

论高等教育的高质量评估*

王建华

【摘要】 伴随我国高等教育从快速增长转向高质量发展阶段,以测量为基础的量化评估需要向高质量评估转变。在对象方面,量化评估关注的是高等教育的质量,高质量评估关注的是高等教育发展的质量;在方法方面,量化评估以相关指标的可测为基础,高质量评估则容纳量化的、定性的以及其他相关的证据;在结果方面,量化评估以测量、描述和判断为主,而高质量评估更加强调以测量、描述和判断为基础的建构和再建构;在功能方面,量化评估强调以绩效为杠杆进行问责,高质量评估则强调以评估促进发展;在风险方面,与问责导向的高利害的量化评估相比,以发展为导向的高质量评估是低风险的。

【关键词】 高等教育;量化评估;高质量评估;高质量发展

【作者简介】 王建华,南京师范大学教育科学学院教授 (南京 210097)

伴随高质量发展从经济领域到社会领域再到教育领域的渗透,高等教育高质量发展成为政策的焦点。为促进高等教育的高质量发展,我们需要弄清高质量发展中“质量”的内涵,弄清楚到底是什么使一个国家的高等教育体系有效,使一个国家的大学能够持续培养出杰出人才,并生产出创新性的知识。如果对这些根本的问题不清楚,只是盲目地追逐更高的排名,会削弱而不是增强高等教育的竞争力,最终也导致“对全球化的知识经济里‘智慧’国家的重要性的彻底误解”^[1]。基于此,在高质量发展阶段对于量化评估的反思不仅是评估方法(论)层面的事且有更广泛的社会和政治含义。“虽然围绕着衡量标准的问题存在诸多口舌之争,争论的重点却不在于此。口舌之争的背后是关于目的和价值的争议。衡量方法一直是反映更深层次价值

的政治冲突的替代品。任何关于政策的意见冲突,其根本问题都会通过我们选择考虑问题维度的方式暴露出来。”^[2]同样的,关于量化评估和高质量评估的争议绝不是评估标准或指标的争议,而是关乎高等教育发展观和价值观。在高质量发展阶段不可能有简单的科学公式可以用来衡量高等教育发展的水平或评估高等教育之于经济社会发展和人的发展的重要性。以对量化评估的反思和澄清为切入点,以高质量评估为方向,一方面可以检视高等教育发展观的偏差,重申人的全面发展在高等教育发展中的中心地位;另一方面也可以揭示在基于知识和创新的社会里高等教育发展阶段的升级。在新的发展阶段和发展范式下,我们需要更加关注那些在评估中不能量化但同样重要甚至更加重要的高等教育能力。实践证明,“信息可用性的指数级增

* 本文系教育部哲学社会科学研究2020年度重大课题攻关项目“教育高质量发展评价指标体系研究”(项目编号:20JZD053)的研究成果。

长使学生记忆和复述知识变得不那么重要之后,培养学生在现实情境中综合、转化和应用知识的能力变得更加重要,这些能力包括学生的创造力、批判性思维能力、团队协作能力、高效沟通能力和适应快速变化的环境的能力”^[3]。基于此,伴随高等教育从高速增长向高质量发展阶段的转变,关于高等教育的评估也需要从量化评估转向高质量评估。在高等教育高质量发展阶段,高质量评估是建设高质量的高等教育体系的核心。具体而言,为适应并促进高等教育的高质量发展,关于高等教育的评估需要从可量化转到高质量,从关注发展的效率转移到发展的范式,从关注发展的结果转向关注发展的过程,从聚焦高等教育机构的质量转移到聚焦高等教育系统的创新。

一、量化的风险

从实践出发,当前高等教育中对于评估尤其是量化评估有两种极端的态度,一种是认为评估至关重要,是高等教育发展的指挥棒。所有的政策、项目或工程都需要有评估与之匹配,评估驱动已经成为高等教育改革和发展的新常态。实践中评估的范围越来越广,评估的方式越来越多。评估指标的重要性正在上升。评估和测评结果的应用也越来越广泛。^[4]另一种是对于评估的蔑视或否定,认为评估,尤其是那些量化的评估,完全不靠谱。以评估驱动高等教育改革和发展将误入歧途。高等教育改革和发展离不开人的理性判断,甚至也少不了基于测量的评估。需要注意的是,“知识工作,如科研(新知识的创造)和学习(新知识在某个人身上的创造),完全不同于工业生产的成品,因此应由一种完全不同的方式来衡量。我们需要的不是技术,而是思维术(thinkology)”^[5]。现在的问题是,那些基于有限信息的量化评估结果被给予了过度的关注和滥用,从而使理性走向

了非理性,科学异化为了科学主义。对于高等教育改革发展而言,评估是从属高等教育的,无论如何,评估都不能比高等教育本身还重要。“我们切莫误认为必要之恶为善。很容易犯这一错误。要是想运送水果,就必须听装,因而也就必须遗失水果的某些好品质。然后总会碰到这样一些人,他们变得喜欢水果罐头,更甚于新鲜水果。”^[6]客观而言,相对于高等教育自身固有的重要性,评估无论如何只能是“次要”的,是“必要之恶”。但实践中为评估而评估、为量化而量化正在僭越高等教育的本质和规律,种种评估指标不是在促进高等教育发展,而是在“规训”和“宰制”高等教育发展。究其根本,一方面,高等教育改革和发展离不开人的理性选择和判断。“我们需要有恰当的评价框架;我们也需要有机构和制度来为促进我们的目标和对价值判断的承诺而工作;此外,我们还需要有行为规范和理性思考来使我们得以实现我们努力争取的目标。”^[7]另一方面,高等教育的改革发展虽然是人为的,但又并非是人的理性可以完全计划的。基于理性的有限,加之高等教育自身复杂性的约束,人类不可能对于特定时期高等教育发展状况做出可量化的精确的评估。除了复杂性之外,实践中高等教育发展的实际效果还具有显著滞后性。实践证明,“在开始改革的初始成本出现的时间和显著的改革预期效益真正实现的时间之间,存在巨大的时间差。教育改革的滞后时间远远超过典型的改革,例如劳动力或产品市场改革”^[8]。基于此,虽然在高质量发展阶段我们仍然需要评估,包括量化评估,但需要的是适度的、有限的、恰当的评估,而不是强制的、武断的、甚至是威权主义的评估。换言之,高质量发展需要的是高质量评估而不是现在正在流行的量化评估。

近年来,受量化评估的影响,学界关于高等教育的讨论越来越关注“一流”和“排名”,高等教育“育人”的本质在评估中被忽

视。“当前的大学氛围‘鼓励学者过度生产包装,为的是掩盖住内容的贫瘠。’‘在学术界,传统的评估品质的方式,是基于同行评审和行业判断,靠的是学术纪律的信条,但现在,这种方式正遇到挑战,新的方法重在科研行动的表现力,强调的是多面手、高效率和透明度。’”^[9]对于“一流”和“排名”的关注是高等教育处于高速增长阶段的显著特征。在高质量发展阶段不是不要关注“一流”和“排名”,而是需要将其置于“次要”的位置。“次要”的位置并非没有价值。相反,“惟其有价值,故而次要——有价值之物,必定不是价值之源头。说某事‘次要’,意思是它不能成为自己的理由”^[10]。“一流”和“排名”对于大学或高等教育发展有价值,但作为量化评估的产物,其只有相对于特定指标体系才有价值,它既不能准确反映高等教育质量的好坏,更不能反映高等教育发展质量的高低。更关键的是,那些来自第三方的量化评估结果或排名只“是评判(judgements),而不是事实陈述(statements about matters of fact)。它们是完全可以逆的(reversible)”^[11]。长期以来,由于受量化评估和高速增长发展观的影响,高等教育发展在价值观的层面上往往过度关注效率而不是效果,增长被等同于发展。大学发展追求的也是可量化的或者是有显示度的科研成果或社会贡献,那些无形的不可量化的教育价值被忽略。“效率开始向内部转移,对价值缄口不语,对目标保持中立,对方式却滔滔不绝。”^[12]高等教育中的量化评估过度强调绩效和成果导向,忽视了高等教育之所以为高等教育的本质。量化评估以存在某种最佳实践为假设,驱动所有高等教育机构围绕相关指标展开竞争。“用当下流行的语言来说,这是‘一个最好的体系’:专家、职业人士、官僚都知道学生们需要何种教育以及提供教育的最佳方式。在这个意义上说,所谓‘一个最好的体系’其实是一个工厂模式:公共教育的组织——即便不是内容——被标准化、均质

化,而且被当作大型市场的公共产品提供给社会。”^[13]在量化评估模式下,与那些显性的成果或可量化的绩效相比,作为过程的人才培养或育人被忽视。

与高速增长阶段对于数量或规模的简单重视不同,当下以高质量发展为名,那些以“一流”命名的各种“工程”或“计划”又开始追逐有质量的数量。各种评价开始强化代表性成果或标志性成果的重要性。与过去对于数量的简单比较不同,高等教育评价中对代表性成果或标志性成果的强化有其积极意义,但也隐藏着更大的风险。表面上看,对于代表性成果或标志性成果的强调可以淡化量化评估的压力。实质上,由于评价成果本身的方法没有变化,量化思维依然盛行。只不过是过去直接计算某种级别或类型的论文的数量,转换成了看某些顶尖期刊的论文,抑或某些高影响因子的论文,抑或高被引的论文,抑或是获了什么级别奖项的论文。由于可以被作为代表性或标志性成果的期刊资源的高度稀缺,绝大多数的成果因为发表的期刊不够权威或没有高被引而被排除在评估对象之外,由于评估对象的范围被进一步缩小,量化评估的风险将进一步增加。因此,从表面上看,突显代表性成果或标志性成果淡化了数量本身的重要性,强化了质量的重要性,但实际上如果确定哪些成果是标志性或代表性成果的思路没有改变,突显标志性成果和代表性成果也可能会抹杀高等教育发展过程和成果的多样性。更何况,那些在现有评价体系中被认为是代表性或标志性的成果未必真的就可以代表或标志高等教育发展的水平。从高等教育的本质或规律出发,高等教育的发展不能轻易用任何指标或成果来代替。无论标志性成果还是代表性成果所强调的仍然只是“最终成果”,忽视了高等教育发展的过程和潜力。

单纯从结果的视角审视,高等教育发展中一切活动都只是实现目的的手段,其自身

固有的重要性被忽视。只有以自由的视角重新审视高等教育发展,高等教育自身固有的重要性才会被凸显。在经济社会转向高质量发展的背景下,高等教育之所以要高质量发展,并不只是为了满足经济社会发展的需要,高质量发展也意味着高等教育发展中人的培养过程需要改进。换言之,高等教育高质量发展不只是为了生产更优质的人力资本或者科研成果,在根本上是为了实现“高等的教育”或者“成人”。基于此,当前,关于高等教育发展的评价,需要超越对于绩效和成果的迷恋,更加突显过程性因素之于高等教育发展的重要性。当然,这并非否定绩效和成果导向之于高等教育发展的积极意义,而是认为,在高质量发展阶段,需要以更广阔和综合的视角,来重新审视高等教育发展观以及大学发展范式。无论高质量发展还是对于高质量发展的评估,都绝不存在一个简单的“公式”,它是一个整体性的概念,与各种各样的活动以及机构和制度有关,高质量发展过程中需要综合考量过程层面以及实质性机会层面的多种因素。^[14]

表面上看,量化评估是一个评估的技术或方法的问题,但实质上量化评估作为一种范式背后有独特的政治内涵。“在任何时代,效率所谓的技术概念都富有政治内涵,我们的时代也不例外。”^[15]如果没有政治和政策的介入,量化评估即便存在也不会对高等教育发展产生深刻影响。和其他各种排行榜一样,在20世纪的大部分时间里,大学排名也主要是媒体的“自娱自乐”。政府和学界并不关心。20世纪末以来,随着世界一流大学建设运动的兴起,排名作为绩效评价的重要指标,成为政府部门高等教育决策和政策的依据。正是由于政治和行政权力的介入,原本漏洞百出的排行榜在高等教育评估中被赋予了权威性,并从根本上改变或形塑了学生、教师、校长、公众和政府对于高等教育发展水平的认知以及各利益相关方接受或发展高等教

育的动机。“评估和测评在教育系统内的扩展也反映了政府努力使公共部门‘现代化’,并将商业实践纳入公共服务管理。”^[16]最终当评估的方式与高等教育发展的目的脱离,对于测量或可量化本身的需求就成为一种目的。以排名为显著特征的量化评估并不会告诉大学或高等教育发展应往哪儿走,它只是告诉政府和办学者应该参照排名的指标在最短的时间内取得最佳的排名,以便在声誉竞争和政府的资源分配中占据有利地位。量化评估作为政府对高等教育资源进行配置的一种政策工具,其本身并不天然具有权威性,甚至天然地不具有权威性;但因为量化评估结果与资源分配密切相关,其才具有了权威性,乃至影响高等教育发展的方向。就高等教育评估而言,首先应关注的是高等教育本身或高等教育发展本身,其次才是评估,最后才是量化不量化。现在价值序列却恰恰倒过来了。“方法”取代“本体”成为政策和实践关注的焦点。为量化而量化,为评估而评估,成为高等教育评估中的常态,而高等教育本身或真正的高等教育发展被忽略。如果为了提升或证实高等教育发展的重要性而在评估中让“一流”和“排名”成为目的,高等教育发展的目标和价值将被“搁置”或“错置”,那些量化的评估结果很容易成为组织和个人追逐的目标,最终数量增长成为高等教育发展目标和价值的替代物,这样不但会导致对于发展的误解,也不利于高等教育从高速增长向高质量发展阶段的转变。基于量化评估的局限性,政府和大学在制定高等教育高质量发展政策时需要摆脱对于数据、证据或指标的迷思,高质量的高等教育政策未必建立在高质量的数据、证据或指标的基础上。本质上,高等教育研究和发展都属于人类实践理性的范畴,未必总是能够找到高质量的数据、证据或指标体系作为决策的依据,那些不可量化的经验、传统和实践智慧对于推动高等教育高质量发展也同样重要。

二、评估的迭代

本质上,评估作为一种理性判断,是人类心智的产物。“评估并不是简单地在某一天就突然出现的,它是众多相互影响的建构和再建构发展的结果。”^[17]关于教育的评估同样如此。“教育领域内评价是不可避免的。教育概念本身就根本不可能没有评价。”^[18]在思维层面上,高等教育评估范式大致可以分为两种,一种是科学范式,另一种是哲学范式。“科学认识论的最基本原则是主客二分。哲学认识论的最基本原则刚好与之相反,是主客合一。”^[19]科学范式的评估强调评估方法的科学性以及评估结果与事实的相符性。在科学范式下,基于主客两分,评估活动和评估结果完全由评估机构和评估专家所主导,为了避免影响评估的科学性,利益相关者的主张、怀疑和焦虑作为消极因素完全被排除在评估之外。与之相较,哲学范式的评估强调评估结果与评估对象的相互建构以及评估的过程性。在哲学范式下,评估的过程也是一个思想涌现和知识生产的过程,而不只是信息的收集和数据处理的过程。遵循科学思维,评估对象是客观的事实,通过不断改进评估指标,最终一定可以发展出科学的评估方法,基于科学的方法就可以得到科学的评估结果;遵循哲学思维,评估对象并非完全客观的事实,而是社会建构的“事实”,评估指标的选择取决于评估的目的,并不存在唯一的科学的指标体系。由于主体与客体间相互建构,无论基于何种指标体系都不可能得到完全与事实相符的评估结果。“评估是人类思想的建构,它与某些‘客观实际’的符合不是也不会成为一个终结。”^[20]当前由于受实证主义、科学主义以及管理主义的影响,高等教育评估中科学思维占主导,无论量化语言多么脱离高等教育实际,使用量化的语言都比使用抽象的概念更容易被政府和社会所接受。

基于实证主义和科学主义的方法论,量化评估假定只要找到适合的指标就可以对高等教育进行精确的测量,通过对测量结果的分析,高等教育中存在的问题可以被暴露,高等教育发展所取得的成果就可以被“计算”。为了便于测量,高等教育自身所蕴含的政治、经济、文化价值等被忽略,所谓评估被“简化”为了一个简单的统计或计算的过程。“对科学范式的承诺不可避免地导致了对正规的定量测量工具的过分依赖。最初,它们被看作科学变量的‘操作化’,但最后,它们自己也成了变量。”^[21]结果就是,量化评估原本是用来“保卫”高等教育,即以数据作为客观的证据来为高等教育改革发展寻求合法性,但最终因为量化的准则违背了高等教育发展的规律,正在“摧毁”高等教育。

与科学思维和哲学思维的两分法不同,古贝(Guba, E. G.)和林肯(Lincoln, Y. S.)将教育评估分为了四代。第一代评估的核心特征是“测量”,强调对调查变量进行技术性的测量。评估的结果就是测量的结果。第二代评估的核心特征是“描述”,评估者的角色就是描述者,评估就是描述关于某些规定目标的优劣模式。此时,“测量”不再等于评估,而是作为评估的工具。第三代评估的核心特征是“判断”,评估者扮演评判员的角色,根据绩效来划分人或事务的等级。^[22]评估实践中“测量”、“描述”与“判断”并非相互替代,今天在高等教育绩效评估中,早期的技术性和描述性功能依然存在。鉴于这三代评估中存在“管理主义的倾向、忽略价值的多元性以及过分强调调查的科学范式”三个重大缺陷,古贝和林肯提出了第四代评估的概念。第四代评估以“响应”和“建构”为核心特征,强调在利益相关者参与的基础上决定要解决什么问题 and 收集什么信息,并从建构主义的视角赋予了评估全新的含义。在目标上,第四代评估聚焦利益相关者(评估中将面临风险的人或组织)的“主张”、“焦虑”和“争议”。所谓“主

张”是利益相关者提出的有利于评估对象的方案。“焦虑”是利益相关者提出的不利于评估对象的方案。“争议”就是理智的人不一定都赞同的某种事情状态。实践中不同利益相关者持有不同的“主张”、“焦虑”和“争议”；评估者的工作就是发现这些不同的因素，并且在评估中解决这些问题。^[23]在方法(论)上，建构主义方法论和传统的调查和统计方法非常不同：后者是线性的、封闭的；相反，前者是往复的、互动的、辩证的，而且有时很直观，多数情况下是开放的。和更严格的、规范的、苛求的传统方法论相比，“没原则、懒散、无知、没有竞争力”的建构主义方法论远不是一种“简便”的方法；它要求甚高，因此多数建构主义者的感觉是焦虑和疲惫。这是一条不同的路，可能点缀着砾石，但是却引向奢华而显然未被欣赏的玫瑰园。^[24]与前三代相比，第四代评估无论从观念上还是程序上都发生了根本改变。第四代评估为评估者提供了一种新的思维和行动指南。“第四代评估者承担了前三代评估者的角色，重新定义他们，并把他们融合成一个比以前更有技巧性的实践者。第一代的技师(测量专家、测试制造者和统计专家)转变成人类设备和人类数据分析师。第二代描述者角色转变成领路人和历史学家。第三代的判断者角色被转变成判断过程中的仲裁者……第四代评估要求评估者从控制者的角色转变成合作者。评估者不是作为调查者的形象而是担当教与学的角色。评估者的新角色不是发明家而是给现实定型的人。第四代评估者将他(她)自己从被动的观察者的角色中跳脱出来，并认同和接受身为变化媒介的角色。”^[25]

第四代评估以改进前三代评估的重大缺陷为目标，具有理念的先进性，也符合教育评估发展的趋势。但和任何一种评估范式一样，第四代评估在其优越性的背后也隐藏着一些难以避免的弱点。“从本质上来说，第四代评估是具有分歧的——它们倾向于提出更

多的问题，提出的问题多于回答的问题。总是有无法解决的主张、焦虑和争议，而且现有建构很可能只能支撑一会儿。可能会突然得到新的信息，复杂性也可能突然增加，这时就要重新考虑那些认为已经解决了的‘主张’、‘焦虑’和‘争议’。重新建构几乎是任何时候都是适宜的。第四代评估永远不会停下来；它们只会暂时停止。”^[26]鉴于此，如果说传统量化评估方法的明显缺陷和优势在于评估结果的确定性和唯一性，那么，第四代评估的最明显的优势和缺陷则在于其评估结果的不确定性和多样性。在建构主义方法论和本体论的指引下，第四代评估强调利益相关者的高度参与，强调评估过程的开放性，强调评估主体与评估对象之间的沟通和协调，从而导致评估的过程十分漫长，且没有明确的结论。在一种理想情境中，或许评估本来就应如此，好的评估应致力于丰富我们对事物复杂性的理解和想象，并让所有的利益相关者在评估中得到成长，而不是经由简单化的程序给出确定性的结论或建议。“他们确信，他们用一种人们永远不知道事情到底是什么的谦逊方式发现了‘实在’；他们确信，有关事物是怎样的建构乃是由调查本身创造的，而且不被某种神秘的‘本质’所决定。用相对性替代绝对性、用授权替代控制、用局部理解替代普遍性诠释，用谦逊替代傲慢，这就是第四代评估者得到的最明显的收获。”^[27]但在政策实践中，那些没有明确结论或建议的评估很难实施，即便实施也很难受到重视，即便偶尔受到重视也很难持续。

当前，伴随高等教育从高速增长转向高质量发展阶段，高等教育的评估也需要从量化评估转向高质量评估。借鉴评估的哲学范式以及第四代评估的新理念，在高质量评估中，“我们要承认矛盾，接受有些东西无法用同一尺度衡量的事实。我们需要注意的不仅仅是我们可以衡量的东西，更重要的还有那些我们无法衡量之物。关于价值和目的讨论

的确艰深而且苛求,但是如果没有这种讨论,效率就沦为崇拜,责任则成为一场误导人的数学练习”^[28]。实践中无论宏观的作为一个系统的高等教育还是微观的作为一个组织的大学都是极其复杂的,涉及人性的、政治的、经济的、文化的、社会的等各种因素;这些因素更多的是社会建构意义上的“主观事实”而非实证主义的“客观事实”。以测量为主导的科学范式的评估无法反映高等教育事实的整体性,在此过程中即便量化评估发现了某些事实,但由于它遮蔽了高等教育本身的意义,评估自身的价值也要大打折扣。在高等教育评估中,受人性、政治、经济、社会、文化等各种因素影响,那些简单的结果不一定被认可,不同的量化评估甚至会得出相互竞争的“真相”,但若仅局限于过程而没有结果,评估活动也将缺乏合法性和合理性。很明显,评估过度关注结果或建议不行,但若完全没有结果或建议恐怕也不行。过度关注结果或建议会导致其他可能的焦点被忽略,但没有明确的结果或建议也会导致评估自身的目标被忽略。第四代评估通过强化利益相关者思维来克服传统评估方法中主客两分的缺陷,这是其优势也是其弱点。不同利益相关者参与评估的意愿和能力存在差异,其利益诉求也可能会相互矛盾。通过不同利益相关者的建构与再建构,评估的确可以增进利益相关者对于评估对象复杂性的理解,丰富其对于评估本身的认识,甚至可以生产出知识,但评估的过程毕竟不同于高等教育的过程,更不同于知识生产的过程。考虑到技术上和制度上的复杂性,过度强调利益相关者的参与和协商并非总是有利于高质量评估,源于不同利益相关者的信息的泛滥也可能会使得那些真正与评估相关的核心信息被遮蔽;与此同时,过度关注评估的教育性功能也会削弱高等教育评估自身的其他功能,高等教育高质量评估不能以评估为目的,但也不能完全不以评估为目的。本质上,高等教育高质量评估应从

高等教育实践出发,经由评估的过程,促进高等教育的高质量发展。

三、高质量评估的要义

无论在高等教育领域还是在社会其他领域,基于量化的评估都不是今天才有,而是有着相对漫长的历史。那么,为什么今天量化评估成了高等教育发展中问题的焦点,引发了广泛的讨论?原因大概有两方面,一个是量化评估的重要性突显,另一个是高等教育自身的重要性突显。这两种重要性并不是孤立存在的,而是相互增强。从高等教育自身的重要性变化来看,在高等教育发展与经济社会发展之间关系较为疏远的时期,即便有量化评估的存在也不会引起广泛关注,量化评估的结果也不会影响高等教育发展的方向。“随着人类社会向知识驱动型经济和社会的发展,教育对国家未来的经济表现和相对的经济地位从未像现在这样重要,而它也使个人能够充分参与经济和社会的发展。在这种情况下广泛参与教育只是一个方面。他们对质量的强调也使评估和测评活动更加突出。”^[29]此外,从高等教育评估的发展来看,非量化的经验评估一直在高等教育中居于支配性地位,从非量化的经验性评估向量化评估的转变,极大促进了高等教育发展。在以大众化和普及化为代表的高速增长阶段,量化评估也曾是一种先进的评估范式,满足了政府和社会对于提高高等教育办学效率,甚至是办学质量的需求。“市场语言同样以品德之名将效率歌颂。它赞美生产效率,而且在后工业时代尤为赞美知识的高效生产。”^[30]因此,基于科学思维的量化评估今天所面临的批评并不能被用来否定其历史的贡献。在高等教育发展的特定阶段追求效率没有错,量化评估也不是完全没有道理。关键是我们需要什么。“衡量标准的选择相当程度上不仅取决于我们所衡量之物,也取决于我们衡量

的原因。”^[31]在高等教育发展需要提高办学效率的时代,量化评估就是一种好的评估范式;当高等教育发展过了高速增长期需要强调内涵和高质量发展时,那种以效率为主导的量化评估就显得不合时宜。过去那种为了高等教育高速增长阶段所设计的评估模式很难适应高质量发展阶段。在高质量发展阶段,“绩效测量应该是宽泛的,而不是狭隘的,既包括定量数据,也包括定性数据,还包括高质量的分析”^[32]。说到底,高质量发展并非不需要效率而是不需要高速增长阶段的那种效率。

教育的本质是“育人而非制器”。^[33]由于人性本身的易变性,教育过程中不可避免地存在风险。“教育不是一种可以预料的活动,它不像工程计划那样,人们能够准确地预测其结果。教育本身实际上离不开冒险。即使是最佳的教育活动也总是伴随着失败的可能性。”^[34]不过,为了实现“成人”的目标,教育是一场值得冒的“美丽风险”。^[35]与“教育自身”的风险不同,评估尤其是高利害评估会把更多的利益相关者置于风险中。在科学范式下,评估尤其是竞争性的绩效评估,常常涉及诸多利益相关者的重大关切。好的评估结果不只可以为评估对象增加社会美誉度,还能带来实实在在的利益;而差的评估结果则会使评估对象陷入危机。就高等教育发展而言,那些以“排名”为特征的第三方评价的结果,经常被政府或大学作为政策工具或治理工具。大学在排行榜上的排名也经常被简单地等同于大学的办学水平,而一个国家在某个排行榜上的整体排名则被认为是这个国家高等教育综合实力的反映。为了规避风险,同时也为了维持大学的声誉,很多国家的很多大学根据排行榜的指标来调整发展的策略甚至是战略。短期来看,这种选择是理性的,似乎也是有效的,可以保障大学在一年一度的排名竞争中取得预期的进展;但长期来看,这种以增长为目标的发展使大学及其利益相

关者处于高度风险中。“评估和测评的发展功能可能会在强调问责功能的过程中受到阻碍,因为其中的风险将限制学校负责人在收到反馈后公开暴露自身的弱点,从而无法借此提升他们的实践能力。因此,合理设计如何使用评估和测评结果来问责十分重要,这样可使非预期的影响最小化。”^[36]高风险或许可以带来高激励,甚至会强制性地改变大学的一些传统或游戏规则以提高效率。但高风险不符合大学的组织特性,也不符合高深知识生产、传播与应用的规律。“随着公共问责制日益普遍,控制差错范围的压力与日俱增,虽然消除了失败的风险,但也让成功的可能随其风险一并消失。用来评估研究提案的质量和重要性的‘标准’实际上系统性地阻断了开创性的学术研究,并偏向支持更可预测的结果导向性研究。这种盲目崇拜数字的做法造成了巨大的代价,特别是对人文学科和社会科学而言,其微妙而复杂的观念与见解在刺眼的‘定量’眼光的审视之下很容易便化为无形。”^[37]即便在最乐观的情境下,那些来自第三方的排名或许可以塑造大学光鲜的社会形象,但亦无助于甚至会损害大学对于杰出人才的培养。因为现有以测量、描述和判断为主要特征的大学排行榜,更加重视量化的科研成果的产出,人才培养由于其难以量化的特征,往往被量化评估所忽视。

基于此,从风险的角度看,量化评估是高风险的评估。这种高风险,一方面是评估本身的风险,即量化评估仅仅考察了可以量化的事项,而忽略了其他不可量化的重要事项,甚至量化本身成为目标。为评估而评估,为量化而量化,极大削弱了量化评估的实践价值。另一方面也是对评估结果使用的风险。量化评估以统计学为基础,对于评估结果进行了数学意义上的化简。基于高度简化后的得分的排名会把大学之间极微小甚至是毫无意义的差异加以放大,最终影响大学的发展。“评估和测评结果的高利害使用,可能

会造成教育进程的走样,因为学校负责人会关注使他们负责的措施。此外,当评估和测评框架倾向于强调其问责功能时,会产生一种风险,即评估和测评主要被视为对学校负责人问责的工具,用来‘控制’和评估其对法规的遵守情况。”^[38]由于高等教育自身的复杂性以及高等教育过程的模糊性、不确定性,那些量化评估的结果并不能准确衡量大学对于人才培养、科学研究和社会服务的贡献,更不要说文化的传承和创新创业,排名所收集的有限的可量化信息只关乎大学里那些最简单的事实。因此,虽然那些好的大学也会在排行榜上名列前茅,但在排行榜上名列前茅本身并不能说明大学在现实世界中的真实发展水平或综合能力。那些最终被排行榜列为优秀的大学可能主要是因为其本来就优秀,这些大学并未从排名中受益;而那些原本优秀但因为技术性原因被排除在某些名次之外的大学,则成为大学排名和量化评估的受害者;而那些原本并不优秀但因为符合排行榜的指标而排名靠前的大学,则会误导公众,影响政府的高等教育投入和政策选择,进而损害公共利益或共同利益。

与量化评估相比,高质量评估是一种低风险的评估。低风险一方面源于高质量评估对于量化指标的超越;另一方面源于对于评估结果的使用。高质量评估将更多的权利(力)赋予利益相关者尤其是高等教育的办学主体。各利益相关者基于责任感和主体性,经过协商和互动共同“创造”高等教育高质量发展的路径。高质量评估虽然也需要对于高等教育的质量进行测量和分析,并基于测量和分析的结果做出价值判断,但这并非高质量评估的最主要的工作。高质量评估关注的主要是“高质量的高等教育”,焦点是高等教育发展的质量而不仅仅是高等教育质量。与质量评估相比,在高质量评估中关于发展的评估需要更加综合的视角,所涉及的因素也更加复杂,需要综合考虑那些可量化的与不

可量化的指标以及其他的相关性证据。与质量评估相比,发展评估的周期相对更长,政府对于评估结果的使用也要更加审慎。当前的质量评估或办学水平评估通常以一年为一个周期,而发展评估则通常以5年为一个周期,甚至还更长。“大学是具有相对惯性的机构,如果每6至8年(甚至10年)对它们的真正变化进行一次观察,会更有现实意义,同时也节省了成本。”^[39]当然,高质量评估的低风险并不等于无风险。实践中没有无风险的评估。无论何种评估,由于关乎理性判断和价值选择都必然伴随一定风险。无风险的评估是无法实现也不值得实现的理想。问题的关键在于,与量化评估相比,高质量评估中的风险是一种理性的可以预期的风险,也是一种为了高质量发展而值得冒的风险。

本质上,评估的风险与对评估结果的使用相关。教育评估通常有两个目的:问责和发展。问责目的:关注学校;关注学校组织机构;数据驱动,强调结果;定量导向;对于如何实施变革策略缺乏知识;更关注学生学习成果的变化;更关注某个时间点的学校;基于研究知识;关注效能较高的学校;静态导向(学校当前的状况)。发展目的:关注教师;关注学校过程;改变效果的实证评估;定性导向;只关注学校变化;更关注学校提升的过程,而非结果;更关注学校的变化;聚焦于实践者的知识;关注学校如何变得更高效;动态导向(学校过去或未来的状况)。^[40]高等教育评估中的风险主要与问责有关。量化评估通常以绩效为杠杆,强调评估的问责功能。在高利害的量化评估中无论个人、组织还是系统都倾向于“自保”,从而使得评估难以获得客观的真实的可靠的信息,不利于促进高等教育发展。实践中由于缺乏客观真实可靠的信息,那些由政府发起的量化评估经常陷入评估主义的陷阱;由于各种“变通”和“妥协”的普遍存在,评估的结果经常不是发现了高等教育系统存在的问题或缺陷,而是“证明”了

“优秀”。与量化评估主要强调问责不同,高质量评估更加强调发展功能,即以高质量评估促进高质量发展。高质量评估的主要目的不在于对高等教育发展的现状做出判断,而是为高等教育发展的未来提供愿景。当然,强化评估的发展功能并不意味着高质量评估不需要问责。政策实践中,没有问责就没有发展。高等教育评估也同样如此。若完全没有问责评估就会流于形式。关键是要保持评估的问责功能和发展功能之间适当的平衡。评估不能“为问责而问责”,而是要“为发展而问责”。以发展为目的,“精心设计有效的问责制时要注意以下七个核心要素:问责制的原因及其预期目标(预期结果);使用的绩效衡量标准正确有效,并且应该尽可能采用多种而非单一的衡量标准;设计符合预期目标的体系(例如,使用状态、改进或增长指标);产生处罚和奖励的后果,并监测其有效性;向学校、学校提供者和公众说明问责制及其结果以及它们的局限性;国家支持学校改进,并评估问责制是否有利于高质量的教学;定期评估、监测与改进体系”^[41]。

除问责与发展的目的不同,量化评估和高质量评估还有一个重要区别。量化评估倾向于为所有高等教育系统、所有大学提供“一个最好的”模板,这个最好的模板就是每个排行榜上的第一名,那些非第一名的高等教育机构或系统的发展方向就是向第一名看齐,而第一名的任务则是保持现有地位。这是一种非常矛盾的逻辑。由于强调机构之间的可比性和竞争性,量化评估的结果基本上是“零和博弈”。部分量化评估项目从高等教育的复杂性出发有时也会考虑分类评估或分学科评估,但量化自身的逻辑总是会导向“化简”和“简化”,以求得出一个总的排名。某种意义上,量化评估是与工业化大生产相适应的。在工业化大生产中,一个最好的模式具有最大的适应性。但在基于知识的创新经济中,个性化、多样化、差异化和网络化成为主

流。在高质量发展阶段,伴随工业经济(社会)向知识经济(社会)的转型,规模和效率不再是关键,质量和创新成为高等教育发展的重中之重。高等教育的高质量发展就是为了适应新的经济和社会发展的需要。此时那种追求最佳实践的量化评估则很难持续。“如果只依赖量化标准,缺乏耗时良久的质化判断作为补充,会使责任的评估产生重大偏差。”^[42]基于此,评估范式的选择体现了高等教育最为基本的价值观和目的。选择什么标准来衡量高等教育发展的质量或水平极为重要。因为这些标准一旦被社会接受就会反作用于并左右高等教育发展的方向。如果为了操作的便捷或数据的可比,而只选择容易量化的指标对高等教育或大学进行评估,会对高等教育体系造成重大伤害。与量化评估相比,高质量评估需要具有更大的包容性和综合性,只有丰富多样的质化和量化标准才能保护并促进高等教育生态系统的多样性。相较而言,如果说量化评估的目的是挑选出一流的高等教育机构,那么高质量评估的目的则是构建高质量的高等教育体系或生态系统。

四、从量化评估到高质量评估

量化评估向高质量评估范式的转变,既受评估范式演变的内在规律制约,也受高等教育自身的发展范式和价值观影响。和高等教育从高速增长向高质量发展阶段转变相适应,关于高等教育质量和发展的评估也可以分为两个阶段。那种基于测量的量化评估可以称之为高等教育评估的初级阶段,其目标是以简单的数字对于高等教育的事实状态进行量化描述;在高质量发展阶段,评估也需要从量化评估走向高质量评估。高质量评估意味着评估并不只是对“高等教育的事实是什么”、“高等教育应如何发展”进行量化描述,而是对高等教育发展提出有意义的解释,并尝试通过评估来更好理解和建构高等教育发

展的制度环境。“评估结果并非终极意义上的‘事实’，而是由包括评估者（以确保客观公正）以及由于评估而处于风险之中的利益相关者通过互动而实际创造的一种结果。”^[43]换言之，量化评估只是尝试反映或描述高等教育发展的现实或高等教育质量的现状，通过给出一系列结论、建议或排行榜来影响高等教育实践，而高质量评估则需要建构一种利益相关者协同机制，共同参与创造高等教育质量或高等教育发展质量。高等教育发展中从量化评估到高质量评估的转变可以理解评估的“范式革命”，抑或科学评估范式向哲学评估范式的转变。

与长期以来量化评估主要考虑评估指标的选取、权重的分配、数据的采集以及结果的呈现等不同，高质量评估更加关注影响高等教育发展的内外部因素及其与高等教育发展过程和结果之间的逻辑关系。立足高等教育的高质量发展，评估不应沦为“数字游戏”，评估结果与高等教育发展之间不应只是一种可选择的“事实关系”，即不同的评估指标可以呈现不同的评估结果；高质量评估不只是评估者对评估对象的评估，也是被评估对象的自我评估，同时也是诸多利益相关者通过协同机制共同参与高等教育发展的过程；高质量评估追求的不是评估结果的精确性或科学性，评估的目的也不只是为了院校的质量管理或政府的科学决策，而主要是为了展示或呈现评估本身与高等教育发展之间的逻辑关系，即“以高质量评估促进高质量发展”。具体而言，一方面，高质量评估要揭示哪些因素可以促进高等教育发展，以评估促发展；另一方面，通过将评估行为与高等教育发展整合或融合成为一个不可分割的整体，其本身也是高等教育高质量发展的一部分。

虽然在理论上从量化评估向高质量评估转变有其合理性，但实践中从量化评估向高质量评估转变并非易事。在旧的范式下，量化评估符合管理主义和科学主义的规范，其

以科学方法论为基础也有现实合理性。此外，范式作为一种基本信念体系，反映了某种固有的偏好。这种偏好不是基于客观的证据，甚至也不是基于理性的分析就可以轻易改变。量化评估的支持者并非不清楚量化评估可能存在的缺陷或弊端，而是认为量化评估虽有缺陷或弊端，但仍不失为一种最好的评估方法。在这种情况下，单纯强调量化评估的缺点或高质量评估的优点都不足以推动高等教育评估范式的转变。从评估研究者的社会化来看，新的评估范式的普及大致需要经历四个阶段：第一，专心于支配性范式。一个人如果不理解旧范式，那么他/她也不可能理解新的范式；第二，发现或意识到支配性范式中的问题，并且承担起解决因这些问题而产生的冲突的义务；第三，通过列举具体的对旧范式的批评意见，以及新范式所受到的广泛支持而强化范式转变的义务，发扬革新的态度和学习新范式方法等形式来解决冲突；第四，在进行新的范式的研究过程中应当更贴近实践和生活；通过反馈循环、修正、与拥有相同观点的人交流的方式来保证研究的进行。^[44]上述四个阶段之间不是线性的或递进的关系，而可能会循环往复。理论上评估的旧范式与新范式截然不同，但是，实践中新旧范式之间并没有明确的分界线，而是存在着大片的模糊地带或者灰色地带。很多评估的研究者和实践者会在两种范式之间徘徊。高质量评估作为新范式的胜出既有赖于其优点的显现，也有赖于量化评估这种旧范式因为自身弊端的难以克服而逐渐退出评估市场的竞争。

当前，高等教育发展中量化评估仍然占据绝对主导地位，属于支配式范式。虽然有研究者和实践者已经发现或指出量化评估中存在的种种问题，并尝试发展新的评估范式；但受高等教育自身发展阶段的制约，即便有政策层面的呼吁和倡导，新的评估范式并没有付诸实践的机会。在范式竞争中，量化评

估的支持者经常以科学范式曾在其他领域取得的成功为自己“打气”，寄希望于通过不断改进评估的技术和方法以实现科学评估；量化评估的反对者则以量化评估实践中出现的各种有违常识的结果来指责其“不靠谱”，寄希望于政府深化教育评价改革，主动放弃以量化为导向的绩效评价，转向一种更温和、更适当、更符合高等教育发展规律的评估。现在的困境在于，几乎所有人都知道并承认过度量化的评估不好，但对于放弃量化评估之后，高质量评估到底该怎么做大家心中并不是十分有数。某种意义上，就像高等教育高质量发展相对于高速增长是一种政策理想一样；作为量化评估对立面而提出的高质量评估现在也还只是一种学术理想，现实中究竟如何操作并没有成功的范例可循。更大的麻烦还在于，高等教育的发展既不是完全可以量化的也不是完全不可以量化的，很多因素在可量化与不可量化之间。如切斯特顿（Chesterton, G. K.）所言：“我们这个世界真正的麻烦并不是其不合理，甚至也不是其合理。最常见的麻烦是这个世界近乎合理，却又还没到那个境界。生活并非不合逻辑，却是逻辑家们容易坠入的陷阱。生活看上去比实际更严谨一些、规律一些；它的精确性显而易见，但它的非精确性却是隐藏的；它的狂野性伺机而动。”^[45]高等教育评估所面临的复杂性也大致如此。量化评估从科学主义和管理主义出发，放大了可量化的部分，甚至是强行量化、滥用量化；高质量评估为避免量化评估的缺陷必须关注那些非量化或不可量化的部分，尝试通过更加综合的视角、收集更加多样的信息来衡量高等教育发展水平。

以高质量评估为愿景，在理论上甚至政策上强调评估要关注那些非量化的部分，要强化综合评价、过程评价、增值评价等都比较容易；但实践中如何对于那些非量化或不可量化的因素进行可靠的评估，如何能够实现综合评价、过程评价、增值评价才是真正的困

难所在。不过可以肯定的是，当前高等教育从高速增长向高质量发展阶段的转变，为评估范式从量化评估向高质量评估的转变提供了机会、创造了条件。此外，随着量化评估的方法越来越精致，评估结果越来越精确，利害关系越来越广泛，其缺陷暴露的也越来越明显。“一个手段的完美是与它的适应性的权衡——手段越完美，适应性越差。”^[46]量化评估在技术和方法上越完美，其评估结果在接触到高等教育实践时的适应性就越差，因为高等教育本身在技术和方法上不可能是完美的。随着高等教育的高质量发展逐渐成为一种社会事实，高质量评估作为评估的一种理想类型将逐渐被广泛接受。实践中高质量评估也不可能完美，但它无疑更适合高等教育高质量发展。无论从高等教育发展阶段的转向还是从评估范式的转移来看，从量化评估到高质量评估的范式革命都不可避免。需要注意的是，在范式层面上高质量评估与量化评估虽不可通约，但这并不意味着高质量评估就必然排斥量化的指标；相反，作为一种方法或技术，量化仍然是高质量评估的一部分。只不过在高质量评估中量化不再一种是主导性和垄断性的技术或方法。在使用量化指标时“我们需要弄清楚我们在计算什么，为何如此计算，谁选择了衡量标准，衡量标准又是如何选择的，衡量者和衡量标准之间有何政治联系，以及这些衡量标准会带来什么样的激励机制”^[47]。简言之，对于高质量评估至关重要的问题是，如何吸收量化评估的优点，同时整合非量化的信息和证据，以促进高等教育的高质量发展。

参考文献：

[1][2][12][13][15][28][30][31][42][45][47] 贾尼比·格罗斯·斯坦. 效率崇拜[M]. 南京：南京大学出版社，2020. 18、236、41、121、88、269、69、233、202、189、236.

[3][4][8][16][29][32][36][38][40][41] 经济合作发展组织. 为了更好的学习：教育评价的国际新视野[M]. 上海：上海教育出版社，2019. 112、1—2、86、30、32、3、12、11、368、564.

[5][9] 玛吉·伯格,芭芭拉·西伯.慢教授[M].桂林:广西师范大学出版社,2020. 51,124.

[6][10] C. S. 路易斯. 荣耀之重暨其他演讲[M]. 上海:华东师范大学出版社,2016. 188,254.

[7][14] 阿马蒂亚·森. 以自由看待发展[M]. 北京:中国人民大学出版社,2002. 251,295.

[11] C.S. 路易斯. 切今之事[M]. 上海:华东师范大学出版社,2015. 185.

[17][20][21][22][23][24][25][26][27][43][44][46] 埃贡·G·古贝,伊冯娜·S·林肯. 第四代评估[M]. 北京:中国人民大学出版社,2008. 2、1、13、5—8、15、131、191—192、165、21、前言·2、45—47、125.

[18][34] O·F·博尔诺夫. 教育人类学[M]. 上海:华东师范大学出版社,1999. 128,48.

[19] 涂又光. 教育哲学课堂实录[M]. 武汉:华中科技大学出版社,2020. 17.

[33] 育人而非制器:杨叔子口述史[M]. 武汉:华中科技大学出版社,2020.

[35] 格特·比斯塔. 教育的美丽风险[M]. 北京:北京师范大学出版社,2018. 1.

[37] 亚伯拉罕·弗莱克斯纳,罗贝特·戴克格拉夫. 无用知识的有用性[M]. 上海:上海教育出版社,2020. 31.

[39] 伊夫斯·金格拉斯. 大学的新衣? ——对基于文献计量学的科研评价的反思[M]. 上海:上海交通大学出版社,2019. 61.

On the High-Quality Evaluation of Higher Education

Wang Jianhua

Abstract: With the development of China's higher education from rapid growth to high quality, quantitative evaluation based on measurement needs to turn to high-quality evaluation. In terms of objects, quantitative assessment focuses on the quality of higher education, while high-quality assessment focuses on the quality of higher education development; in terms of methods, quantitative assessment is based on measurable indicators, while high-quality assessment contains quantitative, qualitative and other relevant evidence; in terms of results, quantitative assessment focuses on measurement, description and judgment, while high-quality assessment focuses on construction and reconstruction based on measurement, description and judgment; in terms of functions, quantitative evaluation emphasizes the use of performance as a lever for accountability, while high-quality evaluation emphasizes the promotion of development through evaluation; and in terms of risks, compared with accountability-oriented high-interest quantitative evaluation, development-oriented high-quality evaluation is low-risk.

Key words: higher education; quantitative evaluation; high-quality evaluation; high-quality development

Author: Wang Jianhua, professor of the School of Education Science, Nanjing Normal University (Nanjing 210097)

[责任编辑:刘 洁]