

# 发展改革动态

2019 年第 10 期 共 104 期

发展规划处

2019 年 6 月 17 日

---

【聚焦内涵发展】

## 中国大学生学业成就评估研究：二十年的回顾（1998—2017 年）

### 一、中国大学生学业成就评估研究综述

#### （一）中国大学生学业成就评估工具及应用

在 20 世纪 70 年代末 80 年代初，美国就出现了一些测评工具，如 McBer 公司的“行为事件访谈”，美国 ACT 和 ETC 关于通识教育的测评工具等。近年来国际上的大学生学业成就评估工具发展迅速，笔者根据评估内容和测评对象的不同，整理出了各国研究者们开发的 27 种测评工具。

美国最多，主要包括合作性的院校研究项目—新生调查（Cooperative Institutional Research Program—Freshman Survey, CIRP-FS）（1966）、大学生期望问卷调查（College Student Expectation Questionnaire, CSXQ）、大学学术能力测评（Collegiate Assessment of Academic Proficiency, CAAP）（1988）、学业表现和进步测量（Measure of Academic Proficiency and Progress, MAPP）（2006）、大学生就读经验调查（College Student Experience Questionnaire, CSEQ）（1994）、大学生调查（College Student Survey, CSS）（1993）、全美大学生满意度调查（National Student Satisfaction Study, NSSS）（2000）、学院学习结果评估（Collegiate Learning Assessment, CLA）、美国大学生学习经历调查（NSSE）（2000）、

研究型大学就读经验调查 (SERU)、CIRP 大四学生调查 (CIRP College Senior Survey, CIRP-CSS) (1992)、全美大型质量评估项目 (the Wabash National Survey of Liberal Arts Education, WNSLAE 或 Wabash) (2006)、专业领域测试 (Major Field Test, MFT) 以及大学结果调查 (College Results Survey, CRS) 和综合性的校友评价问卷调查 (Comprehensive Alumni Assessment Survey, CAAS) 等。澳大利亚次之, 包括毕业生技能评价 (Graduate Skills Assessment, GSA) (2000)、大学生经验调查 (Student Experience Survey, SES) (2015)、课程经验问卷 (CEQ) (1993)、毕业生就业目的地调查 (the Graduate Destination Survey, GDS) (1974)、毕业生收获调查 (Graduate Outcomes Survey, GOS) (2016) 等。还有英国的全英大学生满意度调查 (National Student Survey, NSS) (2005) 和大学毕业生目的地调查 (the Destinations of Leavers From Higher Education, DLHE), 加拿大的青年工作转换调查 (the Youth in Transition Survey, YITS) 和全国毕业生调查 (National Graduates Survey, NGS), 日本的全国大学生调查 (东京大学创成科研项目, CRUMP 调查) 和日本大学生调查 (山田礼子科研项目, JCSS 调查)。国际经济合作与发展组织 (OECD) 2008 年发起的高等教育学生学业成就跨国评估项目 (The Assessment of Higher Education Learning Outcomes, AHELO) 属于涉及的国家数量最多的项目。

其中, 评估内容又有认知和非认知的区别, 前者一般分为共同性知识与领域特定的知识。如美国的 CAAP、MAPP、Wabash 项目测量的就是通识教育教学成效, CLA 测量的是特定领域的学习结果, 而 CEQ、NSSE 主要关注非认知学习结果和一般性能力, MFI 则是两者兼而有之。从测评对象来看, 大多数是关注过程中的大学生就读经验和收获, 如美国的 CSEQ、NSSE、SERU, 澳大利亚的 CEQ、GAS, 英国的 NSS 等, 而 CIRP-FS 和 CSXQ 关注的则主要是新生的基本信息和就读期望。同时, 一些测评已经关注到校友层面的增值和发展, 如美国的 CRS、CAAS, 澳大利亚的 CEQ、GDS, 英国的 DLHE, 加拿大的 YITS 和 NGS, 等等。还有两个测评工具比较特殊: 一是日本的全国大学生调查, 既关注在校生也关注校友; 一是 OECD 发起的高等教育学生学业成就跨国评估项目 (AHELO) 测评多国的大学生增值情况, 可以是一般性能力, 也可以是专业能力。

上述这些测评工具之间存在一些“互相借鉴学习”(或称为本土化)的情况, 如英国的 NSS 就是在澳大利亚的 CEQ 基础上修订而成, 美国的 NSSE 原版调查问卷的 2/3 来自 CSEQ 的

问卷内容，而日本的全国大学生调查参考了美国的 NSSE 调查。之所以花了很大的篇幅先介绍国外的测评工具，是因为发生在英国和日本的这种本土化进程也同样发生在中国。

国内方面，自从有学者开始讨论大学生就读经验，始于 2007 年左右的学业成就评估蓬勃发展，逐步形成了多足鼎力的第三方机构评价局面（见表 1）。主要有北大项目、清华大学的 CCSS、北京师范大学的 CCSEQ、南京大学发起的国内 SERU 调查项目、厦门大学的 NCSS 等。有研究者还提到了北京市学习科学研究会杜智敏教授主持的“北京市大学生学情调查”（2002），山东理工大学谭秀森负责的“当代大学生学习环境与研究”（2007），复旦大学的“上海大学生发展研究”（1998）等都是学生学习与成就相关的测量工具，这些测评工具尽管不完全是建立在学业成就评估基础上的，但依然对了解学生学业成果和发展有重大意义。国内的测评工具全国性调查与地方性调查并存。当然也有很多高校尝试自主研发测评工具主要服务于各自的院校研究，如中山大学，可以称之为“校本模式”。

表 1 国内主要的大学生学业成就评估工具一览

序号	发起者机构及时间	评价工具	借鉴测评工具	测评对象	主要负责人
1	北京大学, 2006	全国高校教学质量与 学生发展监测项目	美国的 NSSE、CSS, 日本的 GHUMP	在校学生和毕 业生	鲍威团队
2	清华大学, 2009	中国大学生学习与发 展追踪调查 (CCSS)	美国 NSSE	在校本科生	史静寰和罗 燕团队
3	北京师范大学, 2007	中国大学生就读经验 问卷 (CCSEQ)	美国 CSEQ	在校生	周作宇和章 建石团队
4	南京大学, 2011	研究型大学学生参与 调查 (SERU)	美国 SERU (或 UCURS)	大二、大三、 大四	龚放和吕林 海团队
5	厦门大学, 2011	中国大学生学习情况 调查研究项目 (NCSS)	自主研发	在校生	史秋衡团队
6	中山大学, 2012	中山大学本科生学习 情况调查	自主研发	在校生	邱琼莹团队
7	西安交通大学, 2004	课堂学习环境与学生 发展关系调查	数学学习经验量表 (ESM)、比格 (Biggs) 的学习过程问卷 (RVLPQ) 和澳大利亚 CEQ	大一学生	陆根书团队

参与测评院校数量而言，以北大项目和清华项目（CCSS）为例，北大项目自 2006 年开始 6 所学校参加，到 2008—2015 年每年参加的首都高校稳定在 50 所上下，十年参与人数超过 30 万人次，2017 年开始推广到全国，参加院校近 100 所。清华领衔的 CCSS，2009—2015 年参加的院校累计超过 100 所，人数超过 30 万次。厦大的 NCSS，2011—2014 年期间每年参加的院校在 50 所左右，参加学生人数亦超过 30 万人次。据估算，2017 年选择参加三个项目的院校数量，预计超过 200 所，占全国高校近 1/10。另外，南京大学、西安交通大学、湖南大学、同济大学参与了 SERU 项目。北京大学也在考虑加入。国内参与学校尽管较少，但可以与美国的 17 所顶尖大学（如加州伯克利大学）、欧洲（如牛津大学）和南美的 6 所顶尖大学、韩国首尔大学等近 30 所大学形成跨国的数据库。在“双一流”建设政策背景下国内大学可以更好地与国外大学在相同指标上进行比较，看出各自大学的长处和短板。国际上测评工具绝大多数都有较好的信效度，在国内学者引入和本土化过程中也进行了信效度检验，相当一部

分在研究中都呈现了较好的结果；多数学者使用的是克隆巴赫  $\alpha$  信度系数或者项目反应理论边际信度，探索性因子和验证性因子验证效度。

## （二）中国大学生学业成就评估测评方法

大学生学业成就评估方法通常采用定量方法：根据信息来源不同，可以分为直接估算法和间接估算法；按照测评学生是否为同一批，可以分为横向评估法和纵向评估法。测评的数据分析方法也各有不同。直接和间接估算法。直接估算法，通常采用直接对比的方法，在大学入学时对学生进行标准化的测试，或者在学生的入学考试成绩基础上预测大学学习结果评估（the Collegiate Learning Assessment, CLA）的成绩作为基础值，在就读的某一阶段，通常是在学生毕业时，进行类似测试以确定结果，两个得分的差异即可视为学业成就的提升。国际上，使用直接估算法的测评工具主要有美国的 CAAP、MAPP、CLA 及 MFT，澳大利亚的 GSA 等。间接估算法则通过学生自我汇报的方法确定学业成就提升的程度。国际上，美国的 NSSE、NSSS、CSEQ、SERU，澳大利亚的 CEQ 即为代表，国内方面，北大项目、清华的 CCSS、北师大的 CCSEQ、厦大的 NCSS 等都属于自我汇报法。

横向和纵向评估法。横向评估法对一年级新生和四年级毕业生同时进行测试，入学新生被视为控制组，其当前状况被近似地等同于现有毕业生的当时入学状况，新生和老生之间的差距就被标识为“大学在学生的学习经验的发展上所起的作用和作出的贡献”。这种评估方式的问题在于，当前的新生状况不能完全地等同于老生入学时的基准状况，诸如转学或辍学学生、招生标准等因为时间因素所导致的诸多变化和影响没有被考虑进来。这种缺陷，可以通过纵向评估或跟踪评估得到很好的解决。纵向评估被用来测量同一组学生在时间发展序列上的学习结果变化，这种方式保证了对入学时的各种输入变量的良好控制，美国的 CAPP、MAPP 和 CLA 即采用此法。构建通识教育的理想教育产出的全美大型质量评估项目（WNSLAE，简称 Wabash）和澳大利亚毕业生收获调查（Graduate Outcomes Survey, GOS）亦采用了跟踪调查。国内北大项目从 2011 年也开始加入纵向评估。尽管如此，纵向评估仍然因为其外部环境因素的不可控性、样本的大幅流失和天花板效应等原因存在着一些局限性。最理想的解决办法以加拿大青年工作转换调查（YITS）为代表，选择一组没有进入大学的高中毕业生，他们与进入大学的且接受测试的学生群体在个人特征和学术背景上基本相同，并将该组学生作为控制组进行评估。

数据分析的技术层面，主要有多元线性回归模型（MLRM）和分层线性模型（HLM）。通过多元线性回归模型得到的残差值，即是大学生在一段时间内学习进步的增值；分层线性模型可以将影响学生学业成果的外部因素（如学生的学习基础、家庭背景等）与学校或教师的效应分离开来，得到学校或教师的“净效应”，从而判断学校对学生的影响力。国内北大项目和清华的 CCSS 采用的均是多元线性回归模型。

综合国内外的测评方法，可以将大学生学业成就评估方法归纳为一个矩阵（见表 2）。国际上早期处于第二象限、第三象限，但现在测评工具多数处在第一、四象限，不管是直接估算法还是间接估算法的测评工具都能很好地实现跟踪调查。而国内主要采用的间接估算法，直接估算法的测评工具研发相对比较落后，大多数院校采用的是横向评估法，很少能像北大项目和清华 CCSS 做到长期跟踪调查。因此，我国大学生学业成就评估方法应该向第一、第四象限努力，如果能够直接将直接和间接估算法相结合并且做到长期跟踪，则效果更佳。

表 2 大学生学业成就评价方法组合一览表

测评学生选择信息来源	横向评估法	纵向评估法	数据分析法
直接估算法	第二象限：直接估算法—横向评估法 (国际上早期)	第一象限：直接估算法—纵向评估法 (国际上当前多数)	多元线性回归模型和分层线性模型
间接估算法	第三象限：间接估算法—横向评估法 (国际上早期，我国多数)	第四象限：间接估算法—纵向评估法 (国际上当前多数，我国少数)	

### （三）中国大学生学业成就评估结果使用

理论上，大学生学业成就评估的结果是用来指导教育教学改革和实践。Astin1985 年就谈到它的功能：学生的前测表现出的优缺点可以用来作为学业建议和职业咨询的依据，后测中对学生成长的性质和程度也可以为学生、教师和管理者所用。对于学业成就评估结果的使用可以从教育行政管理、学校（含教师）和学生三个角度分别阐述。

国际上，对于教育行政管理而言，教学行政部门可以通过 CSEQ 对学校常模的分析，把握学校的实际教学状况，为高校教育质量的评估、认证奠定现实基础，并从根本上提高对高校教学质量干预的科学性。这种“干预”在有些国家是作为资源分配的依据，如澳大利亚政府通过 CEQ 结果对各个大学的教学质量进行排名，排名结果将作为大学拨款的重要衡量指标。在有些国家结果使用却是间接的，英国政府尽管会通过 NSS 测评结果，掌握高校教学工作和人才培养的基本情况，并可以此作为对高校进行质量问责的依据，但坚持“有限使用”的原则，并不以此来对高校的教学工作水平进行排序和奖惩，也不直接以此来干预高校内部的教学质量保障工作。

在学校方面，对于美国的院校，评价结果除了可以衡量内部教学的质量并进行恰当的干预外，还可以向认证机构表明教学工作的成效，满足认证的需要。同时，也可以通过横向比较和纵向比较，发现自己多年来的进步和问题，以及与同行相比的优劣势，便于更好地自我定位。对于教师而言，学生对教师个人教学效果的评估结果是保密的，但会以恰当的形式告知教师个人，并附以改进的意见。

在学生方面，除了 Astin 提到的作用外，美国 CSEQ 官方认为还可以为学生自我反思和评价提供非常有价值的信息，促使他们思考学习的经历及取得的进步；英国的准大学生还可以根据 NSS 提供的质量信息来选择高校与专业；澳大利亚的 GSA 给参与评价学生求职时提供其课程申请简历与 GSA 得分，提供评估院校新生识别和研究生增值信息。

相比较而言，国内大学生学业成就评估结果使用的系统性研究较少。即使产生的影响也是以学校和学生为主。与美国大学生学业成就评估结果可以满足认证的需要不同，在中国对于高校的这种制度性的内部动力并不存在，不过审核性评估为此类似的动力提供了可能性。将本科教学评估结果与大学的资源分配相挂钩的方法在国内一般而言属于政府主导的教育教学评估的功能范畴，而非民间发起的大学生学业成就评估项目。再者，有些项目负责人明确表态需要抵挡住排名的诱惑和弊端，否则数据的真实性将会大打折扣。

一般而言，国内的大学生学业项目的研究结果使用一般交给参与项目的院校自主处理。对于学校而言，理论上可以从学生个体层面监控、保障教学质量，还可以根据需要形成院系、年级、专业等类别的常模，把握各院系、年级、专业的学生整体的学习的动态变化，有针对性地改进教学，不断提高教学质量。而实际上，有关大学生学业成就评估结果对于学校的发展和教师的教学如何影响的研究凤毛麟角。清华大学的 CCSS 研究发现，清华教师对学生的支持度不够，建立了学生辅导中心，参与 CCSS 项目的一些院校也存在类似的问题，也建立了相应的机构；从学生的满意度也可以看到学生对教学、环境等方面的精准需求。也有研究表明，南京大学将 SERU 评估工作作为“三三制”教学改革的一驾马车之一。其项目介绍人所言，SERU 项目扮演了孵化器的角色，引发了一系列类似的实证调研，逐渐在该大学形成了证据文化。然而，对于参与这些大学生学业成就项目的院校而言，对于评估结果的重视程度本身也存在差异。

对于学生而言，像国外大学学业成就评估信息作为学生选择大学、就业或继续升学方面

的功能尚未在国内已有研究中发现。他们在大学生学业成就评估中的角色仅限于填写问卷提供自己的个人信息，至于个人的实际调研结果到底意味着什么很难得知。很多时候，大学生学业成就评估结果可以确定在少数部门、少数群体内部共享，而对于到底如何影响了本科教育教学改革，可以说是一个黑箱。

### 三、中国大学生学业成就评估存在的问题

#### （一）评估工具：标准化、本土化和多样化程度不够

首先是标准化问题，研发测评工具需要非常强的专业性，而且周期很长，即便高校自己可以解决这个问题。完全自主设计的测评工具可以反映院校具体的课程重点和前瞻性风格，却更容易变化或缺乏稳定性，而且无法形成校际、省际、国际之间的比较。因为各自依据的理论框架不同，甚至没有理论框架，最后的数据结果只能在本校进行分析，无法在同类学校之间形成比较。如此学校就没法定位自己在推动学生学业成就提升方面究竟处于怎样的位置，很容易在剧烈的高等教育改革大潮中迷失方向，失去机会（美国、韩国、日本有相关的法律要求，也许可以借鉴）。

其次是本土化的问题。清华、南大分别在 NSSE 和 SERU 基础上进行改进或微调，前者还增加了毕业生的跟踪调查。厦大、中山大学主要是自主开发工具，而北大的首都高校教学质量与学生发展评价项目兼顾了本土化和国际化的双重特征。

再次是单一化问题，即测评工具选择单一化，如未参加国内几大测评工具联盟的院校仅选择自己开发的工具，或者即便是选择了国际化的测评工具，也存在的问题在于每个测评工具都有不同的侧重点，如 NSSE 和 Wabash 项目比较倾向于通识教育，而澳大利亚的 CEQ 和 OECD 主导的 AHELO 比较全面。比较理想化的方法是，选择至少两种测评工具。参与美国 Wabash 项目的成员学校学生除了填写该项目的调查外，还被要求填写 NSSE 和 CESQ 的问卷，这将大学教育产出的测量和学生学习过程的测量紧密地结合在一起。

然而，在工具的选择上面，无论选择哪种方案，院校都需要明白测评的共性和个性相结合的问题，前者倾向于通过院校联盟的方式采用同一种测评工具，而后者则针对本校特色的校本测评工具。国内也开始出现一种模式，即参与北大项目的院校可以选择除了参与统一的测评外，还可以委托项目组提供“定制服务”，目前已有两所院校进行了实践。

#### （二）评估方法：过于依赖间接估算法和横向评估法

直接估算法和间接估算法各有优缺点，目前国内的问题在于集中在使用间接估算法，即依靠自我汇报法获得数据，这种方法适用于非认知变量的测量，但对于认知能力的测量则需要直接估算法，如美国的 CAAP、MAPP 测量的通识能力。估算法的选择背后是评估内容的选择，这是理论的本土化问题。选择间接估算法意味着认知变量无法很好地兼顾，对于测量学生学业成就来说是不全面的。

而在横向评价法和纵向评价法或跟踪评价法中，很显然纵向评价法对于同一组学生的测评能够控制可能影响学业成就提升的输入变量，可以更好地解释增量。横向评价法的局限是只能进行学生发展的描述性分析和解释性调查研究，无法动态地探究高校人才培养质量与学生发展的变化趋势和因果关系，也无法充分考虑高校学生不同学年过程中的发展和行为特性等。选择横向和纵向评价法的差异核心在于，研究者认为不同时期学生的特征是否一样。美国几十年的 CIRP 项目研究证明，“不同时期学生的特征是有差异的”。国内测评方法比较显著的问题，在于主要实施横向评价法，原因可能在于跟踪研究（follow-up studies）的难度较大和样本丢失问题。然而，如果不是建立在面板数据的分析基础上，所谓的纵向比较难以经得住考验。学校决策者就没法获得确切的证据来了解自己的本科教育是否获得了进步以及获得了多大的进步，因为研究结果无法剥离学生群体之间的能力差异。

### （三）评估结果使用：信息透明度不够

关于质量的信息应该对公众开放到何种程度？这个问题是有争议的。然而，国内有的院校可能还停留在“信息是否公开”的问题上。可以说评价结果使用几乎是黑箱。中国大学生学业成就评估项目负责方通过无偿或有偿的方式为参与院校撰写调研报告供其推进本科教育教学改革作为参考，是否使用以及如何使用完全取决于院校自身，这是市场所决定的。因此，可以说，不只是评价结果是否到达院级二级单位，而是在学校层面是否很好地接受和消化的问题。也就是说，测评工具的功能使用面临虚化和悬置的尴尬境地。另一个问题是，问卷调查是为教育教学改革服务，但实际情况可能恰恰相反，“某些决策者不是通过调查以及由此得来的数据来了解学校的总体情况和学生对于学校运作的基本看法和评价，而是将调查当成是一种应急的工具”。

为什么学业成就评估结果使用处于黑箱？可能存在以下几个原因：第一，传统公立院校的绝大多数经费来自中央部委或地方政府，故由上级管理部门发起的评价，要拿到资源就必



须按照教育质量评价的一套制度来办学，资源依赖理论（Resource dependence theory）已经能支撑这个观点，显然，学业成就评估的实施并不能帮助学校直接获得经费，反而要使用经费。第二，评估属于部门发起者“自导自演”，未能调动全员。收益没有考虑清楚，学业成就评估往往成为学校本科教育质量成绩的依据，而不是促进教育教学质量的工具，目的和手段本末倒置。师生未能充分地参与，学生参与的形式停留在填写问卷和提交问卷上，甚至都无法获知自己填写的结果到底意味着什么。从组织层面来说，院系作为学校的二级单位，也是被动参与自我评价。参与评价主体缺失严重的表现之一还在于，教师没有参与到测评工具的开发和完善中来，如商用测评工具关注的通识教育是否是教师认为的呢？第三，院校的学业成就评价并未成为常规的本科教育质量的的活动，带有一定的偶然性，上级决策层觉得很重要，那对于结果相对重视，即便如此，信息公开非常有限，特别是如果结果不利于学校声誉，那更难对公众开放结果。第四，就国内而言，政府有自己的一套评价制度，而学校发起学业成就评估是研究范式与国际接轨所致，我们政府的问责和公众的问责尚属于“同床异梦”，或者说，政府问责是强势，而公众问责是弱势，强势一方对于学业成就评价还不能真正用于其重大决策，其质量评价实际上还停留在所归纳的“四种传统质量观”。简言之，国内这类评价结果使用缺乏强有力的内在动力。

除了上面讨论的学业成就评估工具、评估测评方法的选择与使用、评价结果使用等方面的问题以外，评估团队的专业化等问题也是需要院校推进和解决的实际问题。

### 三、中国大学生学业成就评估的展望

学业成就评估已经是国际上发达国家非常认可的质量问责的途径，我们的任务在于如何让它更好地服务于提高国内的本科教育质量。目前由北大、清华、南大、厦大等高校发起的评价联盟已经让人们看到了更大进步和更多希望，初步形成了北大模式、清华模式、南大模式、厦大模式和校本模式，究竟其各个模式特征如何有待进一步研究。笔者提出如下未来需要努力的方向。

#### （一）建立学业成就评估系统性的制度

上文讨论的评估内容、评估方法、工具选择与使用，以及结果使用都属于学业成就评估制度的内部组成部分，未来院校还需要将这些组成部分通过系统的方式有机地形成一个整体。澳大利亚从2015年开始进行了一系列的调整，将原来相对分散的测评工具通过 QILT 的项目

方式联系在一起综合衡量学生学业成就。也就是说，澳大利亚高校的质量保障形成了“组合拳”，其理念是教育增值。国内大学将学业成就评价作为单项或专项的本科教育质量评价，已经远远不够。

## （二）建立新的学业成就评估范式

对过去二十年间的评估实践略加考察我们便会发现，高校评估所面临的问题远不是挑选一两个评估工具或创建一个数据库就能解决的，高校必须首先在更深的层面上回答传统范式和新范式之间的争论和整合问题。传统的质量保障环节看起来是闭环管理，但信息完全的假设是不成立的，导致结果不是“O”型，而是“C”型。程星提出的新范式是“三级金字塔”：

（1）对直接参与教学活动的团体负责，负责将学生学习的成果交代清楚；（2）对直接和间接支付教育费用的团体负责，负责将他们投资高等教育所能得到的回报交代清楚；（3）对直接和间接管理与规范高等教育的团体负责，负责将高校的社会效益交代清楚。在这个三级范式中，评估活动根据信息接收端的需求展开，并且上一级的评估总是建立在下一级评估所提供的信息基础之上。

该新的范式最大的进步是利益相关者理念再次得到体现。然而，根据对我国特有文化的了解，传统范式和新的范式都面临信息公开的问题。将学业成就评价结果等信息对公众公开估计还需要漫长的道路，当前比较有希望实现的则是“内部科学有效使用”。在笔者看来，这种新范式的形成还需要漫长的过程，比较务实的做法还是倾向于在传统范式上的微调，尤其是通过立法对于信息的公开进一步推进。时机成熟再考虑向程星所提倡的新范式过渡。

## （三）学业成就评估和其他学习结果评价结合使用

评估的多元化包括两种情况，一种是我们需要采用多元化的测评工具，典型的澳大利亚的学习与教学质量指标（QILT）包括学生经历调查（SES）、毕业生雇佣（GOS）、毕业生满意度（CEQ）和雇主满意度（ESS）来综合衡量学生学业成就。另一种是指学业成就评估和其他学习结果评价结合使用形成的更大结构的多元化。尽管我们讨论了学业成就评估的种种优点，但并非否定传统的结果评价的方式，如学业成绩、毕业率、就业指标等，这些多元的指标在评价内容的探讨部分已经可见一斑。当然，需要提醒的是，关于学习结果评价数据的统一性（和真实性），我国高等教育主管部门则需要努力改进。否则省际、校际之间就无法实现比较，也就失去在横向和纵向比较中进步的机会。在数据端口的统一性方面，美国的做法值得借鉴，

如美国高等教育委员会向大学推荐 CLA 和 NSSE 两个测量工具，自愿问责系统（VSA）允许参与的院校使用 ETS 成绩包、CAAP 和 CLA 三种测量工具。尤其需要注意的是，高校如何向相关利益群体证明数据的真实性和可靠性，也是另一个令人头疼的问题。

#### （四）对现有学业成就评估进行元评价

所谓元评价（meta-evaluation），就是对现有评价的评价，即“按照一定的理论框架和价值标准对教育评价所进行的评价”。“国内做学业成就评价已经做了很多年，研究成果也那么多了，那么效果到底怎么样？还是需要有相关研究关注。”通过评估才能发现现有的评价哪些部分是值得保留的，哪些是需要改进的。如英国高等教育研究院组织专家于 2006 年年底对质量信息收集工具——NSS 的影响和成效进行评估。其评估结果既认可了调查本身对于高校教育质量保障的作用，也发现在行动方面真正付诸实施的高校也不多。国内的学业成就评估可以评价结果使用的跟踪作为起点，搞清楚结果是如何被学校和二级单位使用的，在实际的整改方案中是否体现了专家组的意见。在一定时期内，通过一定的评价标准或工具来评价整改的效能。

#### （五）学业成就评估理论和实践的本土化

谢立中归纳了四种社会科学本土化类型，即对象转换型本土化、补充—修正—创新型本土化、理论替代型本土化和理论方法、全面替代型本土化，这里的分类只是事实判断，并无价值判断。所幸，我国的专家学者早已认识到本土化的重要性，北大项目、清华的 CCSS、厦大的 NCSS 等项目测评工具方面的本土化则可见一斑。早期参与北大项目的卢晓东研究员的观点则具有代表性：“中国和国外的教育发展阶段不一样，所以国外的大学问题有些和中国不是一回事。所以这些量表也不可能问的是中国的问题。评估量表应该问中国的问题，反映教育管理人员、教师和学生的问题，应该用中国的教学管理老师、教学老师和学生的话来编制，框架可以借鉴，必须本土化。”目前，中国大学生学业成就评估的对象转换型本土化、补充—修正—创新型本土化形式已经基本实现，第三种有关理论方面的本土化，则处于较为缓慢的阶段，主要还是借助于西方的院校影响力理论，但这也是最难的，需要一个发展的过程。我国高等教育进入后大众化发展阶段中，涌现出很多特有的现象和问题，需要更具有解释力的本土理论指导实践，以学业成就评估为契机提高我国高等教育的竞争力。

#### （六）评价部门人员进一步专业化

评价部门人员的专业化是院校开展包括学业成就评估在内的各种评价活动的基本性前提，尽管如今的社会机构测评比较成熟，但院校自身应该有专门的研究人员能够甄别工具的优劣，能够解读最终的结果，并能够深入浅出，说服相关高校决策层重视学业成就评估的作用，推动该评估在本科教育质量保障过程中的功能发挥。