，1991年的“新生录取”权重为25%，1995年起被下调至15%；1991年的“教师资源”权重为25%，1993年起被下调至20%；1991年的“经费资源”权重为20%，1992年、1993年中分别被调至18%和15%，直至1994年最终被稳定为10%。与此同时，输出性指标的权重却在逐步提高。例如，“毕业与新生保持率”起初并未被纳入评估指标体系中，而1993年添加的“毕业与新生保持率”权重为10%，到1994年时权重被提升至15%。1995年虽然《美新》以“学生保持率”取代了“毕业与新生保持率”，却赋予其25%的高权重。1998年起，这两项评估指标被合并成“毕业与新生保持率”成为一项一级指标，但权重系数仍然非常高。“毕业与新生保持率”自1996年被引入评估指标体系以来虽然权重仅为5%，但其在2010年也被提升至7.5%。《美新》评估指标体系愈加重视输出性指标的主要原因在于：越来越多的高等教育研究者指出，输出性指标才能反映一所高等学校的真实教育水平，输入性指标最多只能反映高校能够以社会名气、高薪等因素吸引优质生源和顶尖教师队伍的情况。而顶尖学者与本科教学质量的相关程度也还存在质疑，因为并非所有的优秀学者都愿意真正参与到本科教学中去。

4. 重视分类排行。虽然《美新》本科院校排行榜的具体评估指标和权重系数是学者们关注的重点，但其评估指标体系中另一个不得不提及的重要特色就是分类排行。美国拥有世界上最庞杂的高等教育系统，只有对同一种类型的高等学校进行比较，才能使排行结果更具信度。《美新》信奉“分类排行”，主张对同

一类型的高等学校进行比较。具体来说，《美新》杂志是以卡内基教学促进基金会的高等学校分类标准为依据，将全美高等学校分为全国性大学（National Universities）、全国性文理学院（National Liberal Arts Colleges）、地区性大学（Regional Universities）、地区性学院（Regional Colleges）。其中，地区性大学和地区性学院进一步按照学校地理区域分别分为北部、南部、中西部和西部四大子项（具体对照信息参见表3）。全国性大学是指有资格为学生授予所有学科的学士、硕士及博士学位的大学，这类大学一般具备很强的科研能力。全国性文理学院的重点主要放在本科教育上，它所授予的学位至少有50%集中在文理学科。地区性大学在本科教育阶段所开设的学科数目跟全国性大学一样，但是不同之处在于其只有一部分学科可以授予硕士学位，极少数学科有资格授予博士学位。地区性学院的办学重点也集中在本科教育，但这类学校在文理学科授予学位的比例一般低于50%。因此，《美新》本科院校排行榜向人们标榜的排行理念是，排名应该在同一类型的高校之间进行才有意义。

表3：2005版卡内基大学分类与2012年《美新》本科院校分类对照表

2005 版卡内基大学分类		2012 年《美新》本科院校分类	
一级分类	二级分类	一级分类	二级分类
博士学位授予院校 (Doctorate-granting Universities)	科研能力很强的研究型大学 (Research Universities-very high research activity)、科研能力强的研究型大学 (Research Universities-high research activity)、普通研究型大学 (Doctoral/Research Universities)	全国性大学	—
硕士学位授予院校 (Master's Colleges and Universities)	大型硕士学位授予院校 (Master's Colleges and Universities-larger programs)、中型硕士学位授予院校 (Master's Colleges and Universities-medium programs)、小型硕士学位授予院校 (Master's Colleges and Universities-smaller programs)	全国性文理学院	—

学士学位授予院校 (Baccalaureate Colleges)	授予学士学位的文理学院 (Baccalaureate Colleges-Arts and Sciences)	地区性大学	北部 (North)、南部 (South)、中西部 (Midwest)、西部 (West)
学士学位授予院校 (Baccalaureate Colleges)	多学科学士学位授予院校 (Baccalaureate Colleges-Diverse Fields)、 学士/副学士学位授予院校 (Baccalaureate/Associate Colleges)	地区性学院	北部 (North)、南部 (South)、中西部 (Midwest)、西部 (West)

四、结语

《美新》最佳本科院校排行榜作为世界大学排行榜历史上重要的里程碑，经过多年的调整，已逐渐建立了一套较为成熟的评估指标体系。其建立主客观多种评估指标和权重评价大学、输入性指标和输出性相结合并逐渐提高输出性指标的权重、重视分类排行等做法，把大学排行建立在相对客观、全面评价大学的基础上，在一定程度上提高了排行的准确性和可信度，同时也有利于促进高校分类发展，注重提高教育质量，办出特色。实际上，本科院校排行榜存在的本身还可以起到纠正高等学校特别是研究型大学普遍存在的重视研究生教育轻视本科生教育、重科研轻教学的弊端，提醒人们重新审视本科生教育，努力提高本科生教育质量。

需要指出的是，《美新》本科院校排行榜的评估指标体系并不完美，其在发展中仍然受到多方指责。其中，《美新》频繁变动评估指标一举最受诟病。高校管理者们认为此举致使他们所在学校的名次也随之频繁地上升或下降，尤其是对于那些本身质量并未发生变化，却纯粹因为评估

指标的变化而引起排名变化的学校而言是十分不公平的。因为一般的社会大众不会花太多的精力去研究排行指标的变化，在他们眼中学校在排行榜中的名次直接与学校的质量对等。如1996年以前“新生保持率”在《美新》评估指标体系中的权重是15%，而编辑们在1996年时将其权重提升至25%，却又在1997年将其降至20%。如此频繁的变动不仅引起高校管理者的不满，甚至连《美新》杂志自己委任的全国民意调查中心（National Opinion Research Center, NORC）都对其排名方法颇有微词。1997年，NORC的评估报告指出：

“《美新》本科院校排行榜目前最主要的问题就是用于联接各项评估指标的权重分布缺乏理论基础。这并不意味着现在所使用的权重就一定是错误的，只是表明目前的这套权重分布是《美新》杂志的工作人员所能做出的最好判断，但是很难与任何理论基础或经验性基础产生联系。”此外，还有许多研究者尖锐地指出“这是经济因素在作祟”。教育是一项长效事业，很多改革举措的成效并不是短时间内可以显现的。然而，《美新》排行榜作为一件商品，如果其连续两年的排名结果完全保持不变的话，消费者根本就没有购买的必要。也就是说，如果排行榜的结果不发生变化，《美新》本科院校排行榜的销售量就会急剧下降，这将会对《美新》杂志造成严重的经济损失。一部分学者强调，《美新》本科院校排行榜会因销售量而调整评估指标体系，这种会“为五斗米而折腰”的商业游戏根本不具备对高等学校进行排名的资格。因此，人们在借鉴或使用《美新》评估指标体系时应正视其所存在的问题。

（李函颖，教育部人文社会科学重点研究基地北京师范大学国际与比较教育研究院博士研究生，北京 100875；刘宝存，教育部人文社会科学重点研究基地北京师范大学国际与比较教育研究院院长、教授，北京 100875）

（原文刊载于《中国高教研究》2013年第11期）

QS大学排名体系剖析

刘 强 丁瑞常

2004年,英国《泰晤士报高等教育增刊》(Times Higher Education Supplement, THE)联合国际高等教育资讯机构QS(Quacquarelli Symonds)推出了“THE-QS世界大学排名”(THE-QS World University Rankings),与我国上海交通大学2003年发布的“世界大学学术排名”(Academic Ranking of World Universities, ARWU)并肩成为两大具有世界影响的全球性大学排名项目。2010年, QS公司改与《美国新闻与世界报道》、英国《周日时报》、韩国《朝鲜日报》合作,以THE-QS排名为基础研发了“QS世界大学排名”(QS World University Rankings)系统。之后根据市场需要逐步建立起一个以其世界大学排名为核心的跨国型大学排名系统,其中包括QS世界大学排名、2009年发布的QS亚洲大学排名(QS University Rankings: Asia)、2011年开发的QS拉丁美洲大学排名(QS University Rankings: Latin America)以及QS世界大学学科排名(QS World University Rankings by Subject)和2012年启动的QS年轻大学50强(QS Top 50 Under 50)项目。本文将以此四大跨国型大学排名项目为研究对象,对QS大学排名体系的对象、原理、方法做一个全面的剖析,旨在为高校正确认识QS大学排名体系、有针对性地制定战略规划提供参考。

一、QS大学排名体系的评价对象

在QS世界大学排名中,全球所有符合QS公司基本要求的大学均可申请参与排名。这些基本要求包括:参与排名的大学至少在“艺术与人文”、“工程与IT技术”、“生命科学”、“自然科学”以及“社会科学”中的一个学科领域开设了本科、研究生项目(联合培养项目不计入其中);大学能够提供师生规模、学费方面的准确数据,以及相关自述材料以论证自己有资格参与排名,比如提供该大学在地方大学排行中的名次。然后, QS公司根据申请名单,同时参考全球学术调查及雇主调查的反馈结果并适当兼顾地域

均衡等因素,最终选定参与世界大学排名的大学名单。

按照社会认同理论,“个体需要通过实现或维持积极的社会认同(social identity)来提高自尊,而积极的自尊来源于内群体与相关的外群体的有利比较”。因此,大多高校都本能地希望跻身世界大学排名,从而对QS世界大学排名的发展提出了扩大评价对象规模的市场需求。为此, QS公司多次扩大世界大学排名规模,如今能进入其世界大学排名的高校已扩展到700所。QS指出,“学术引用”(academic citations)数据表明这些高校已基本能覆盖学生、雇主和研究机构可能感兴趣的所有一流大学(除了少部分不设本科或是只专攻少数学科的一流机构)。尽管有学者指出,“全球大约有17,000所大学,但世界大学排名却只能涵盖极少的精英式大学,这使得世界大学排名对于大多数大学毫无意义”,但事实上,任何排名都不可能容纳进世界上所有的大学。因而, QS公司开发了两个区域型大学排名,在满足择校生消费需求的同时,也满足了那些难以进入世界大学排名的高校对于社会地位认同的诉求。QS公司选定其他大学排名体系评价对象的思路与世界大学排名基本相似,只不过因各自排名的特定目的增加了一些特定的条件限制,如参与“年轻大学50强”评选的大学必须建于1962年以后。

二、QS大学排名体系的评价指标体系与权重系数

在QS大学排名中,假设用 y 表示一所大学的综合表现,用状态向量 x_i (i =教学、科研、国际化…)表示该校在 i 方面的运行状况。QS首先将 x_i 分解为若干能从某一侧面反映该校 i 方面现状的分向量 x_{ij} (如同行评议、论文引用等),然后通过一定的渠道获取 x_{ij} 观测值并将其转化为无量纲的标准分,最后通过集结模型 $y=f(\omega_{ij}, x_{ij})$ (ω_{ij} 即是与指标 x_{ij} 相应的权重系数, $\omega_{ij} \geq 0$ 且

$\sum \omega_{ij}=1$) 将各指标观测值标准分合成为一个整体性的综合评价价值, 并根据y值对各个大学进行排序。从集结模型 $y=f(\omega_{ij}, x_{ij})$ 很容易看出, 决定最终y值及排序结果的自变量就是指标 x_{ij} 以及相应的权重系数 ω_{ij} 。

QS 世界大学排名的指标体系及权重系数多年以来没有发生明显变化, 后来陆续推出的亚洲大学排名、拉丁美洲大学排名、世界大学学科排名的指标体系基本也是衍生自世界大学排名。(具体参见表1)

(一) 声誉

QS公司将大学的“声誉”(如今包括“学术声誉”和“雇主声誉”两项)视为大学排名中最重要的评估指标, 最少权重40%(亚洲大学排名), 世界大学英语学科排名甚至将声誉作为唯一的评价指标。然而, 相比“客观性”的硬数据, 声誉这种软数据不免显得主观性太强, 这在很大程度上使得QS公司的大学排名方法饱受批评。在这一问题上, 哥伦比亚大学教授保罗·瑟曼(Paul Thurman)认为, 客观性的测量固然重要, 但它无法捕捉到大学在质量方面的所有差异。想要全面评价一所大学, 并在一系列复杂机构间形成综合比较, 了解学者和雇主对该大学的看法至关重要。尽管见解看似主观, 但当大量有识之士的看法高度一致时, 这些见解便值得相信。此外, 人们对于大学排名信息的需求不仅是一些硬性数据, 也包括相关群体的见解。进行大

学排名时, 不考虑关键客户群体、雇主的看法可能会导致更令人怀疑的论断。因此, 尽管QS公司针对各种反馈意见逐年对声誉评估的具体操作方法相应地做出了一些调整, 但始终坚持赋予“声誉”以极高的权重, 并将QS全球学术声誉调查和雇主声誉调查作为评估大学声誉的唯一数据来源。某种程度上来讲, 重视大学的“声誉”是QS大学排名区别于其他大学排名体系的关键特征。

(二) 师资与生源

师生的质与量基本上是所有大学排名通用的评价指标, QS大学排名也不例外。QS世界大学排名、亚洲大学排名均将生师比作为评价教学质量的唯一指标。世界大学排名与亚洲大学排名对于大学国际化的评估也是以国际师生或出入境交换生的比例作为唯一的测量指标。

QS排名指标中的“教师”包括规划、指导以及承担教学、科研工作的所有职工, 含校长、副校长、常务副校长、学院首脑、教授、副教授、助理教授、首席讲师、导师以及博士后研究员, 但不包括研究助理、参与教学的博士生、住院医师、交换学者以及客座(访问)教职。计算学生数量时一般将本科生和研究生分开统计再相加, 其中本科生指那些攻读学士学位或同等学位的学生, 不含副学士学位, 研究生指攻读硕士或博士学位的学生, 如果无法获取这些明细数据, 那QS只能代以所有学生的总数(即包括参与非学位项

表 1. QS 公司四大跨国型大学排名的指标体系与权重系数

评价领域(x_i)	评价指标 (x_j)		权重 ω_j			
			WR	AR	LR	SR
科研质量	同行评议(学术声誉, Academic Reputation)		40%	30%	30%	40%~90%
	论文及引用率	师均引用率(Citation per Faculty)	20%	-	-	-
		篇均引用率(Citation per Papers)	-	15%	10%	0~50%
		师均论文数(Papers per Faculty)	-	15%	10%	-
	网络影响力(Web Impact)		-	-	10%	-
教学质量	生师比(Student-Faculty Ratio)		20%	20%	10%	-
	PhD 教职比例(Staff with PhD)		-	-	10%	-
国际化	国际师资比例(International Faculty)		5%	2.50%	-	-
	国际学生比例(International Students)		5%	2.50%	-	-
	入境交换生比例(Inbound Exchange Students)		-	2.50%	-	-
	出境交换生比例(Outbound Exchange Students)		-	2.50%	-	-
毕业生质量	雇主评议(雇主声誉, Employer Reputation)		10%	10%	20%	10%~40%

目的学生)。国际师生的“国际”指的是其“国籍”，欧盟中来自不同成员国的师生均算国际师生，香港学校中来自中国大陆的师生也算国际师生，对于具有双重国籍的人员，则以出生国籍(首个护照国籍)为准。

QS公司统计国际学生时并不计入交换生，但鉴于亚洲大学往往更注重交换生项目而非招收全日制外籍学生，在亚洲大学排名中单独列入“入境交换生比例”和“出境交换生比例”作为评估大学国际化的重要指标，并分别赋予2.5%的权重，同时将国际师生比例权重减半。

(三) 论文与论文被引用情况

论文及其被引用情况是评估大学科研质量的惯用指标。QS四大排名的指标体系共涉及三个与此相关的指标：师均论文引用率、篇均论文引用率和师均论文发表数。

许多学者指出，采用文献计量法评估大学科研产出会因国际论文数据库偏重收录英文文献而存在严重的英语偏向弊端，致使非英语地区的大学在排名中处于不利地位。在这一问题上，QS公司认为，世界大学排名看重的是大学科研成果的国际性，文献计量偏向英文期刊情有可原；但区域型大学排名注重反映机构在所在地区的地位与贡献，在文献计量时就不能再偏向英文期刊。因此，QS在亚洲大学排名和拉丁美洲大学排名中采用师均论文数而非师均引用率，并尽可能计入地方语言发表的论文以全面反映大学科研产出，同时通过测算篇均论文引用率以衡量机构论文的质量。二者的结合既兼顾了大学科研成果的数量与质量，又减少了文献计量中的语言偏见。此外，QS世界大学排名中以相对性的“师均论文引用率”替代绝对性的论文引用数以防止大学规模影响文献计量结果。

许多研究者都认为论文及其被引用情况是反映大学科研产出的最佳指标，因而许多大学排名体系赋予了这类指标以极高的权重，然而QS却只赋予论文和引用率方面的指标以20%~30%的权重。QS认为，采用文献计量法不可避免地存在着学科偏见，过分注重论文及其被引用情况对于那些偏向人文社科的院校是不公平的。因此相比看似“客观”的文献计量，QS更注重“主观”的学术声誉在评估机构科研质量中的作用。

(四) 网络影响力

2004年起，西班牙国家研究委员会(CSIC)每6个月发布一次“世界大学网络排名”(Webometrics Rankings of World Universities, WRWU)，评比全球大学对于网络学术资源的贡献。目前该排名体系设置了四个主要指标：Google、Yahoo、Bing和Exalead搜得的机构网页数量(进行对数标准化后取中值)占20%权重，用同样方法搜得的丰富文件(包括pdf、doc、ps和ppt格式)数量占15%的权重，被Google Scholar收录的机构论文数量占15%的权重，Yahoo和Exalead显示的外部连接数占50%权重。QS认为，从全球层面来看，英语国家的大学在这种网络计量排名中可能占有很大优势；但从区域层面来看，尤其在一个非英语地区，语言偏见问题已得到控制，那大学的网络影响力便可以成为一个有意义的比较点。基于此，QS公司在其拉丁美洲大学排名中设置了“网络影响力”指标，并赋予10%的权重，但鉴于WRWU排名中有关学术文章方面的数据与Scopus有所重叠，QS在使用网络计量时，删去了Scholar部分。

三、QS大学排名体系的指标观测值获取途径

(一) 通过QS全球大学声誉调查获取学术同行评议和雇主评议方面的数据

诚如保罗·瑟曼所言，“独特的世界学者和毕业生招聘调查是QS世界大学排名的基石”。2012年，QS公司的学术声誉调查收到了46,079份回复，雇主声誉调查收到了25,564份回复。这些调查数据直接决定了所有参与排名的机构在“声誉”指标上的最终得分，甚至在某种程度上因其极高权重而决定了其最终排名。

QS的全球大学学术声誉调查和雇主声誉调查均分为五大模块：第一部分要求受访者注明自己的姓名、工作单位等基本个人信息；第二部分要求受访者表明自己熟悉的国家、地区及学科和专业领域，以防出现“外行人妄评内行人”的情况；第三部分和第四部分是调查的核心部分，要求受访者在自己熟悉的国家、地区和学科、专业领域按要求列出一些自认为表现卓越的院校，但自2007年起不再允许受访者在此列举自己所在机构；第五部分为附加信息部分，主要要求受访者针对以往出版物以及

各种大学评价措施的重要性提出反馈意见。

既然声誉调查能对排名结果产生如此之大的影响，那受访者的来源、权威性和代表性势必成为人们关注的问题。2004年，QS主要从“世界科技出版公司”（World Scientific）数据库获取参与学术同行评议的学者信息。经过逐年的改革，如今参与QS两大声誉调查的受访者总共有七大来源。（具体参见表3）

从2011年QS大学学术声誉调查数据来看，52%的受访者已在学术界工作20年以上，52.7%为教授或副教授。这些受访者分别来自80多个国家和地区，其中65.7%的学者表示熟悉美洲，48.6%的学者熟悉亚洲、澳大利亚和新西兰，75.3%的

学者熟悉中东和非洲。就学科分布而言，受访者覆盖了艺术与人文（17%）、工程与IT科技（23%）、生命科学与医学（16%）、自然科学（19%）、社会科学与管理（25%）五大学科领域和50多个专业领域。

在选择雇主声誉调查受访者时，QS关注的是受访者所熟悉的行业与专业领域，而并非其所从事的工作或所在机构的性质，并且要求所有受访雇主必须有在大学招收毕业生的经历。从2011年的有关数据来看，27.6%的受访雇主来自人力资源部门，超过50%的受访者担任总裁、副总裁、首席执行官、总经理等要职。这些受访者分别来自110多个国家和地区，其中47.6%的雇主表示熟悉美洲，

表 2. QS 全球大学学术声誉调查和雇主声誉调查内容

模块	学术声誉调查内容	雇主声誉调查内容
个人信息	姓名,机构,职称(职位)及类别,所在部门,进入学术界年限	
知识说明	国家	列出自己最熟悉的国家(不一定是自己所在国家)
	地区	从“美洲”,“亚洲、澳大利亚和新西兰”,“亚洲、中东和非洲”三个选项中选出自己熟悉的机构所在地区(可多选)
	学科领域	从艺术与人文、工程与技术、生命科学与医学、自然科学、社会科学五个选项中选出自己专长的学科领域(可多选)
	专业领域	选择自己专长的专业领域(最多可选两项)
国内顶级机构	在每个自己所属学科领域中选出 10 个科研最佳的国内机构(不允许选择自己所在机构)	选出 10 个适合用来招聘毕业生的国内机构
国际顶尖机构	在每个自己所属学科领域中选出 30 个科研最佳的国际机构(必须在自己熟悉的地区中选,不允许选择自己所在机构)	选出 30 个适合用来招聘毕业生的国际机构(必须在自己熟悉的地区中选)
附加信息	对以往出版物、大学评价措施的重要性的反馈意见等。	

表 3. QS 全球大学学术声誉调查和雇主声誉调查的受访者来源

来源	解释	AR	ER
历届受访者	邀请之前已接受过 QS 大学声誉调查的受访者继续提供他们对所处领域大学质量的最新见解。	√	√
世界科技出版公司	QS 基于学科和地域代表性,从世界科技出版公司获取了 18 万份数据。但由于这一途径的有效性逐年下降, QS 自 2011 年起更多的从 Mardev 名单中获取信息。	√	×
Mardev- DM2	QS 公司 2011 年从中抽取了 200,000 条记录。	√	×
学术注册 Academic Sign- up	QS 公司在 2010 年开发了一个学术注册程序,吸引学者积极参与学术调查。为了防止某些机构通过大量注册来人为操控数据, QS 要对注册者进行筛选。	√	×
大学提供的 名单	2007 年起, QS 公司要求受访大学提供一份雇主名单,以便邀请他们参加雇主调查。2010 年起,要求这些大学也提供相关的学者名单。所提交的名单都要接受审查,反馈信息如超过 400 条便要对其进行随机抽样。	√	√
QS 数据库	QS 在 20 多年的运作中已建立一个全球主要市场的雇主信息库。	×	√
QS 合作伙伴	QS 有包括国际媒体组织和工作门户网站在内的广泛的合作伙伴,他们可以帮助 QS 发放调查邀请。	×	√

40.2%的雇主熟悉亚洲、澳大利亚和新西兰，52.1%的雇主熟悉中东和非洲。就行业分布而言，受访者覆盖了20多个行业领域（具体数据参见表4）。

（二）通过高校和第三方提供的统计资料获取师生数量方面的数据

在QS的大学排名系统中，多个指标涉及到师生数量问题。如生师比、PhD教职比例、国际师生比例、出入境交换生比例，在计算师均论文引用率和师均论文发表数时同样要用到教师数量方面的数据。QS大学排名中的这些数据主要由参与排名的院校提供，所以QS规定申请参与QS大学排名的院校有义务提供相关数据。同时，为了防止高校人为操控数据，QS公司还从全球56个国家的139个教育、统计部门及相关网站上获取这方面的相关数据，通过多方信息的比对来提高数据的可信度。

（三）通过“爱思唯尔”（Elsevier）的文献计量数据库“斯高帕斯”（Scopus）获取论文及引用率方面的数据

全世界有三个主要的出版物及引文数据源：汤森路透社的“科技网络”（Web of Science）、爱思唯尔的Scopus和谷歌学术。2007年以前，QS通过统计科技网络“基本科学指标”（Essential Science Indicators, ESI）数据库中近10年的有关数据来评估大学论文发表量及论文被引情况。但后来有人指出，尽管拉长统计年限能更好的体现大学在科研方面的丰硕成果，但这对于年轻院校来说是不利的。因此QS后来将数据统计范围缩短至最近5年，更加注重机构的最新科研产出。自2007年起，QS转向使用Scopus，最主要是因为Scopus的学术刊物覆盖范围更广泛。截止2012年11月，Scopus已涵盖5,000多家国际出版社出版发行的20,500多种刊物，其中包括19,500种同行评审期刊（含19,000种开

源期刊）、400种商业出版物、360种丛书、3,850多种期刊的“在编文章”（Articles-in-Press）、4.9亿项1823年以来的档案，覆盖科学、技术、医学、人文、艺术和社会科学多个学科领域。此外，Scopus收录了大量的非英语文献（只需有英文摘要），这可以减少文献计量过程中的语言偏见。除了转换数据源，QS自2011年起不再将“自引”（self-citations）算入文章引用率。

四、QS大学排名体系的数据处理与计算

（一）声誉调查数据的筛选与加权处理

1. 数据的筛选

为了增大样本的规模和稳定性，QS公司会参考近3年的声誉调查数据，对于接受过一次以上调查的受访者，QS会取用最新数据。此外，QS公司设置了大型的垃圾数据过滤程序和异常数据检测程序，以此来识别和剔除那些无效的和投机性的回复，从而消除无效数据的干扰和防止某些机构人为操控调查数据。

2. 加权处理

第一，根据受访者熟悉区域的分布及国内外的回复结果，对五大学科领域的国内外调查数据分开进行加权处理，以保证每组数据都具有均等的区域代表性。第二，分别将五个学科领域国内外调查数据的加权结果转化为百分制得分。第三，对国内数据和国外数据进行加权汇总。对于学术声誉数据，国际数据和国内数据按17:3的比例进行加总；对于雇主声誉数据，国际数据和国内数据按7:3的比例进行加总。QS指出，这一比例是根据QS分开收集国内数据和国外数据以前的调查数据分布设定的。第四，对五个学科领域的加总数据分别进行开平方根处理。QS认为，如果大学在某一个学科领域表现的尤为突出，那该

表4. 受访者行业分布

行业	比例	行业	比例	行业	比例
航空 / 国防	1.10%	IT / 电脑服务	7.50%	人才招聘 / 人力资源服务	5.40%
建筑 / 房地产	3.80%	法律	1.50%	零售	2.80%
咨询 / 专业服务	12.30%	制造 / 工程	9.30%	电信	2.50%
电子 / 高科技	3.70%	媒体 / 娱乐与艺术	2%	交通运输 / 分销	2.10%
能源	3.50%	金属 / 采矿	1.40%	旅游 / 休闲 / 酒店	2%
金融服务 / 银行	10%	制药 / 生物技术与医疗保健	5.60%	公共事业	1.10%
快速消费品	4.40%	公共部门 / 政府 / 非营利机构	8.60%	其他	9.30%

学科领域的卓越表现应当对最终得分产生一定影响，但影响也不能太大，而开平方处理恰好能够达到这一目的。第五，将五个学科领域的数据分别转化为百分制得分。第六，将五个学科领域的得分按同等权重加总出最后的声誉调查得分并进行标准化处理。

(二) 师生“全日制当量”(Full Time Equivalent, FTE)的换算

不同国家、不同大学中有着不同比例的兼职师生，在计算师生人数时，如果简单的计算人口总数必定会对一些大学造成偏见。基于此，自2007年起，QS公司将所有有关教师和学生人数的数据折合成全日制当量。换算过程中，QS公司采用的是美国的经典FTE计算公式，用全职人员数加上兼职人员数的三分之一，即默认兼职人员的平均工作量为全职人员工作量的三分之一。如果大学没有提供师生全日制当量，QS会参考机构所在国或所在地区的其他大学所提供的师生总数与师生全日制当量的比例估算有关当量。

(三) 各项指标数据的标准化处理与加总

在运用综合评价函数计算最终得分之前，要先将每项指标的得分都转化为无量纲的标准分。在2007年以前，QS公司采用了简单的标准分计算公式： $Score = \frac{x_i}{x_{max}} \times 100$ ，式中Score表示标准分， x_i 表示机构i的某项指标得分， x_{max} 表示参与排名的所有机构在该指标的最高得分。但QS意识到这种标准分计算方法存在着三个主要的弊端：不恰当地使用权重、缺乏对离群值(outliers)的控制、得分最高机构的成绩中的微小误差可能会对排名结果产生极大的影响。因此，从2007年起，QS改用Z分数计算指标标准分。首先，对所有指标的得分取自然对数，以防非显著性差异数据对总分产生过大影响；然后，计算每项指标的平均得分和标准差；最后，根据公式 $Z_{score} = \frac{x - \bar{x}}{\sigma} (\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{N}})$ 计算出Z分数，式中Zscore表示Z分数， x 表示某指标得分， \bar{x} 表示该指标的平均分， σ 表示该指标得分的标准差， N 表示样本容量，即参与排名大学的总数。

当所有指标得分均转化为Z分数之后，通过线性函数 $y = \sum \omega_{ij} \cdot Z_{xi}$ (Z_{xij} 表示指标 x_{ij} 的Z分数， ω_{ij} 表示与指标 x_{ij} 相应的权重系数， $\omega_{ij} \geq 0$ 且 $\sum \omega_{ij} = 1$) 算出最终整体性综合评价值 $y = f(\omega_{ij}$,

x_{ij}) ($y \in [0, 100]$ 且精确到小数点后一位)，并根据y值对各个大学进行排序。

五、启示

QS公司作为一家商业性质的高等教育资讯机构，进行大学排名活动的初衷是为国际择校生提供大学质量方面的信息，但也正是QS大学排名结果对国际学生择校行为的指向作用，使之对各国高校的声誉、生源、经济资助等各个方面都产生着不可低估的影响。因此，高等教育决策者、高校管理者有必要更加审慎、系统地认识QS大学排名体系，并有所选择地将其应用到高校的管理和决策活动中去。

第一，高校应全面了解QS大学排名的原理与方法，而不仅仅是关注QS大学排名的最终结果。从理论上讲，对类型各异的大学进行跨学科、跨地域的统一化质量评估和排名是不可能的。QS大学排名所谓的“综合”评价，事实上只是对不同意义的指标得分进行主观性的加权汇总，而这本身就存在着理论上的缺陷。如果不全面了解QS大学排名的原理、方法，其排名结果很可能会误导读者。

第二，QS大学排名作为一种第三方质量评估机制，可以为高校分析自身优势与不足提供参考信息。高校应当有选择性地根据QS排名方法采取相应的措施，将提升排名、建设世界一流大学和履行大学基本职能有机地结合起来。比如：鼓励师生在国际知名期刊上发表英文文章，加强国际交流，提高国际师生比例，尽可能地降低生师比，提高师资队伍的整体水平，加强与用人单位的沟通与合作，加强数据统计工作并积极向排名机构提供可靠的数据源等。

第三，包括QS大学排名在内的所有大学排名只能成为高校制定发展战略规划的参照物而不能成为指挥棒，提升排名只是高校实现更好发展的途径而非终极目的。大学在发展过程中注重通过提升排名来获取更多的资源、得到更好的发展本身无可厚非，但千万“别让排名挤掉大学使命”。

(刘强，教育部人文社会科学重点研究基地北京师范大学国际与比较教育研究院讲师，北京100875；丁瑞常，教育部人文社会科学重点研究基地北京师范大学国际与比较教育研究院硕士研究生，北京100875)

(原文刊载于《比较教育研究》2013年第3期)

我国重点大学建设世界一流大学的奋斗方向研究

——基于两个著名世界一流大学排行榜的视角

田 锋

当前，我国大学需要在哪些方面与世界一流大学缩短差距，我们建设世界一流大学的奋斗方向是什么？对此，研究一些著名的世界大学排行榜的评价指标可能有助于我们加深对这个问题的思考。

一、两个世界一流大学排行榜2012年排名

（一）《美国新闻与世界报道》

表1 2012年世界顶尖大学400强之前10及我国排名前5高校排名

排名	学校	总分	学术声誉得分	雇主声誉得分	师生比	国际教师得分	国际学生得分	每一个教工的论文引用得分
1	麻省理工学院	100	100	100	99.9	86.4	96.5	99.3
2	剑桥大学	99.8	100	100	98.3	98.2	96	97
3	哈佛大学	99.2	100	100	98.6	90	78.4	100
4	伦敦大学学院	98.7	99.6	95.6	98.4	96.3	99.9	94
5	牛津大学	98.6	100	100	100	98	95.8	89.4
6	帝国理工学院	98.3	99.8	100	99.8	99.8	99.6	87.3
7	耶鲁大学	97.5	100	100	100	92.7	63.9	93.3
8	芝加哥大学	96.3	99.9	96	95.6	68.9	77.8	96.4
9	普林斯顿大学	95.4	100	88.5	90.7	85.5	63.5	99.8
10	加州理工学院	95.1	99	51.3	100	96	91.1	100
44	北京大学	78.8	99.7	96.6	82.6	55.7	35	36.8
48	清华大学	77.5	98.8	97.6	86.1	50.6	26.3	32
90	复旦大学	68.3	92.5	84.8	42.4	18	49.8	51.4
125	上海交通大学	59.9	80	82.3	42.4	20.6	17.5	43.3
168	南京大学	53.2	70.8	39.4	49.3	41.8	5.8	40.9

“2012年世界最好大学排名”是采用了QS (Quacquarelli Symonds) 世界大学排名提供的数据库资料。在前400名之中，美国大学有83所，英国45所，德国35所，澳大利亚22所，法国19所，加拿大和日本各16所，荷兰12所，韩国11所，中国

9所。前十均为美英两国的大学，总分所在的区间为95.1~100。其中美国有6所，英国有4所。美国的麻省理工学院排名第1，英国的剑桥大学排名第2，美国的加州理工学院排名第10。我国排名靠前的5所大学分别是北京大学（第44位），清华大学（第48位），复旦大学（第90位），上海交通大学（第125位），南京大学（第168位），总分区间为53.2~78.8。

（二）《英国泰晤士报—高等教育专刊》

泰晤士报高等教育专刊从2004年起每年进行世界大学排名，由汤姆逊——路透社集团所开发，用于评判世界级大学在所有他们的核心任务——教学、研究、知识转移以及国际化等方面的表现。2012年排行榜中，前10均为美英两国大学，总分所在的区间为90.4~95.5。其中美国有7所，英国有3所。美国的加州理工学院排名第1，英国的牛津大学排名第2，美国的芝加哥大学排名第10。我国排名靠前的5所大学分别是北京大学（第46位），清华大学（第52位），复旦大学和中国科学技术大学（第201~225位），南京大学（第251~275位），其中北京大学和清华大学的总分分别为70.7和67.1，其他几所大学的分数均显示为数据保留。

共有12所世界大学进入这两个大学排行榜前10的名单之中，其中在两个排行榜中都出现的有8所大学，占总数的66.7%。这说明，尽管两个排行榜所采取的评价指标有所不同，但是仍然有超过半数的大学均满足这些不同指标且能取得较好的成绩，这些大学也是当之无愧的世界一流大学，均能经过不同维度的考察。这也说明，两个排行榜的指标能够在一定程度上较准确地反映出世界大学的发展情况，如麻省理工学院在两个排行榜分别排名为第1、第5，剑桥大学的排名分别为第2、第7。

表2 2012年世界顶尖大学前10及我国排名前5高校

排名	学校	总分	教学	国际化	企业经费	研究	论文引用
1	加州理工学院	95.5	96.3	59.8	95.6	99.4	99.7
2	牛津大学	93.7	89.7	88.7	79.8	98.1	95.6
3	斯坦福大学	93.7	95.0	56.6	62.4	98.8	99.3
4	哈佛大学	93.6	94.9	63.7	39.9	98.6	99.2
5	麻省理工学院	93.1	92.9	81.6	92.9	89.2	99.9
6	普林斯顿大学	92.7	89.5	54.5	79.5	99.4	99.8
7	剑桥大学	92.6	91.2	83.6	59.1	95.6	96.2
8	帝国理工学院	90.6	88.0	91.4	87.5	90.9	93.0
9	加州大学伯克利分校	90.5	85.1	49.7	65.4	99.3	99.3
10	芝加哥大学	90.4	89.6	55.3	数据保留	92.9	98.7
46	北京大学	70.7	81.5	54.1	99.9	67.9	64.3
52	清华大学	67.1	74.5	37.2	99.2	76.4	55.2
201-225	复旦大学	数据保留	数据保留	数据保留	数据保留	数据保留	数据保留
201-225	中国科技大学	保留	数据保留	数据保留	数据保留	数据保留	数据保留
251-275	南京大学	保留	数据保留	数据保留	数据保留	数据保留	数据保留

二、两个排行榜的排名分析

(一) 指标分析

《美国新闻与世界报道》排行榜所采用的一级指标为：学术声誉、雇主声誉、师生比、国际化、每一位教工的论文引用率，占总分的权重分别为40%、10%、20%、10%和20%。而《英国泰晤士报—高等教育专刊》排行榜所采用的一级指标为：教学、国际化、企业经费、研究、论文引用，占总分的权重分别为30%、7.5%、2.5%、30%和30%。

这些指标归纳起来主要有四类：第一类是学术类指标，主要是学术研究方面，包括研究环境、研究产出以及研究声誉三个方面，其衡量方式为在世界权威学术期刊上发表论文的数量，以及发表的论文被引用的次数，学术产出及其影响在排行榜中所占的权重均达到60%，甚至在有的排行榜中还超过了60%。排名前10大学均在这些方面取得了令人瞩目的成就。第二类是主体类指标，包括生师比，即学生与教师数之比。这可以从一个侧面反映出一所大学所有的基本教学条件。第三类指标是国际化类指标，主要通过学校所拥有的国际学生、国际

教师、国际合作的学术研究项目以及学术期刊来度量。第四类指标是质量类指标，包括教师质量、学生质量（主要是毕业生）。通过获取教师获奖种类、次数及学术成就的资料，通过对雇主的调查访问以及了解学校校友的获奖种类及次数，可以度量这些指标。

可以看出，这些排行榜对一些可以度量的指标十分重视，因其具有可获得性易于比较，能够在一定程度上反映出大学在某几个方面的实力，但是也存在过于依赖数字的弊端，对于一些难以度量而对于一所大学的发展十分重要的指标则关注不够，如大学办学思想、大学精神、大学文化、大学组织结构以及本科生的培养过程等等。因而，从某种意义上讲，现有世界大学排行榜在本质上是学术研究方面的比拼，尽管学术研究是现代大学的一项重要职能，但并不是唯一的职能。大学的宗旨是什么，值得每一个关注大学排行榜的人士深入思考。

(二) 我国高校在排名上的差距

从上述两个排行榜的2012年排名及我国排名前5位高校的排名情况可以发现，我国高校与世界顶尖一流大学的确存在较大差距。

如《美国新闻与世界报道》的QS排行榜中，我国排名前3的北京大学、清华大学以及复旦大学在前两项（学术声誉、雇主声誉）的得分和世界排名前10高校的差距并不大，大多只有5分左右的差距，甚至在雇主声誉方面我国的北京大学、清华大学得分均高过普林斯顿大学和加州理工学院；但从国际化这个指标开始差距明显加大。我国的5所高校在国际教师和国际学生方面得分均未超过60分，而世界前10大学这两项的得分大多在90以上。由此足见我国高校的国际化学程度有待大大加强。而在最能代表一所学校学术能力的指标“每一个教工的论文引用”上，我国5所高校的得分大多只有世界前10高校得分的1/2，有的只有1/3。这也说明我国高校整体的学术水平和质量在世界上还不够强，虽然在论文的总量上我国并不差。

英国《泰晤士报—高等教育专刊》的排行榜中，我国大学仅有北京大学和清华大学有具体的得分数据。在“企业经费”这个指标上，我国的两所大学均高过前10大学。这也说明我国一些国

家重点建设的985高校的办学经费以及获得的社会资助经费是非常充足的。在教学方面,虽有一定的差距,但并不大,只有10~15分的差距。在国际化这个指标方面,差距较大,相比于大多数前10高校都有20~30分的差距,但是也和少数高校的国际化程度相似。而在反映学术水平的两个指标“研究”及“论文引用”方面,世界前10高校均在90分以上,我国的两所高校显然差距十分明显,即差距在30~40分。这也再次验证了QS排行榜的相关得分具有一定的可信度,同时再次提醒我们:我国高校的学术水平还有待大大提高。

经过分析可以看出,我国追赶世界一流大学的高校与世界一流大学差距的重点表现在教师的学术成果方面,即论文发表与论文被引用上面。这一方面说明我国高校特别是追赶世界一流大学的重点高校的学术研究与世界主流学术研究仍有距离,具体表现在研究领域、研究内容、研究视角、研究方法等方面;另一方面也说明我国学者的学术成果传播能力亟待加强。由于我国大量的学术成果并非以世界通用的语言来撰写,导致有些学术成果虽质量很高,但由于语言障碍而没有被世界同行所知晓,更谈不上被引用了。同时,我们也不可忽视提高我国高校国际化程度的重要性,特别是吸引国际上的教师和学生来我国的大学进行中长期的任教和学习。

三、我国重点大学建设世界一流大学的奋斗方向

(一) 准确研判形势

担负追赶世界一流大学任务的重点大学要正确分析世界高等教育以及国际著名大学排行榜的发展态势,准确判断自身在当前阶段所面临的主要问题,不断加强对世界一流大学内涵的认识,结合自身的优势与特色,不断完善自身的战略定位,大力加强内涵建设,努力构建既顺应国际高等教育发展趋势又合乎中国特色社会主义本质要求的办学体制机制。要努力在一流特色办学方面形成自己的理念和风格,在为国家提供高质量的人才、科研成果、社会服务以及文化传承的过程中实现世界一流,在服务于国家和地方经济社会发展的过程中实现世界一流,引领我国高等教育的发展。

(二) 明确当前追赶的主要方向

国内一流大学如清华、北大等“985”、

“211”工程高校,并不缺乏建设的资金。其资金来源既有国家的专项建设经费;也有各级政府的财政经费,同时还有社会的捐助与委托项目经费;另外还有通过各种方式获得的创收经费。从这一角度来看,经费并不能成为建设世界一流大学的障碍。目前,迫切需要建设一批世界一流水平的学科、教师队伍、学生群体、学术成果,尤其要特别重视我们与一些世界一流大学存在较大差距的一些指标方面的建设。同时,也需要重视科研、教学科研人员的招录与管理、合理的生师比、充足的经费来源、充分的学术自由、先进的办学理念等等一些基础条件的建设。根据上述分析,当前我国高校的建设重点是大力提升学术研究能力及传播学术成果能力。

(三) 大力建设学术研究平台

相关高校要充分把握2012年出台的《教育部财政部关于实施高等学校创新能力提升计划的意见》所带来的重大契机,挑选若干个在国内同类高校中有鲜明的特色、实力较强、水平较高,且有突破的潜力的学科、实验室和工程技术中心,努力构筑一批复合型、开放式、具有先进水平的科技创新大平台或协同创新中心,在资金、条件建设上予以重点支持,使它们率先达到国内领先、世界先进的水平,以高水平的平台开展各类学术研究,创造发挥作用的舞台,吸引和留住一流的教师,以此形成吸引国内外一流人才的“磁场”,同时加大中青年优秀拔尖人才培养的“孵化地”。

(四) 积极推进学术工作的国际化

这里的国际化指的是学术要立足于全球的眼界,要以国际通用的标准来提高学术研究水平。同时还要加强学术的国际化交流,既要借助国外优秀的智力,也要让我们的学术研究成果走出去。相关高校需要把大学的人才培养、科学研究、社会服务以及文化建设放在国际视野中去开展,以国际一流大学的相应绩效表现为目标,采用国际认可的标准来评判自身的工作。例如,在对教师个人进行学术评价时,采用学术声誉、学术产出、学术影响等国际化的指标,使用国际国内同行评议的方式对教师学术声誉进行评议;以国际通用数据库上的数字来增加评价的公正性、透明性与客观性。长此

高等教育机构排名的柏林原则

联合国教科文组织—欧洲高等教育研究中心和华盛顿高等教育政策研究所于2004年共同发起成立了大学排名国际专家组（International Ranking Expert Group, IREG）。2006年5月18至20日在柏林召开的IREG第二次会议上，讨论通过了一系列高等教育排名的质量原则和操作范例——“高等教育机构排名的柏林原则”（Berlin Principles on Ranking of Higher Education Institutions）。

我们希望这项工作可以为各国、地区和全球性排名的制作和发布提供基本框架，最终形成一个使排名方法可以得到持续提高和完善的机制。虽然各个排名采用的方法差异巨大，但是这些原则将会对排名的改善和评价发挥积极作用。

排名应该：

A) 排名的目的

1. 排名应该成为对高等教育的投入、过程和产出评价的众多方法中的一种。排名可以提供有关高等教育的可比信息和对高等教育更深入的了解，但是不应该成为判断高等教育是什么和做什么的主要标准。排名是从市场的角度来看待高等教育机构，这可以作为政府、认证机构以及其它独立评价机构所从事工作的补充。

2. 排名应该明确自己的目的和目标群体。排名的设计应该和目的相一致。为特定目的或特定群体而设计的指标可能并不适合其它目的或目标群体。

3. 排名应该认识到高等教育机构的多样性并考虑到它们不同的使命和目标。例如，对研究导向型院校和面向大众化教育的普通院校的质量评价标准就截然不同。排名应该经常咨询被排名院校和相关专家的意见。

4. 排名应该清楚数据来源范围和数据背后的含义。排名结果的可靠性取决于对信息的加工和信息来源（比如数据库、学生、教师、雇主等）。好的做法应该是将来源于不同渠道、代表不同立场的数据组合起来，从而形成一个对被排名院校的更全面的看法。

5. 排名应该考虑被排名院校所处教育体系的语言的、文化的、经济的以及历史的背景。特别是国际排名应该意识到可能产生的偏向以及结果的精确性。并不是所有的国家和教育体系在对第三级教育机构的“质量”认定上都有共同的价值取向，排名系统不应该被设计用来强制进行这种比较。

B) 指标设计与权重分配

6. 排名方法应当清楚透明。排名方法的选择应该清楚、明确。透明包括指标的计算和数据的来源。

7. 指标的选择应该基于指标的恰当性和有效性。数据的选择应该基于对质量和学术优劣的各种评价指标的认识，而不是仅仅考虑数据的可获取性。排名应当明确告知为什么包括这些指标和这些指标反映的是什么。

下去，我国高校才能与国际一流大学站在同一大平台上对话，把我们的研究成果传递出去，并不断拓宽传播的范围。

（五）抢占学术制高点

在追赶的过程中，积极的学术竞争不可缺少。我国大学要结合自身特点和优势，争取在一些领域的学术研究方面占据世界的一席之地，如航空航天、海洋深潜技术、信息技术等一些高精尖端并带有战略性的研究领域，彰显我国在这些方面的学术水准与学术

高度，为国家学术独立与学术繁荣作出贡献。在这方面，相关高校一方面要利用好国家给予的各项优惠政策，另一方面也要充分把握一些国际交流平台所提供的机会，参与国际学术前沿领域内的知识合作研究与成果分享、研究方法的创新、知识成果的创新，以尽快缩小差距，赢得追赶的时间。

（田锋，广东省教育研究院高教室助理研究员，广州 510035）

（原文刊载于《江苏高教》2014年第1期）

8. 尽可能优先评价产出而不是投入。投入数据反映的是机构现有的、全面的情况并且通常都可以获取。然而，产出评价是对特定机构或专业的声望和/或质量更为精确的评价，排名的完成者应该确保这两者的权重达到恰当的平衡。

9. 指标的权重分配（如果有的话）应该非常明确并且尽量保持稳定。权重的变化会使顾客难以明确地判断究竟是实力的变化还是排名方法的变化导致了机构或专业的位次变化。

C) 数据的收集与处理

10. 排名应该有一定的道德标准并吸收好的操作方法。为确保每个排名的可靠性，数据的收集、使用和网络访问的提供应该尽可能的客观和公正。

11. 排名应该尽可能地使用审核过的、可核实的数据。这样的数据有许多优势：比如机构已经接受了这些数据，这些数据在机构间是可比的、口径是一致的。

12. 排名使用的数据应该是按照科学的数据收集过程所获得的。从不具代表性的或有缺陷的样本（学生、教师或其他群体）中获得的数据可能不能

准确地反映一所机构或专业的实际状况，这样的数据不应被用于排名。

13. 运用各种手段对排名活动本身进行质量保障。应该考虑采用评价高等教育机构的方法去评价排名本身。排名应该不断的根据专业意见去改进排名方法。

14. 采用体制性的措施增强排名的可靠性。这些举措可以包括设置顾问委员会甚至是监督委员会，委员会成员中最好包括一些国际专家。

D) 排名结果的公布

15. 提供有关排名制作的所有信息，使得客户能清楚的理解排名是如何得到的，并且允许客户自由选择排名结果的展示方式。这样排名的使用者可以更好地理解用于排名的指标。此外，他们有机会自行决定指标的权重分配方案。

16. 通过一定的处理消除或降低原始数据中的误差，并且通过恰当的组织 and 公布方法使得错误可以被校正。机构和公众应该被告知排名中曾犯过的错误。

（原文刊载于《大学排名国际专家组》官方网站，<http://www.ireg-observatory.org>）