

## 编者语：

大学生的发展质量是高等教育质量的最终体现，高等教育学习成果是高等教育质量评估的重要内容。如何测评学生的学习结果才能符合时代对大学生的要求是高校必须研究和实施的内容。本期高教动态特选取 OECD 所提出的高等教育学习成果评价项目的概述、历史演变、特征、组织架构以及评估工具等相关研究，以飨读者。

# 目 录

美国高校“学生学习成果评估”的历史演进	黄海涛
美国高校“学生学习成果评估”的特点与启示	黄海涛
美国高校“学生学习成果评估”的组织架构及其职能	黄海涛 常桐善
美国大学生学习成果标准化评估工具分析	周廷勇等
OECD 国际高等教育学习成果评价项目 ——AHELO 解析	廖 青



# 美国高校“学生学习成果评估”的历史演进

黄海涛

“质量保障”已成为 21 世纪世界高等教育发展的鲜明主题,世界各国都在探寻更为有效的质量管理范式。20 世纪 90 年代以来,美国高等教育高度重视“学生学习成果评估”(Student Learning Outcomes Assessment),即“评估主体运用各种测量方法和测量工具,持续收集、分析有关知识、技能等学生能力增值和情感、态度变化情况的信息,并以这些信息为直接证据,对照学习成果的最初设定,进行评价、判断教育教学的成效,找出学生学习和教师教学及相关学生服务等各个教育环节中的问题并有针对性加以改善,最终实现提高学生学习成果、促进学生个体发展的高等教育质量保障之目的”。

当前,这种注重“产出”的质量管理方式因其“直面学生学习增值、强调教育成效证据”而备受各国关注,并且成为世界大学教育质量保障范式的变革趋势。本文着力于在美国高等教育发展的宏观背景下,统筹考虑重要的影响因素和重大的历史事件,勾勒出学生学习成果评估的基本发展脉络及其大致全貌图景以呈现其历史的“实况”,以便我们更加全面、准确地理解美国高等教育中的这一质量保障方

式,更好地为我国改进大学教育质量保障体系提供有益的借鉴。

## 一、口试、论文等早期综合考核: 1636~1904 年

美国高校“学生学习成果评估”可以追溯到 17 世纪 40 年代的哈佛校规和 17 世纪 50 年代哈佛学院的第一次毕业课,那时候以学生演讲辩论为主要形式的口试已经非常普遍。比如,哈佛学院 1646 年的校规已明文规定:学士学位候选人能够用拉丁文读出新约全书和旧约全书的原文,并能按照逻辑解答其中的问题,在公开答辩中,得到董事会、教师的认同,方可被授予第一级学位(学士学位)。哈佛学院 1678 年毕业生科顿·马瑟(Cotton Mather)的论著详细叙述了当时哈佛学院对学生学习进行评估与考核的情形:学生要参加每周一次的演讲会,还要参加由校长或教师主持的公开辩论会。临近毕业的学生在长达 3 周的“考核周”(Weeks of visitation)内必须接受语言能力和学科知识方面的考核,学位候选人要在地方官员、牧师和绅士们前来参加的授予学位典礼上公开口头答辩自己的论文,通过答辩的学生在典礼末尾获得校长颁发的学位

证书。可见,在美国殖民地学院时期,特别是18世纪以前,由于学院学生人数很少,学习内容为具有浓厚宗教色彩的古典课程,这一时期的学生学习成果评估的特点也就相应表现为:评估目的主要是检查学生是否能够“回忆起”学习的内容,并能“优美”的表达;评估人员主要是教师和董事会成员,地方官员、牧师和绅士们等校外人士起到一定的见证作用;对学生学习的评估内容主要为文学、修辞、教义等陈述性知识;评估方法主要是口试,学生根据教师的要求进行演讲、辩论等。

18世纪的课程改革拓宽了学生学的内容,学习成果评估的内容也随之有所突破。18世纪早期,由于受到欧洲启蒙运动的强大影响和英国非教派学院与苏格兰课程设置的影响,美国的许多学院进行了课程改革,实用性课程开始出现。到了18世纪中叶,近代科学更加有力地冲击着美国学院的课程和教学,自然科学很快成为美国学院课程的组成部分。在学生学课程逐渐摆脱了经院哲学和宗教神学束缚的过程中,学生学习考核的内容也随之改变,不再仅仅局限于以教义为主的古典课程知识。同在这一时期,美国高等教育随着社会工业化而开始转型,作为对一个班级所有学生进行统一论文考试的笔试方式开始流行,这种笔试方式已远不同于先前的单个学生口试。但从总体而言,在19世纪,高年级学生在获得学士学位的最后学习阶段,主要仍是通过演讲和辩论等综合考试来证明自己确实为进入社会做好了智力上的准备。

## 二、标准化考试源起:1904~1932年

19世纪末20世纪初,随着学生人数

的增长和课程门数的增加,教师难以对每一个毕业生都进行演讲和辩论的综合考试。为了解决这个问题,针对课程设定的学分制开始实行。学分制虽然比综合考试简便易行,但在实施后便面临诸多困境,遭到了一些改革者的抨击,如亚伯拉罕·弗莱克斯纳(Abraham Flexner)认为学分制的实施使得大学学位的质量正在丢失。恰在此时,以1904年爱德华·桑代克(Edward Lee Thorndike)发表《心理与社会测量理论导论》一书为标志的教育测量技术第一次飞跃为如何适应新的形势,有效、高效地对学生学习情况进行科学的评价提供了新的理论视野,标准化的、由是非题和选择题组成的客观题考试作为一种全新的尝试登上了历史舞台且日渐显著。其中,卡内基基金会的改革尝试和宾夕法尼亚研究最具影响。

卡内基基金会对学生学习评估的改革作出了开创性的贡献,其第一任主席亨利·普瑞切特(Henry Pritchett)在“客观题考试”出现之初就认识到这种考试变化在监控高等教育质量方面可能具有很大的潜在影响,“运用客观题测试学生的能力来说明学生在大学里的学习是否成功”这一见解得到了卡内基基金会同仁的认同,引领着学生学习评估的改革实践:1916年,威廉(William S. Learned)在密苏里大学(University of Missouri)“使用经过验证的试题内容、考试程序、评分标准和统计方法,测试学生在数学、拼写、写字、阅读和作文等领域的学习成果。虽然这次客观题测试还不够精确,但却有标志性意义”;爱德华·桑代克(Edward Lee Thorndike)紧随其后,进行了类似测试。

在标准化考试方面另一个具有突出贡献的当属规模浩大的宾夕法尼亚研究(Pennsylvania Study, 1928~1932)。宾夕法尼亚研究的有二:一是了解“学士学位”作为清晰的、可以明确感受到的重要概念到底等同于什么?二是了解“学士学位”作为正确的辨别能力和缜密的思考能力到底等同于什么?为达成这些目的,1928年至1932年,威廉和本·伍德(Ben D. Wood)一起在宾夕法尼亚州进行了一场大规模的大学学习评估研究工作,其特点为:1)测量工具采用客观试题,且题目总量大;2)被试样本数量大;3)测试内容广泛,几乎涵盖大学课程的所有领域;4)采用纵向度的跟踪测试。参加1928年测试的高中高年级学生,到1930年作为大学二年级学生再次被测试,到1932年作为大学四年级学生再次被测,然后进行比较。这项研究认为,学习成就是学习行为的结果,学习行为表现为学生不断增加专业知识的广度和深度,而专业知识应该全面涵盖陈述性知识和程序性知识。研究结果在一定程度上也证实了“如果课程的知识能够被理解,那么客观题测试就能有效地证实知识的累积”这一学习理论假设。宾夕法尼亚研究具有里程碑的意义,特别是在“学生究竟应掌握哪些知识,学习应该如何衡量”这一核心问题上的研究对此后的测试变革产生了重要影响。

物之初生,其形必丑。标准化测试在美国高校中启用虽然还很粗糙,但它确是学生学习成绩评估发展历程中的一个重要里程碑。与早期的口试相比,标准化测试扩大了评估规模,评估模式从个体评估转向集体评估,测试的效率也随之大大提

高;与使用广泛的论文考试相比,标准化测试增加了评估内容,拓宽了考核内容的覆盖面;而且“客观题”测试明确、细化了评分标准,为主观打分趋向客观的量化评分提供了科学依据,测试成绩的统计和比较等方面也都显示出较好的信度,自此,学生学习成果评估步入了近代实证科学的轨道。

### 三、通识教育和研究生教育评估: 1932~1948年

随着美国通识教育的兴起,高等教育评估摆脱了只重视认知知识评估的束缚,开始强调对学生接受通识教育后所形成的“软技能”(Soft-skills)进行评估。在这一历史时期,美国研究生入学测试也开始发展。

#### (一)通识教育评估

20世纪二三十年掀起的美国第一次通识教育“高潮”无疑对美国高等教育的人才培养目标产生了重要的影响,同时也对如何评估学生学习成果提出了新的挑战。因为评估除了关注学科知识以外,还要关注学生个人品性、人际关系及道德伦理观念等通识教育的成果。早期的宾夕法尼亚研究已关注到这一变化,并证明对学生学习进行综合性评估是可行的。20世纪30年代以后,诸多研究项目都对此进行了相关的探索性研究。其中,代表性的案例主要有芝加哥大学方案(University of Chicago)和通识教育合作研究(Cooperative Study in General Education)。

如何对学生在大学通识教育中取得的成果进行评估,芝加哥大学进行了积极的尝试。“对学生学习通识课程的成绩进行评估并非由每一位教师在他们所教授

的课程中进行,而是主要由大学测试中心办公室负责命题、考试和评分”。试题测试更为广泛的知识 and 技能:在多种不熟悉的状况下运用知识的能力;运用原理来解释现象的能力;能够预测结果、确定行动过程、欣赏评价艺术作品的的能力。芝加哥综合测试的特点在于要求学生理解、综合和应用新课文(原始资料),进行开放性的论文写作、完成多项选择题。这一评估方法预示着当代评估的发展。

为了促进院校之间发展通识教育的合作,改善高校的通识教育,1939年1月至1944年9月,在美国教育部的赞助和支持下,拉尔夫·泰勒(Ralph W. Tyler)博士领导了通识教育合作研究。为了使参加研究的被试能够代表美国高校的整体情况,项目组挑选了各种类型的高校参加课题研究,研究的设计帮助这些高校回答“通识课程应包括哪些内容、应该使用什么样的教学方法和教材?应该如何评价和记录学习成绩?”等十大问题。事实证明,该研究获得了巨大成功。虽然每所高校的研究结果有所不同,但参与研究的高校获益颇多,认识到“尽管各校教育使命不尽相同,但某些基本问题具有很大的相似性,因此可以通过合作去解决,即通过分享经验、材料、理念和各自的努力,能够更容易地解决某些复杂的问题”。更为重要的是,通识教育合作研究的很多评估内容关照着“人的发展”这一进步概念,预示着对“批判性思维、社会责任心、美好的心灵”等“软技能”进行评估的要求。

## (二) 研究生教育的发展和 GRE 的早期发展

20 世纪初期美国社会的工业化和城

市化进程加快对各行业高层次人才的需求,促使美国研究生教育进入第一个发展高潮。如何保障研究生教育质量很快被提上议事日程。从入学源头上完善对研究生申请者的甄别无疑是保障研究生教育质量的第一步,而研究生入学考试正是对本科生接受专业教育和通识教育而获得的学习成果进行的一次全面综合的评估。当时,威廉和本·伍德等人提出,学士学位已经“不再是是否适合攻读研究生的入学标准”,而是需要新的标准。1937年10月,为了确定研究生的质量指标,威廉领导的“研究生测试合作项目”课题组和哥伦比亚大学、哈佛大学、普林斯顿大学、耶鲁大学的研究生院一起开展课题研究,执行了7项测试,这便是著名的“研究生入学考试”(GRE)的第一次施行。

最初的 GRE 考试是一个综合的、客观的测试,试题主要关注学生的学科知识,但它也涉及口头推理能力。1939年,16门主要学科的学术测试(Advanced tests)加入 GRE,用来评估专业领域的知识,这些测试专门针对申请读研的高年级学生。到1949年,为了测量“表达的有效性”、提供“通识教育指标”,通识教育知识作为试题的一部分加入能力测试。应该说,“研究生测试合作项目”研究及其后的不断改进是成功的,而且取得了突飞猛进的进展,但迫于财政压力,卡内基基金会最终将 GRE 移交给了教育测试服务中心(ETS),完成了它在学习测试方面的征程,但它在选择题测试等方面的探索为我们留下了宝贵的遗产,且影响至今。

这一阶段的学生学习成果评估的特点主要表现为评估内容发生了显著变化,

即不仅对学生在一定学科领域内的专业知识成果进行测试,而且对通识教育所强调的学生个人成果、社会责任成果、道德品质成果等综合素质进行测试,而与之相应的评估方式也在不断尝试、积极应变,因为与专业知识评估相比,对学生的价值观、责任感等态度、情感的变化进行有效测量是极富挑战性的。

#### 四、测试机构快速发展和大学内部学习评估改革:1948~1978年

二战后,由于军人权利法案(G.I. Bill of Rights)的强有力推动,加之婴儿潮和公民权利法案(Civil Rights)的影响,美国高等教育的发展进入黄金时期。高等教育的快速发展,特别是入学人数的倍数增长和课程变化对传统的学生学习评估提出了新的要求。

##### (一)专业测试机构快速发展

电子计算机的出现和使用使得教育测量的理论和技术在20世纪40年代末得到了第二次飞跃性发展。各种专门从事编制和开发标准化测试的专业组织机构和由它们组织的标准化考试应运而生,而且发展迅速,从而对大学学生学习成果评估的新要求进行了积极的回应。其中,最著名的当数1948年成立的教育测试服务中心(Educational Test Service,简称ETS)和1959年成立的美国大学测试中心(American College Testing,简称ACT)。

ETS的成立标志着美国教育测试大规模标准化、专业化、商业化的开始。1949年秋季ETS引进了GRE的能力倾向测试(Aptitude Test),使测试内容从学科知识转向一般推理,这是学生学习评估的又一个重要里程碑;1954年的“领域测试”(Area

Tests)更加强调关注于学生在社会学科、自然学科和人文学科领域的阅读、理解、诠释等学术能力,而不是评价学生在某一领域的知识,相反,“领域测试”通常只需要少量必要的学科知识。到20世纪70年代后期,部分教师对多项选择测试感到不满,他们想要了解学生是否获得了更为广泛的能力。为了回应这一新的发展变化,ETS究了结构性回答测试(Constructed-response tests),涉及交流技能、分析思维、综合能力和社会/文化意识等。而ACT在对大学生的测试项目中明确声明,坚决反对仅仅使用多项选择题测试学生整体能力的趋势,开始尝试以开放的、自由回答的、学生表现为基础的评估,借此追求测量学生社会体系中、在成人生活中如何使用科学技术、人文知识和实用技能。其研发的“大学成果测量”(COMP)成为“以通过直接观察现实情境中学生的表现为基础的评估”的开路先锋,运用多项选择题、简答、论文写作和口试等多种形式,测试学生三项过程技能:交流交际能力、问题解决能力和价值判断能力。然而,备受关注的COMP评估改革由于测试的时间和评分成本非常昂贵于20世纪90年代后期全部终止。无独有偶,其他许多学习评估的操作模型由于类似原因,从而导致要么逐渐消失,要么以某种变形回到多项选择测试。

##### (二)大学内部的学习评估改革

这一时期,同样由于高等教育结构、入学学生、课程等发生一系列变化,大学自身也在积极应变以求发展。20世纪70年代初的非传统教育和教学专业化等各种高等教育改革运动成为刺激各种评估

技术在大学校园使用的直接诱因,美国大学内部开始了一些重要的评估项目改革,以帮助了解、改善学生的学习情况。

艾维诺学院(Alverno College)是早期改革学生评估措施的创新者之一。其突出贡献在于“综合测量评估中心项目”设定了八项核心能力:有效交流沟通能力;分析能力;解决问题能力;价值判断能力;有效的和社会互动能力;良好的人际关系;国际眼光;审美能力。该校学生在整个四年的学习中通常要接受 100 多个表现评估,必须达到八项核心能力的一项或多项。评估方式采用综合评判的办法,由教师、同伴、社区成员、专业人士和其他一些人员作为评估者,公认的评价标准应用于学院的所有学科。这些成功经验为其他小型文科院校所效仿。东北密苏里州立大学(Northeast Missouri State University)的“增值评估”(value-added)更具深远的影响力。1974年,东北密苏里州立大学开始运用多层面的“增值”方法来确定学生学习收获的程度如何,包括对新生和老生进行标准化考试、对毕业生进行主要学科的考试、对在校生的主观感受进行调查等。增值评估取得了显著的成效,学生标准化考试的成绩有了明显的改善,课程质量也开始提高,这些斐然的成绩使得东北密苏里州立大学招收到了更好的学生。

这一阶段的学生学习成果评估的特点主要表现为评估主体的变化。不难看出,数量众多的专业测试机构通过评估工具的开发,特别是评估量表的研发和完善,积极投身于大学的学生学习评估实践,正式成为学生学习成果评估的主体之一,并发挥了重要的作用。而作为学生学

习成果评估的第一主体——大学自身,在大学内外的各种变化和压力推动下也积极地开展了不同程度的学习评估改革,并开始迎接卓越、公平和效率之间的张力挑战。

## 五、回应外部问责:1979年至今

1979年开始至今,美国社会各界对高等教育的问责此起彼伏。迫于政府和公众的要求,各州高校纷纷投身“评估运动”(Assessment Movement),改革学习评估势在必行。20世纪90年代以后,美国高度重视大学的“产出”,学生学习成果评估在赞誉与挑战中走向专业化。

### (一)评估运动时期学生学习成果评估的展开

美国高等教育自20世纪80年代起渐渐地失去了往日的公信力,社会问责之声四起,“评估运动”一触即发,政府、高校、市场都纷纷迅速投身于教育评估,学生学习成果评估的观念在整个美国高等教育界再起波澜并得到广泛推行。

美国各州政府对学生学习评估的回应各不相同。新泽西州制定了第一个全州范围的新生基本大学技能评估项目,佛罗里达州正式命令高校对新老生都进行知识测试,田纳西州提出了鼓励高校进行二年级和四年级学生成果评估的经济激励措施,还有一些州只是鼓励而不是正式要求高校开展学生评估项目。佛吉尼亚州则比较特别,它虽然正式命令高校进行学生评估,但是允许每个高校开发、选择最符合他们自身多样性特征和使命的评估方式。这一评估政策使得高等教育对大学自治的需求和州政府对资源使用问责的需求之间保持适度平衡,为当时其他还没有



进行学术评估立法的州提供了很好的范例。时至1990年,至少15个州在财政预算中使用学生成果的测量结果,并以此作为对州内高等教育机构进行评估的基础,13个州已经正在通过立法、特派工作组、各种委员会回应学生成果评估的政策发展。

各州高校(特别是公立高校)几乎立刻对学生学习评估进行了回应,大多数学院和大学愈来愈关注学生成果评估,同时也有许多高等教育机构抵制学生学习评估的推行进程:一些高校声称他们卓越的声誉、学生学习质量和教师教学质量在评估过程中被消解了;州政府对学生评估程序的正式命令是对大学自治的威胁;学生评估可能带来应试教育、影响高校个性发展。这些批评对学生学习成果评估有效深入的展开造成了一定的阻碍。尽管如此,学生学习成果评估在各州高校还是不同程度的展开了,并越来越受到重视。伊莱恩(Elaine EL-Khawas)1989年的调查显示,3/4的大学管理人员声称他们所在的高校在决策的过程中正在使用评估信息。此外,六大认证机构也开始要求高校在自我评价的过程中使用学生学习成果评估的结果。

如何高效和有效的收集到作为学生学习成果的各种数据成为高校实施学生成果评估并借此向公众提供教育成果证据的重要工作。这也为专门从事教育测量和评估的社会机构提供了广阔的市场需求,测试机构在此背景下开发了许多新的评估工具,比如美国大学测验中心不断改进“大学成果测量”(COMP)并发展成后来的“大学校际学生学术能力评估”

(Collegiate Assessment of Academic Proficiency,简称CAAP)。很多测试机构还提供了测量结果的常模,便于参与评估的学校和学生进行比较。这些商业开发的工具逐渐得到高校的认同,一些大学开始使用它们测量在校学生的学习过程和学习成果。

## (二)20世纪90年代以后学生学习成果评估的进展

自20世纪90年代后期以来,在全球化导致竞争激烈、青少年价值观不断变化、职业社会对人才培养提出新要求、网络教学对传统教学方式提出新挑战、高校积极回应问责等背景下,美国的高等教育质量模式逐渐由原先传统的院校认证和专业课程认证转移到注重学生学习成果评估上来。各州高校纷纷成立专门的评估机构,配备专业的评估人员,鼓励师生积极参与学习成果评估,提升了学校的教育质量,促进了学生的学习和发展。

进入21世纪以后,学生学习成果评估更是备受美国重视,成为政府、企业界、高校、团体协会关注的焦点。2002年,大学学习全国论坛(The National Forum on Col-lege-Level Learning)强调收集可以进行比较的学生学习成果资料。2006年,美国教育部高等教育未来展望委员会(Commission on the Future of Higher Education,简称CFHE)发表的改善高教报告草案再次强调学生学习成果评估的重要性,并提出诸多具体实施的方案。比如该报告呼吁使用相关测验指标以评估高等教育的附加价值,建议成立一套“全国大学生个人教育记录数据库”(National Student Unit Record)系统,以纪录、追踪全

美大学生的教育成就与发展。

在大学内外都注重学生学习成果评估的背景下,测量学生一般能力、学科成绩、学习行为等各种旨在测量学生学习成果的评估工具的研发和改善在 20 世纪 90 年代后特别是新世纪以来如雨后春笋。其中,教育援助委员会(the Council for Aid to Education,简称 CAE)所研发的“大学校际学生学习评估”(Collegiate Learning Assessment,简称 CLA)充分运用现代计算机技术,增加了广泛的评估内容,而且评估方式也不限于标准化测验,增加了情境式问题分析与论述,其评估结果可以用来评价校际之间的差异,具有较好的信度和效度,赢得了很好的知名度和美誉度,有力地引导了大学评估走向深入。

美国高等教育中的学生学习成果评

估工作不断深入展开的同时,也遭受到了一些批评与质疑,但在持续不断地强调资源与效率、质量与问责的竞争环境中,大多高校已纷纷采用各种不同的评估方式和评估工具,诊断学生能力与学识基础,提供学生学习提高、教师教学改进和其他相关学校措施改善的重要依据,从而确保学生到毕业时能达到相当程度的能力水平。正是诸多赞誉与挑战的交织,推动着学生学习成果评估走向了专业化和科学化,学生学习成果评估的应用领域和正向功能因此也得以不断扩展,改进了高等教育质量管理水平,提高了大学人才培养质量,为美国大学回应外部问责、赢得社会支持发挥了积极的作用,也为世界各国完善大学教育质量保障体系提供了有益的启示。



# 美国高校“学生学习成果评估” 的特点与启示

黄海涛

内涵发展、提升质量已成为 21 世纪世界高等教育改革的鲜明主题。为促进大学教育的卓越,世界各国纷纷探寻更为有效的质量管理范式,作为高等教育强国的美国在此方面更是保持着积极的作为,1980 年代广泛、持久的社会问责和新公共管理思潮推动“直面教育成效”的“学生学习成果评估”(Student Learning Outcomes Assessment)走到了历史的前台,并迅速成为美国高等教育质量保障的重要方式。当前,这种注重“产出”的质量管理方式因其“注重学生学习增值、强调教育成效证据”而备受世界各国关注。美国高校“学生学习成果评估”及其所反映的质量证据文化,如果要有利于我国大学教育质量保障模式的完善则需要我们更好地梳理其历史进路,把握其主要特点,体会其各种争议,合理借鉴其成功经验。

## 一、沿革与走向

### (一)历史沿革

美国高校“学生学习成果评估”虽然只有短暂的历史,但是却有漫长的过去。测量学生通过接受高等教育学到了什么,可以回溯到 1640 年的哈佛校规和 1650 年哈佛学院的第一次毕业课,那时,口试

和高年级学生演讲辩论已经非常普遍。19 世纪末 20 世纪初,随着学生人数的增长和课程门数的增加,对每一个毕业生都进行演讲和辩论的综合考试已变得十分困难。在卡内基教学促进基金会的倡导和努力下,由是非题和选择题组成的客观标准化考试作为一种全新的尝试登上了历史舞台。

20 世纪二三十年掀起的美国第一次通识教育“高潮”修订了美国高等教育的人才培养目标,同时也改变了学生学习成果评估的内容。评估不仅对学生在一定学科领域内的认知成果进行测试,而且对通识教育下学生的个人成果、社会责任成果、道德品质成果进行测试。“二战”后,美国高等教育的发展进入黄金时期,入学人数的倍数增长和课程变化对传统的学生学习评估提出了新的要求。电子计算机的出现和使用使得教育测量的理论和技术在 1940 年代末得到了第二次飞跃性发展,各种专门从事编制和开发标准化测试的专业组织机构和由它们组织的标准化考试应运而生,而且发展迅速,其中最著名的当数 1948 年成立的教育测试服务中心(ETS)和 1959 年成立的美国大学测试

中心(ACT)。在学生学习评估机构蓬勃发展的同时,美国大学内部也开始针对高等教育结构、入学学生、课程等一系列变化进行了一些重要的评估项目改革,以帮助了解、改善学生的学习情况。艾维诺学院“综合测量评估中心项目”设定的八项核心能力和东北密苏里州立大学进行的“增值评估”都具有深远的影响力。

美国高等教育自1980年代起渐渐地失去了往日的公信力,社会问责之声四起,特别是对大学生素质下降的问责直接推动了美国“评估运动”的展开,政府、高校、市场都纷纷迅速投身于教育评估,学生学习成果评估的观念在整个美国高等教育界得以更深度的重新讨论和更广泛的实践推行。进入21世纪以后,学生学习成果评估更是备受美国重视,成为政府、企业界、高校、团体协会关注的焦点。2002年,伊利诺伊等5州首度合作召开大学学习全国论坛,强调收集可以进行比较的学生学习成果资料。2003年,美国高等教育认证评估机构联合会(CHEA)特别发表了《共同促进学生学习成果宣言》,强调学习成果在美国大学中的重要性已日益增长,正逐渐成为评估高等教育效率的主要标准。2006年,美国教育部高等教育未来展望委员会(CFHE)发表的改善高等教育报告草案再次强调学生学习成果评估的重要性,建议成立一套“全国大学生个人教育记录数据库”系统,以纪录、追踪全美大学生的教育成就与发展。在学生学习成果的评估工具的研发和改善方面,教育援助委员会(CAE)研发的“大学生学习评估”(CLA)充分运用现代计算机技术,增加了广泛的评估内容,增添了情境式问题分析与论述的评估方式,其评估结果可以用来

评价校际之间的差异,具有较好的信度和效度,赢得了很好的知名度和美誉度,有力地引导了美国高校学生学习成果评估走向深入。

## (二)发展趋向

### 1.“全人”评估理念将占主导地位

心智、道德、情感及体魄等各方面都能得到健全均衡发展的“全人”教育理念具有时代的进步性,其合理内核同普遍提倡的“学会学习”、“学会生存”以及培养学生的独立性、创造性等理念相一致。“全人”教育的理念贯穿于大学学生学习成果评估的整个过程,是21世纪人才观对高等教育的责任与使命进行重新定位的必然要求。事实上,美国通识教育评估已预示了这种评估理念。当前,基于“大学学习成果就是大学生就读期间的各种经历”这一论断的各种大学生就读经历调查项目非常盛行,这反映“全人”教育理念已广泛应用于评估实践之中。需要强调的是,测量学生的情感或态度变化是“全人”评估理念不可或缺的主要内容之一,但也是学生学习成果评估的难点所在。

### 2. 评估的时间界限向未来延伸

随着学习成果评估越来越重视学生适应社会和应用知识的能力,评估的时间界限也必然向未来延伸。换言之,评估已不仅仅局限于学生在校期间的表现,还要对学生毕业后的工作绩效和实践成果进行更多的了解,并与大学时期的表现进行比较,进而判断二者的关系,为提高高等教育质量提供更加翔实的依据。虽然美国许多大学对校友的调查评估已有多年历史,但是随着网络技术和各种全国性学生数据库的快速发展,大学可以通过更多的方式获得校友的联系信息,数据收集的质

量和效率也随之大大提高。

3. 全国性的统一评估工具被广泛采用并趋于国际化

与标准参照性评估方法相比,利用统一、有效的评估工具进行常规参照性评估则可以更好地进行同类院校之间、相同专业之间的比较。近年来,美国大学利用常规参照性评估工具的趋势越来越强劲。例如,全国大学生学习投入问卷调查(NSSE)适用于各级大学,到目前为止已有1300多所大学使用;“大学生学习评估”(CLA)项目的参与大学已从前几年的60多所骤增到400多所,并成为经济合作与发展组织(OECD)“高等教育学习成果评估”(Assessment of Higher Education Learning Outcomes)项目的理念来源和方法基础。与此同时,全球化推动评判“人”的标准在国家之间的差别越来越小,学习成果评估的标准也在向国际化方向发展,许多学者和大学正在寻求学习成果的常模评价标准。例如,加州大学伯克利分校发展的研究型大学本科生就读经验调查问卷(SERU)正在向欧洲、亚洲、南美洲等国家推广,我国南京大学等高校已正式加盟。

4. 政府部门介入评估事务日益凸显

虽然政府部门,特别是州政府在大学评估中的介入已经遭到了许多学者的批评和诟病,但政府部门进一步介入学生学习成果评估已是不可抗拒的发展趋势。首先,联邦政府会采取各种措施,从政策方面予以支持和号召,并在可能的情况下设立专项资金对全国性的学习成果活动进行调查和研究,为制定政策提供依据。其次,各州州政府也将通过各种方式介入大学学习成果的评估活动中。当然,各州介入的模式和程度会有所不同,但总的发展

趋势是:州政府对公立大学的问责力度加大,学习成果评估指标必须是问责指标的组成部分;部分州政府教育部门设立专门从事评估工作的高等教育评估局;在州政府财务预算力所能及的范围内,加大学习成果评估方面的投入,如肯塔基、佛罗里达州的模式,即州政府通过立法要求大学使用统考评估学生的学习成绩,可能会得到进一步推广。

5. 计算机和网络技术在评估过程中将发挥更大作用

计算机和网络技术在学习成果评估中的使用已经非常广泛,随着新技术的不断出现,这些技术在提高评估效率方面会发挥更大的作用,其主要使用范围将仍然表现在数据收集、分析以及评估结果的广泛传播等方面。可以预见,计算机商业智能的功能将被有效地应用在学习成果评估中,如数据汇报、在线分析处理、商业运行绩效管理、基准分析、分析智能、商业运行管理、预测分析、数据挖掘等。这些功能将帮助评估者提高信息处理和报告生成的自动化与时效性,实现信息收集和数据分析的多维化与动态性,凸显对学生未来发展趋势预测的个性化与准确性。

## 二、特点与争议

(一)大学内部评估为主的多主体评估

美国学生学习成果评估的组织架构以高校内部组织评估为主体,政府部门积极推动、民间测评机构积极参与为辅助。作为人才培养主体的大学是学生学习成果评估最重要也是最直接的组织者和执行者。美国各个大学内部学生学习成果评估部门的设置和名称不尽相同,主管评估的校领导岗位设置也没有一致的模式,有

的学校由教务长或主管教学的副校长负责,有的学校由主管学生工作的副校长负责,也有相当一部分大学设置专门的副校长或者副校长助理职位管理学校的评估工作。他们不仅承担评估的管理或协调工作,而且越来越多地直接参与评估工作,如协助授课教师共同开发考试试卷和调查问卷、与专业管理人员和教师共同开发专业学习成果及评估计划等。大学全体师生的积极参与则是学习成果评估顺利有效开展的关键。

政府和中介机构一直积极参与学生学习成果的评估。美国联邦政府和州政府并没有集中评估权力,而是通过立法、规章制度、拨款等方式影响评估;民间评估机构同样扮演着重要的角色,比如许多认证机构已将学生学习成果评估结果作为认证指标体系的重要组成部分,教育援助委员会、公共政策与高等教育国家中心等社会团体发展大量评估工具直接为高校进行学习成果评估提供服务。与此同时,一些商业性质的评估机构(可能仍然被称做非营利机构)也在通过研发各种评估工具开拓学习成果评估的服务市场。推动学习成果评估的外部力量,特别是政府干预的日益凸显,引起了人们对这一评估方式的担心,即学生学习成果评估是否与大学的学术自由相悖。一些人认为,学习成果评估受到了大学外部的“指手画脚”,从而影响了大学自治与学术自由,与政府教育经费预算挂钩的评估结果将滋生“应试教育”,破坏大学的多元性和个性化,导致大学趋于同化,并抑制高等教育改革。哈特对此深表担忧:“我们(美国大学)所骄傲的多元性以及世界对我们的称赞将会遭

到严重的破坏”。

(二)关注学生能力提高和个人发展的产出性评估

美国高校学生学习成果评估着重考查大学对学生进行课程教学、专业培养、公民教育等一系列“加工”后到底有何“产出”,除了毕业率、学位授予率、职业资格考试成绩、研究生入学考试成绩等校院层面的整体指标外,重点测量大学生在接受某一阶段的高等教育后,知识、技能、能力有何提高,态度、情感、观念有何改变,毕业后的生涯如何发展,对社会作出哪些贡献等多种成果。这些学生个体层面的学习成果是评估活动的主要目标,虽然其指向的具体内容纷繁复杂,且在不同的学校、学科、专业、课程等背景中具有明显的“情境差异”,但在具体的评估中,评估主体对学生学习成果的内容进行“情境选择”、构建可以测量的评估内容体系,从而使学生学习成果评估成为可能。比如,在通识教育学习成果评估中,人文知识、科学知识、社会科学知识等是知识层面评估的主体内容,而写作能力、批判性思维能力、推理能力则是技能评估的主要内容;大学生就读期间的各种经历及其发展变化受到评估者的广泛关注,这些内容都关系着大学教育活动的产出情况,决定着大学教育目的的实现程度。

与重视教育产出的学生学习成果评估相并行的还有注重条件投入和教学过程的评估。注重教学条件的投入评估可以从教育“开端”上就对质量进行核查,确保大学教育运行具有“基准”的条件保障;教学过程评估可以加强对教学过程中出现的质量问题进行跟踪监控,但对于大学人

才培养质量到底如何,这两种评估方式往往是“说也说不清楚”。直接反映教育成效的产出性评估同样受到了一些学者的质疑,亚历山大·奥斯汀认为,理解教育的结果必须了解成果、输入、环境这三者之间的关系。他构建的“投入—环境—产出”评估模型说明:知道学生达到什么结果只是评估的一部分,而关于学生起点的信息和他们在求学过程中所经历的事件同样也是必不可少的。没有这些背景,要使用成果评估的结果来提高教育质量是十分困难的。事实上,高等教育评估是一个体系,以上三种评估并非彼此独立、相互排斥。相反,每一种评估模式都会不同程度地涉及其他两种评估类型的评估内容,比如,教学条件的投入评估也会考查教学投入的成效如何,学习成果评估也同样关注学习的基本条件和学生的学习行为。

### (三)注重测量数据的综合性评估

由于大学生的学习是多维、综合、长效性的行为,知识和技能等学生学习成果具有各种“情境差异”,因此,美国学生学习成果评估采用多元评估方法和多种评估工具,以便能全面系统的反映学生学习成果的全貌。定性评估通常采用学生表演、发展档案袋、顶点课程、行为观察以及调查访谈等评估工具;课程考试、标准化测试、量规、毕业后的继续学习和职业发展跟踪数据等则是常用的定量评估工具。运用这些评估工具持续收集到的各种学生学习成果信息是评估运行的基本“原料”。评估主体将这些“原料”转换成初始数据,并运用数据分析技术和理论假设对初始数据进行加工、分析、提炼和挖掘,从而对教育目标实现程度、教学中的现存问

题及其解决路径进行诊断和评价,并反馈给相关部门和人员,使得评估结果得以发挥其重要的价值和作用。

从技术层面讲,评估和测量、评价密不可分。学生学习成果评估依赖于各种有效的教育测量,是一个不断收集数据、分析数据、应用数据的过程。因此,美国综合性的学生学习成果评估非常注重各种具有统计价值的测量数据,并最大限度地利用这些数据。面对大量的学习成果评估数据,一些教育者抱怨他们早已淹没在数据之中,但是很多数据并没有为教学提供有效帮助。一些学者提出学习成果评估不能陷入玩弄数据的技术主义,不能沉迷于数字的慰藉而不能自拔,毕竟数据不是评估的最终目的。对此,马尔凯塞一再强调:“评估不只是收集数据。为了进行评估工作,教育者必须明确他们收集信息的目的。数据收集后,教育者必须研究并使用这些结果,从而能够改善教育活动”。事实证明,打开科学的数据之门,得到可信的分析结果以驱动教学活动的正确决策确是学生学习成果评估面临的一项巨大挑战。

### (四)强调不断改善教学的持续性评估

美国学习成果评估不是对学生掌握知识程度的一次性检测,也不是简单地将一系列调查堆砌而成,而是以理解并改进学生学习为目标,由精心设计的、符合大学教育目的和教学理论的一系列持续不断的活动组成。需要指出的是,强调评估过程和评估的持续性是这一活动的基本要求,即学生学习成果评估的运行机制表现为一个持续不断的循环过程,在整个评

估活动中,评估者付出持之以恒的努力,不断收集和分析各种学生学习成果数据,解释测量结果,并应用于课程的重新整合和专业教育的调整,在下一轮评估中再次测量其变化,以达到不断改善教育教学、帮助学生发展之目的。作为一个操作性概念,美国学生学习成果的内涵本身就具有动态性,它既包括学生学习行为发生之前学习成果的预设,也包括学习行为发生期间学习成果的实施以及学习行为完成之后成果的检验,这些步骤又都和学习成果评估的运行密切相连。

持续性的学习成果评估与“一考定终身”的总结性评估相比,可以帮助授课老师和学生在教学过程中发现教学方面存在的问题,为双方提供改进的机会,从而克服评估结果不能为当时受评的学生服务这一缺陷。正如帕特里克·特伦津尼所言:“学习成果评估关注过程而不是成果。评估成果暗含着一种最后的结果;评估过程则意指有进行改善的时间和机会”。美国高等教育学会(AAHE)组织有关教育专家研究总结的“大学学习评估操作九项原则”也强调“只有把评估看做是持续的,而非间歇性的工作时,评估的实施才能最顺畅,结果最有效”。然而也有教师抱怨周期性的学习成果评估增加了大量的额外工作量,而效果并不明显。帕特·哈群斯的调查显示,教师未能积极参与学生学习成果评估的主要原因之一便是“他们也许还没有看到评估的效益。许多大学将评估看成是‘多余’的活动,是对现有许多活动的重复,没有产生任何额外的效益”。

#### (五)凸显证据文化的绩效性评估

美国学生学习成果评估主要是一种评价目标达成度的绩效性评估,即评估主

体通过适当的大学生学习成效测量数据,判断预设的学生学习成果是否实现或实现的程度,借此对高等教育机构和课程的有效性进行评判。作为绩效考核指标所指向的预设的学生学习成果是多方利益相关者共同协商、重叠共识的结果,且具体明晰,易于观察,便于测量。在崇尚实证科学主义的美国,有关学生学习成果的测量数据让社会公众、政府官员、学生家长、用人单位等了解学生在经历高等教育后的“增值”情况,从而判断预设目标达成情况,这些评估结果都将成为高校办学绩效的有力“证据”,“暗示着‘决策和支持’的背景,决定利益相关者进行持续投资的必要性,提供教育教学过程中哪些方面需要改善的信息”。这种强调教育成效证据的评估范式充分凸显了美国高等教育质量保障体系所秉承的证据文化。需要指出的是,作为绩效指标的学生学习成果评估结果与政府的拨款预算相关联,但不与教师的奖惩直接挂钩。

事实上,美国20世纪80年代的各种问责已经形成不可忽视的强大力量,检视着高等教育的一言一行,影响着高校的声誉和资源。为了回应问责,高校必须积极采取各种措施和方法来保障质量,并向政府、公众展现他们所做出的努力和取得的成效。在此背景下,学生学习成果评估作为一种合适的绩效管理方式无疑得到了广泛认同,因为有关学生学习成果的评估数据无疑可以直接展示大学教育取得的产出成效,提供明了的质量证据。这种证据文化促进了大学发展自我反省、自我检讨的能力,为大学完善教学、改进服务、修订评估提供了管理基础。同时,正如其他管理领域的绩效评估一样,急功近利、评



估目的与手段倒置等问题同样成为人们对大学采用这一评估方式进行批评的主要原因。

### 三、启示与借鉴

美国学生学习成果评估对于促进高等教育质量保障理念和方法的发展有着积极的贡献。它反映了现代大学在“质量危机”时代开始审视自身的弊病,反思自身存在的合法性基础,重新回归以培养高质量的学生为大学卓越标志的正确道路上来。当然,学生学习成果评估也遭受到大学自治和学术自由的理念挑战,面临着教师参与积极性不够、经费投入不足等现实困境。美国学生学习成果评估为我国高等教育领域优化人才培养质量保障体系提供了有益的启示。

(一)推进教育教学理念和评估理念的转向

借鉴美国学生学习成果评估范式迫切需要我们改变传统观念。首先要强化“以学生为本”的教育理念。尊重学生的主体地位、促进学生的个体发展、相信每个学生都能成功既是美国学生学习成果评估的出发点,也是其最终目标所在。随着教育改革的进一步深化,我国高校大力倡导“以学生为本”的教育理念,但我国高等教育大众化进程速度过快,客观上造成高校对学生的关注不够,“科研至上”导致的“教学漂移”进一步忽视了人才培养的质量。因此,对学生学习情况进行持续性的评估要想落到实处、取得成效,必须要将“以学生为本”的教育理念体现在教育者的行动之中。

注重学生学习成果评估需要重新思考我们的教学理念,我们必须挑战我们原

先的有关教师应该如何教学、学生应该如何学习,以及他们在教学中的既定角色等基本假设,重新思考在学习过程中教师的角色和学生的角色。我们要从“教为中心”转向“学为中心”,更加关注学生的学习,教师的角色从主要的信息给予者和评价者转向帮助学生学习的引导者,成为和学生一起评价学习的伙伴。

评估理念从注重教学条件的“投入”性评估转向注重教学成效的“产出”性评估。我国传统的各种评估的基本假设是“投入决定质量”。因此,无论是合格评估、优秀评估、随机评估,还是将上述三种评估方案整合形成的普通高等学校本科教学工作水平评估,其评估指标主要注重条件性的、保障性的教学投入。这种评估模式对重视教学投入、规范教学管理起到了一定的促进作用,但是学生培养质量到底如何往往被忽略。因而,借鉴美国学生学习成果评估,必须在评估理念上转向学习产出性评估,评估学生实际上发生了什么变化,从这些变化中间再来讨论学校提供的条件。需要说明的是,评估理念转向产出性评估并不是否认对教学投入、过程以及教育环境的评估,而是要认识到对教育产出——学习成果的评估非常重要,因为学习成果评估的结果可以成为教育质量最直接的证据,更具有说服力。

(二)改进学生学习成果评估运行机制

#### 1. 预设明确的学生学习成果

预设明确的学生学习成果是进行有效的学习成果评估的起点,是评估具有明确指向的前提。认真思考并明确预设学生学习成果是我国改善学生学习成果评估

体系的首要任务。高校要在教师、管理人员、校友、业主、学生家长和学生自身多方交流、协商的前提下预设学生学习成果；学生学习成果预设的陈述要明确具体；学生学习成果的预设还应明确相应的评估方式。

## 2. 使用多元的评估方法与评估工具

我国目前高校内部对学生学习成果的评估主要对课程、专业学习的评估和对综合素质的评估。评估方法多为直接评估、定量评估、正式评估、总结性评估、终端性评估等,传统的考试方法几乎“一统天下”。高校外部对学生学习成果的评估方式则更为松散,一些社会机构进行的各种资格证书考试由于不同的价值取向而良莠不齐。对照美国常用的评估方式,我国目前的学生学习成果评估在实践中应补充间接评估、定性评估、非正式评估、学生批判性评估、校外第三方组织的评估等评估方式,构建多元化的评估方法体系。在评估工具的选择和使用方面,应采用其他多样评估工具。比如量规、问卷调查、项目研究、面试、为每个学生建立学习成果评估档案袋等。同时,鼓励教师和社会评估机构研究开发新的评估工具。

## 3. 加强过程评估和评估的持续性

我国高等教育大众化后学生数量的激增客观上为教师对学生学习成果进行细致有效的过程评估带来了困难。为了不断收集和分析各种学生学习成果数据,学习成果评估应做以下改进。(1)增加持续性评估。采用不同的评估形式对教学进行周期性的评估,为师生双方、特别是处于学习进程中的学生提供改进的机会。同时,和终端性评估合理结合。(2)强化形成

性评估。加强对学生知识和技能构建进展过程进行评估,对学生平时的作业、专题论文、实验报告进行及时批阅并附上评语返还给学生。(3)增加评估人员。聘请研究生担任助教,在大班教学的基础上,组织小班讨论,为学生答疑,并反馈给任课教师,以改进教师之后的教学。

## 4. 建立评估数据库,公开评估信息

我国构建学生学习成果评估体系迫切需要建立基于一定理论框架的、功能强大的数据库,整合与学生学习相关的各类数据。高校可以借鉴罗索·艾科夫提出的“数据—信息—知识—智慧”模型理念,在“决策支持系统”中构建“智能型”学生学习成果评估数据库。高校之间根据自愿的原则搭建联盟数据库,评估院可以协调高校和其他组织通过网络对接形成更大范围的数据库。这些数据库的信息应按照使用对象的权限向需要的部门和人员开放,为有关管理部门以及大学有效开展高等教育质量评估以及高等教育管理提供数据信息服务,同时也为学生、家长、用人单位和社会各界提供了解高校学生学习成果的“窗口”。

(三)建立专门的组织机构,配备专业的评估人员

现代大学已超越传统的“自治”而进入一个质量“共治”的新时代。学生学习成果评估在大学内部和外部影响的张力场域中应处理好政府和大学的关系,坚持内部评估为主,坚守大学自治底线,发挥外部力量的正向功能,增加评估的活力和动力,共同推动评估发展。这一目标的达到需要在政府部门、高校内部和社会中介之间组建“三位一体”的评估组织机构,配备

专业化的评估人员。

政府部门无需单独成立学生学习成果评估机构,可以以现有的评估院为组织载体,拓展其学生学习成果评估功能,但要慎重处理好外部评估和内部评估之间的关系。高校教学质量的提高在根本上还是取决于高校自身的努力,人才培养质量的责任主要在高校,因此,高校始终应是学生学习成果评估的主体。由于我国目前尚未形成专业化的院校研究机构,因此建议在学校管理职能部门中单独设立大学评估和质量保障办公室,专门负责全校的教育质量评估工作。评估和质量保障办公室和现有的教务处、学生处、规划办、政策研究室等职能部门之间如何协调,则需要全新的制度设计进行保障。这必定是该项工作起步的难点之一。同时,支持建立专业化、组织化程度都很高的民间第三方非营利性评估组织,要鼓励其开发科学有效的学生学习成果评估工具。对于民间第三方评估组织的资质认定,可以借鉴美国的认证制度,由教育主管部门通过认证手段,确保取得资质的民间第三方评估组织达到独立、公正、专业、权威的基本标准。

能否培养专业化、高素质的评估人员关系到学生学习成果评估能否有效开展。我们可以借鉴美国的经验,在研究生培养目标体系中,在教育学等专业下增设“教育评估与评价”方向,培养专业硕士和博士,从而实现在近期内配备专业评估人员的目标。在评估活动中,教师的评估能力和积极作为至关重要,这意味着大学教师既是学科专家,又是具有一定专业水准的

评估实施者。因此,高校对教师进行经常性的评估培训是非常必要的。

#### (四)形成学生学习成果评估动力机制

激发并维持评估主体实施学生学习成果的行为动机,是学生学习成果评估运行机制成为现实可能的必要条件。在美国教育领域,学生学习成果评估的外在推动力主要来自政府和公众对大学人才培养质量问责的社会环境。总体而言,我国高等教育的办学主体以政府为主,办学经费主要来自政府拨款和学生学费,因此,回应政府和公众对大学资源使用绩效如何的问责理应是大学不可推卸的责任。

教师把教书育人作为第一要务是保证教师进行积极的学生学习成果评估的假设前提之一。而这些动机的产生除了问责等外部环境的推动外,更为主要的动力应来源于教师需求学生学习成果评估这一内在驱动力。这种内在驱动力首先来自教师的敬业精神。注重教学、关注学生学习和成长应是教师的本职,也是教师这一职业的本质使然。反思我们当前的教师评价标准,培育教师的敬业精神,增强教师的教学责任心刻不容缓。否则,关心学生学习和成长,对学生学习成果进行持续评估很难成为教师自我价值实现的内在需求。可通过适度的奖惩激发教师参与学生学习成果评估的内在驱动力。调整教学和科研的奖惩力度,出台鼓励教师投入学习评估的激励措施或许能够改善现状,但我们必须警醒绩效评估的负面效应。

# 美国高校学生学习成果评估的组织架构及其职能

黄海涛 常桐善

提升质量是 21 世纪高等教育发展的鲜明主题。为了监控与保证高等教育质量,评估已成为主要的手段。在世界各国的高等教育实践中,正在开展着各种各样的评估活动,其中直接指向学生的学习状况、直接反映高校教学质量本质的“学生学习成果评估”(Student Learning Outcomes Assessment)已成为美国高等教育质量保障的重要范式,为美国高校回应外部问责、赢得社会支持发挥了积极作用。美国高校学生学习成果评估可以分为内部自评和校外评估,因此,评估的组织架构也可相应分为高校内部评估机构和高校外部评估机构,高校外部评估机构又可以分为政府部门和民间组织对学生学习成果进行评估的机构。高校、政府、民间组织发挥着各自的职能并相互配合,形成了三位一体的组织架构,为学生学习成果评估的运行提供了有效的组织保证。

## 一、高校内部的学生学习成果

评估机构及其职责高校作为教育产出的“工厂”,有责任对“产品”的质量进行检验,所以高校在学生学学习成果评估中担负着最重要的职责,其内部相应机构也是

高校评估重要的组织部门。美国高校学生学习成果评估的专门机构兴起于院校研究的形成和发展。20 世纪 70 年代中期到 80 年代中期,院校研究开始对学生学习成果评估进行探讨和实践,院校研究办公室同时兼任院校研究和学生评估的双重职责。时至今日,这种模式在美国许多高校依然存在,并成为高校内部评估组织的主要模式之一。

### 1. 高校内部评估组织架构

目前,无论规模大小、公立还是私立,美国几乎所有的高校都有相应的机构管理学校层面的学生学习成果评估工作。各个高校内部评估部门的名称千变万化,有的称之为评估办公室,如佐治亚理工学院(Georgia Institute of Technology)就专门设有评估办公室;有的则兼顾院校研究、规划和评估等工作,称之为院校研究规划评估办公室,如马里兰东海岸大学(University of Maryland Eastern Shore)的评估办公室;也有的称之为评估和质量保障办公室,如在加州州立大学斯坦尼斯劳斯分校(California State University Stanislaus);还有的则在院校研究办公室或者规划办公室下设一个分支机构专门协调和管理学

校层面的评估工作。

主管评估或者院校研究办公室的学校领导岗位设置也没有一致的模式,有的学校是由教务长或主管教学的副校长负责,有的学校是由主管学生事务的副校长负责,也有相当一部分高校设置专门的副校长或者助理副校长职位管理学校的评估工作。是否设有专门的校级领导职位负责评估工作主要受学校规模的影响,同时也与学校的办学性质有关。以教学为主的高校以及社区学院更倾向于设立专门的校级领导岗位负责评估工作。例如,查尔斯顿学院(College of Charleston)就设有专门管理规划、院校研究和评估的副院长职位。有的高校还在院(系)设立学生学习成果评估协理员职位,以协调学校和学院层面的评估工作。

另外,为了加强学生学习成果评估的有效性,许多高校还成立了校级评估委员会。例如,锡拉丘兹大学(Syracuse University)设立了“全体大学学生学习成果评估委员会”(All University Student Learning Outcomes Assessment Committee)。校级评估委员会通常是一个独立的临时机构,委员会主席由教师担任,成员主要包括教师和管理人员。有的校级评估委员会也吸纳学生、校友、企业管理人员、社区人士参加。评估委员会参与制定学校的有关评估政策,进行调研,特别是开发评估项目和促进教师参与评估活动。评估办公室通常为评估委员会的活动提供数据支持和相关服务。

## 2. 评估办公室的主要职能

美国高校内部的学生学习评估工作主要由教师完成。校级评估机构——评估办公室主要是管理、服务、协调的机构,

通常履行以下职能:制定校级评估目标、计划、实施步骤以及对计划实施情况进行追踪评价,及时发现问题,为学校领导和评估委员会提出改革建议;管理和实施校级学生学习成果评估工作;负责评估数据的收集、数据库的维护;承担学校内部评估培训工作,培训教师和学院的评估协调人员;参与和准备学校的战略规划、专业评估、问责报告等;支持院系评估委员会的工作,为校评估委员会服务,提供相关数据并组织安排有关会议;支持、协调外部评估机构的工作。当然,由于学校规模和对学生学习成果评估的重视程度不同,各个高校的评估办公室的职责也不尽相同,在每个方面的重视程度也存在很大的差异。

## 3. 评估岗位设置及其人员技能要求

根据学校规模以及评估的具体任务,美国各高校校级评估机构的岗位设置以及管理和研究人员的任命级别也存在显著差异。如果评估办公室独立于院校研究处室,则设置主任职位一名、副主任职位一名、协理员和分析研究员若干名。有的规模较小的高校通常不设副主任职位。如果学习评估是院校研究或者规划等处室的下属机构,则设置副主任或者协理员职位分管评估工作,另外,配备分析研究员若干名。许多规模较大的学校,主任职位级别较高,有的高校则由助理副校长兼任。协理员和分析研究员则根据工作经验和教育背景分为不同的级别。一般来说,无论是主任还是其他人员,都要求了解高等教育政策、掌握学生学习成果评估的专业知识、具有很强的统计知识和数据分析技能、熟悉数据库的构建和应用,具有独立工作、团队协作能力以及协调和交际能力。

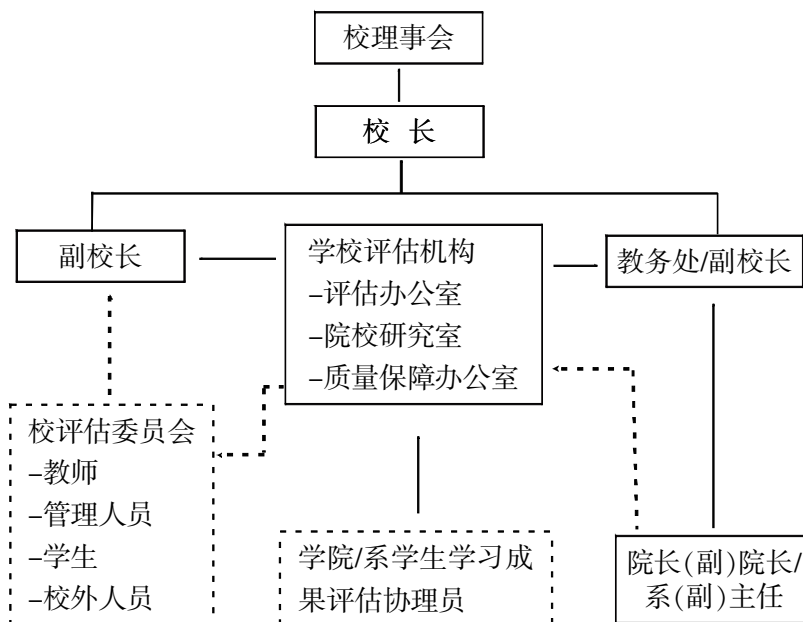


图1 美国高校内部学生学习成果评估机构示意图

## 二、政府部门的学生学习成果评估机构及其职责

### 1. 联邦政府机构在学生学习成果评估中发挥的作用

美国联邦政府没有类似于中国教育部评估中心这样的专门机构,也不以任何形式直接参与高校学生学习成果的评估工作,而是通过对认证机构的资质认定、学生资助评价标准的制定、政策研究与制定、数据收集与分析等对学生学习成果评估发挥重要影响。

第一,由于从事学校层面和专业层面的认证部门的合法身份必须得到联邦教育部的认可,因此联邦教育部实际上通过认证机构间接地影响高校的教学质量和各项评估。1989年,联邦教育部便要求地方认证机构必须将学生学习成效纳入再次认证中的必要一环。近些年来,由于联

邦教育部的明确要求,许多认证部门已将学生学习成果评估作为认证指标体系的重要组成部分。毋庸置疑,学生学习成果评估工作的好坏已成为高校能否通过认证最重要的因素之一。斯坦·伊肯柏里等人的调查表明,认证是目前推动高校内部进行学生学习成果评估工作最主要的驱动力。

第二,联邦政府资助学生的评价标准对学生学习成果评估起到一定督促作用。美国在校本科生中大约有一半学生从联邦政府接受不同形式的资助。联邦政府评价申请资助学生的合格性标准虽然主要是家庭收入,但同时也考虑学业成绩等方面的因素。例如申请资助的学生其注册修课的目的必须是获取学位或者某种证书,入学后必须保持良好的学业成绩等。如果享受联邦政府资助的学生达不到学业成

绩要求或有不良恶习(如吸食毒品),联邦政府将停止对其资助。显然,联邦政府用于评价学生资助的有关标准对高校的学生学习成果评估起到了间接的督促作用。

第三,联邦政府通过对全国性的学生学习质量及其评估进行考察和研究,对其存在的问题进行梳理和剖析,并提出改进建议和措施。比如联邦教育部2005年成立的高等教育未来发展委员会(CF-HE)经过近两年的听证、调查研究,完成了一系列报告,在谈及高等教育质量时明确表示,高校缺乏可行的评估学生学习成果的方法,更缺失比较不同学校学生就读经验的有效措施,因此强烈建议州政府教育官员重视和支持学生学习成果评估;建议构建完整的全国高校评估网络,包含一系列测量学习成果的各项标准;要求高校评估过程必须透明、公开等。这些建议引起了全国高教界的高度关注。

第四,为全国性的评估以及高校之间的比较收集和提供数据是联邦政府在学生学习成果评估中所发挥的另一重要作用。联邦政府利用其特有的权力,并依据相关法律收集全国性的高等教育数据,免费提供给管理和研究人员,以开展相关的政策分析和项目评价,其中许多研究成果与学生学习成果密切相关,为改进学生学习成果评估提供了有效的依据。

## 2.州政府和立法机构在学生学习成果评估中发挥的作用

除了部分州(如佛罗里达州)在州政府教育部门设立了专门从事评估工作的高等教育评估局以外,美国大多数州政府教育主管部门并没有专门的机构从事学生学习成果评估工作。但在管理实践中,州政府和州立法机构对高校特别是公立

高校的质量监控影响深远。20世纪80年代中期广泛的社会问责推动州政府开始通过立法、规章制度、拨款等方式对大学生学习成果进行直接或者间接的评估。

州政府通过立法要求高校进行学生学习成果评估发挥着最为刚性的作用。如加利福尼亚、佛罗里达和肯塔基等州政府通过立法要求高校使用统考评估学生的学习成绩。肯塔基州1997年通过《高等教育提高法案》,并明确规定,州高等教育理事会将开发和实施在以下几个方面测量高等院校的问责体系:(1)教育质量和教育成果;(2)学生入学后的进展;(3)研究和服务活动情况;(4)资源的利用情况;(5)其他支持完成战略规划的性能或者成果,如参与提高基础教育质量的成就;(6)理事会认为有用的其他指标。自此,肯塔基州的所有高校都通过统一考试(从2008-2009学年开始使用美国大学考试中心开发的大学校际学生学术能力评估——CAAP)、全国学生学习投入调查问卷(NSSE)等多种形式对学生的学习成果进行评估。

州政府的各种评估政策直接影响着教育经费的划拨。比如州预算委员会是划拨州税收的决定性机构,其制定的政策就是高校评估的重要影响因子。到1990年,大多数州都制定了本科教育评估政策,要求高校对学生的学习成果进行评估。虽然学习目标和评估方法由高校自行设定,但学校必须向州政府汇报评估结果。到了20世纪90年代中期,州政府层面的学习成果评估由于经费紧张而受到一定影响,评估力度有所削弱。近几年来,受联邦教育部有关高等教育未来发展报告的影响,学生学习成果评估再次在州政府掀起波澜,

成为州政府在教育发展方面重点关心的政策领域之一。

### 3. 地方政府和社区对学生学习成果评估的影响

就地理位置而言,地方政府和社区距离高校最近,与之关系也最为密切。然而,地方政府和社区一般不直接参与高校管理,也很少给高校提供经费和学生资助,但对公立高校而言,社区住户的部分税收会用来支付学校开支,地方政府也为学校提供许多公共服务设施。因此,尽管地方政府和社区对大学生学习成果评估没有直接的“问责”权,但其密不可分的“伙伴关系”影响着高校的发展,包括大学生学习成果评估。当然,更确切地说,地方政府和社区对学生学习成果的评估是一种“声誉”和“服务质量”的评估。

美国绝大多数高校坐落在中小城市或市郊,而且多数情况下校际距离相对较远,社区居民以及地方政府一般很容易识别大学生的学校身份,学生的一言一行都代表着各自学校的形象。地方政府和社区公众的反馈信息便成为对各个高校教育成果的评估信息之一。社区则是大学生从“象牙塔”通向劳动力市场的一座桥梁,是一个提高他们智慧和实践经验的“天然”实验室。社区通过为学生提供义务或有偿的服务机会,促使学生成为有敬业精神的良好市民,积极参与社会活动。而社区人士和雇主对学生参加服务和工作的鉴定和评价又为学校更加全面地评估学生学习成果提供了非常重要的信息。

### 三、民间组织的学生学习成果评估机构及其作用

美国许多民间组织在高等教育质量

保障体系中发挥着重要的作用。具体到学生学习成果评估这一质量保障方式的运行,同样如此。美国涉及学生学习成果评估活动的民间组织数量众多,以下选取最有影响力和代表性的机构予以简要述评。

#### 1. 直接参与评估工作的民间组织及其作用

##### (1) 认证机构。

认证制度是美国最为重要,也最具特色的高等教育质量保障制度。美国高校认证机构繁多,而且其数量几乎每年都以8%-10%的速度递增。如前文所述,学校层面的认证和专业层面的认证目前已成为推动高校进行学生学习成果评估工作最主要的驱动力。认证机构在学生学习成果评估中的作用主要表现为认证指标体系的调整。目前,许多认证机构已将学生学习成果评估作为认证指标体系的重要组成部分。认证机构对高校进行学生学习成果评估的要求主要包括:学校的办学目标,希望学生经过高等教育以及学科训练所达到的标准的明确表述;有关学生学习评估活动的计划;用于改善教学和学生学习的支撑材料;将有关学生学习评估的材料作为院校评估支撑材料的部分证明等。20世纪90年代后期开始,美国各种专业认证委员会也要求考核类似的学生学习成果评估内容。例如,20世纪90年代末,美国工程认证委员会制定了新的“工科标准2000”(Engineering Criteria 2000),该标准不再侧重于对课程形态的认定,而是关注学生学习成果的认定,即从教师“教什么”过渡到学生到底“学习了什么”。此外,新标准还明确提出了所有的工科学生必须达到的11种基础能力。

##### (2) 公立大学与政府赠地大学协会。



组建于 1887 年的公立大学与政府赠地大学协会 (Association of Public and Land-Grant Universities, APLU), 是由美国州立大学和政府赠地大学自愿组成的非营利性公立大学协会。该协会的主要职责是提高高等教育质量, 帮助高校提升学生学习成绩以及参与程度。始建于 2007 年的“自愿问责体系”(Voluntary System of Accountability, VSA) 是该协会的主要成果之一。到目前为止, 近 400 所高校自愿参与了该项目。“自愿问责体系”的主要目的之一便是支持高校进行教育成果评估, 并帮助高校探索和实践提高高等教育质量的经验。该体系包括很多学生学习成果评估的内容, 诸如高校评估学生学习成果的方式(如州政府要求的统一考试和专业评估数据等), 学生能力增值评估的结果(如批判性思维能力、分析能力、写作交流能力)等。此外, “自愿问责体系”还组织“夏季学生学习成果评估研讨班”, 与会者相互交流并共同分享评估学生学习成果的有效经验和新的策略, 以及如何利用评价结果改善高校教学质量。

### (3) 公共政策与高等教育国家中心。

公共政策与高等教育国家中心(The National Center for Public Policy and Higher Education, NCPPHE) 成立于 1998 年, 是一个独立的、非营利、非政府、非党派组织。其主要目的是通过实证性的政策研究和信息分析为提高公共政策的制定效率提供依据, 同时为社会公众了解高等教育发展提供信息资源。“美国国家高等教育报告卡”(The National Report Card on Higher Education) 是公共政策与高等教育国家中心最有影响的成果。从 2000 年起发行的报告卡包括美国 50 个州的高等

教育信息, 而学生学业完成情况、高校教育成果、大学毕业生的能力等学生学习成果则是报告卡的重要内容。

## 2. 开发和提供学生学习成果评估工具的民间组织及其作用

自 20 世纪 40 年代中期以来, 许多民间测试机构通过开发和提供评估工具在学生学成果评估中扮演着重要的角色。具有领跑性质的美国教育考试服务中心(Educational Testing Service, ETS) 和美国大学考试中心(American College Testing, ACT) 无疑是其中最具影响力的机构。教育考试服务中心开发的有关大学生学成果评估的测量工具主要有: 通识教育技能测量(Measure General Education Skills, MGES), 大学专业知识测量(Measure Subject Knowledge, MSK), 课程效果调查(Survey Students on Course Effectiveness, SSCE), 信息和交流技术技能测量(Measure Information and Communication Technology Literacy Skills, MICTLS), 写作能力考核工具(Provide Detailed Diagnostic Writing Feedback, PDDWF) 等。美国大学考试中心的一些项目同样致力于提高学生学习成果, 特别是通识教育学习成果。如 CAAP 就是美国大学考试中心开发的标准化考试评估项目, 其结果通常用于以下几个方面: (1) 认证和问责报告; (2) 测量学校层面、专业层面及个人层面的学习成果; (3) 对不同高校、不同高校相同专业的学生学习成果进行比较; (4) 评价学生通识教育知识和技能等。

在开发学生学习成果评估工具方面, 另一个非常有影响的组织机构是成立于 1952 年的教育援助委员会(Council for Aid to Education, CAE)。该委员会是一个

非营利机构,是美国唯一的拥有教育援助数据的机构。目前,在美国最有影响的学生学习成果评估工具“大学校际学生学习评估”(College Learning Assessment, CLA),就是教育援助委员会和独立学院理事会(Council of Independent Colleges, CIC)联合开发的。“大学校际学生学习评估”的主要目的,就是通过直接测量学生在分析理解、批判思维和写作等方面的学习成果增值来评估高校教育质量。教育援助委员会从2005年起,每年夏季都要召集成员探讨和研究评估项目的有关问题。此外,教育援助委员会还单独为社区学院开发了“社区学院学生学习成果增值评估工具”(Community College Learning Assessment: Value-Added Assessment)。

### 3. 提供学生学习成果评估研究和交流平台的民间组织及其作用

美国大多数民间评估组织提供学生学习成果评估研究、交流的平台,以促进和提升高校学生学习成果评估工作的专业化水平。如学习成果评估国家研究所(National Institute for Learning Outcomes Assessment, NILOA),院校研究协会(Association for Institutional Research, AIR),美国高等教育学生学习评估协会(The Association for Assessment of Learning in Higher Education, AALHE),美国评价协会(American Evaluation Association, AEA)等。以上所列机构只是参与和研究学生学习成果评估活动的主要民间机构,还有许多类似机构虽然影响力较小,但也在快速发展。以下简要介绍前面两个民间组织。

学习成果评估国家研究所成立于2008年,旨在帮助高校进行学生学习成果评估活动。它的基本任务是开发并有效地

向社会各界传授利用评估信息提高本科教育质量的方法。学习成果评估国家研究所希望和政策制定者、学生家长以及对学生学习成果评估感兴趣的各方人士进行交流,近两年的研究成果非常丰硕。

院校研究协会成立于1966年,正式运行于1974年,主要通过召开年会、出版刊物、举办培训班来支持高校提高数据质量,促进高等教育的交流和研究。其年会上交流的论文中有相当一部分与学生学习成果评估有关。它还专门为从事评估工作的管理和研究人员举办旨在提高评估技能和应用能力的培训班。近几年出版的与学生学习成果评估有关的著作主要集中在专业领域的学习成果评估。

### 四、美国高校学生学习成果评估组织架构的主要特点与启示

总体而言,美国高校学生学习成果评估的组织架构形成了以高校内部组织评估为主体,政府部门积极推动、民间组织积极参与为辅助的三位一体的组织模式。提升教育质量、培养优秀人才是高校作为教育机构的永恒使命和不懈追求,而评估学生接受课程教学、专业训练等高等教育后有何收获则是学校教育环节不可或缺的重要组成部分,因此高校内部的教学评估机构必然是学生学习成果评估组织的主体,担负着学生学习成果评估的协调组织和具体实施。政府以教育经费为杠杆,通过制定、实施相关政策引导、影响高校实施学生学习成果评估。需要指出的是,虽然美国联邦政府没有设立专门的评估机构,但加强高校评估,提高学生学习成效一直是包括总统在内的政府官员们所关心的重要问题之一。民间组织积极参与

为学生学习成果评估提供了多方位的服务,与此同时也为自身的存在和发展寻求更多的合理性基础。

美国分权式的高等教育体制对某高校评估的组织架构起着关键性的影响和作用,为学生学习成果评估三位一体的组织模式的构建提供了可能,这也充分体现了美国高校办学的自主性,说明了内动力对高校发展的重要性。在这种三位一体的组织模式运行机制中,各组织之间遵守契约关系,即政府制定的评估政策是政府和高校、民间团体之间必须履行的契约;高校因购买民间测评机构的评估工具和服务而与之签订的合作合同则是它们之间必须履行的契约。

然而,基于不同的组织特性,作为学术组织的高校、作为行政组织的政府和作为商业组织的部分社会团体(如测评机构)之间的组织冲突一直客观存在,高等教育质量内部管理与外部监控的关系这一涉及大学自治和学术自由的争论似乎也从未停止,即便组织领导共同倾向于实证主义管理范式,以使他们进行快速、高效的决策。在高校内部,管理人员、教师和学生基于不同的认识逻辑和价值立场,对学生学习成果评估的参与度与支持度也不尽相同。比如在管理者提出诸多评估要

求和进行各种评估培训的同时,一些教师感觉被评估“绑架”了,还有一些教师担心应试教育应运而生,从而破坏了高校的多元共存与个性发展。

不可否认的是,现代大学已超越传统的“自治”而进入一个质量“共治”的新时代。事实上,美国早期殖民地学院的学生学习评估就有外界的参与,如在17世纪70年代哈佛学院的毕业典礼上,地方官员、牧师和绅士等校外人士都会前来见证学位候选人的论文答辩,这说明对学生学习成果的关注从来就不仅仅是学校内部的事情。在高等教育与社会发展息息相关的今天,高等教育质量已是众多利益相关者共同关心的焦点,他们对教育质量的诉求和参与质量管理的热情更是日益增长。高校、政府、民间组织虽然因不同的组织特性和价值立场而客观存在着一些组织冲突,但基于促进学生发展和成长这一共同的组织目标,回应社会问责这一共同的组织责任,三方求同存异、重叠共识,确保学生学习成果评估在高校内部和外部影响的张力场域中坚守学术自由底线,坚持内部评估为主,同时发挥外部力量的正向功能,形成合力,增加评估的活力和动力,共同推动学生学习成果评估的良性发展,从而推动高等教育质量的不断提升。

# 美国大学生学习成果标准化 评估工具的分析研究

周廷勇 杜瑞军 张敬雨

## 一、美国大学生学习成果标准化评估工具的发展背景

在高等教育问责和认证意识的推动下,美国联邦和各州政府的政策制定以及社会公众讨论中的质量话语的不断汇集都要求高等教育机构向社会提供证据以表明其“教育质量”。大学生学习成果成为美国高等教育机构自我改进及其向社会提供证据的主要依据。许多测量和评估大学生学习成果的工具由此产生。据统计,当前美国大学生学习成果的测量和评估工具大约有 252 个。从开发工具的主体和运行机制看,主要有学术性、商业性以及学术与商业相结合三类。这些工具中,有的用于大学新生,有的用于大四学生,还有的用于全体大学生,也有的用于不同学科和专业的学生;有的关注大学生的通识教育学习成果,有的关注大学生在校学习经验和投入-产出,有的关注大学生在价值观、态度和精神等方面的学习成果。从测量和评估的性质看,主要分为直接测量(标准化测试)和间接测量(问卷调查)两类。它们在美国高等教育界都有相当大的影响。问卷性质的学习成果评估工具,如

“大学生就读经验问卷”(College Student Experience Questionnaire, CSEQ)和“大学学习投入问卷”(National Survey of Student Engagement, NSSE)等,都被相继引进到中国。标准化的测试工具主要包括:美国大学考试中心(American College Testing Program, ACT)的“大学学业水平评估考试”(Collegiate Assessment of Academic Proficiency, CAAP);美国教育考试服务处(Education Testing Service, ETS)的“能力测试”(Proficiency Profile, EPP);美国教育资助委员会(Council for Aid to Education, CAE)的“大学学习评估”(Collegiate Learning Assessment, CLA)。大学生学习成果标准化测试的发展是美国大学学习成果评估发展史最重要的组成部分。美国教育部将教育考试服务处的“能力测试”和教育资助委员会的“大学学习评估”奉为众多学生学习成果评估工具的典型,认为它们能够为美国高中后教育提供有意义的学生学习成果的评估报告。

本文主要对 CLA、CAAP 和 EPP 的评估指标、评估形式与方法以及评估结果的运用等进行比较分析。研究文献来源于这

三个评估工具官方网站的相关材料和美国学术界的相关研究文献。这三个评估工具植根于美国客观化、标准化测试的发展历史与传统。1937年开创的美国研究生入学考试(Graduate Record Examination, GRE)是这三个工具的直接影响源。GRE最初由卡耐基基金会承办,1948年被移交给美国教育考试服务处。在教育考试服务处的管理之下, GRE发生了从测试学生对知识内容的掌握到测试学生一般推理能力的转向。这一转向影响了后来这三个工具重视学生思维能力和发展水平的测量倾向。这三个工具也得到了由美国公立及赠地大学协会(Association of Public and Land-grant Universities)和美国州立学院和大学协会(American Association of State Colleges and Universities)联合成立的美国自愿认证体系(Voluntary System of Accountability, VSA)的认可和推荐。自2010年美国自愿认证体系成立之初,它就以这三个工具开展评估实践及研究,并试图将结果放到《大学详解》(College Portrait)中去。

“大学生学业水平评估”(CAAP)的历史可以追溯至20世纪70年代美国大学入学考试中心开发的“大学学习成果测量”(College Outcome Measures Program, COMP)。经历20来年的发展后,美国大学入学考试中心根据高校在使用过程中的反馈对COMP进行了改进,使其在信息量增大的同时,也更为简洁且易于实施,并将之改名为“大学生学业水平评估”。“能力测试”(EPP)最早可以追溯到1987年美国教育考试服务处推出的“大学生学术能力测试”(The Academic Profile Test)。2006年,该工具被更名为“学术熟练程度

与进步测量”(Measure of Academic Proficiency and Progress, MAPP)。2009年8月,ETS将MAPP再次更名为EPP,但这次更名并没有改变这个测评工具的形式和内容,而是沿用了2001年至2006年使用的版本。EPP与MAPP除了名称的变更之外,没有其他差异。“大学学习评估”(CLA)是美国教育资助委员会负责开发的,它于2000年首次在美国高校发起评估。CLA在经历了由美国教育资助委员会和兰德公司共同管理的一段时期之后,于2013年又由美国教育资助委员会独立管理。CLA产生的时间虽然比CAAP和EPP要晚,但它具有后发优势,其影响比CAAP和EPP更大。经济合作与发展组织(Organization for Economic Co-operation and Development, OECD)在“高等教育学习成果评估”(Assessment of Higher Education Learning Outcomes, AHELO)项目的开发过程中,专门邀请开发和管理CLA的美国教育资助委员会的专家参与了AHELO的可行性研究。

## 二、CLA、CAAP和EPP三个评估工具的异同分析

### (一)CLA、CAAP和EPP的共同关注点

尽管CLA、CAAP和EPP这三个评估工具产生于不同的时期,由不同的机构或组织开发和管理,但它们具有四个相同之处。

其一,三者都是从大学生通识能力的角度理解学习成果的概念。从目前的相关讨论来看,大学生学习成果的定义主要有两个视角。一是学科和专业素养的视角。该视角假设不同学科门类的知识有其自

身的边界,由这些知识组建起来的能力和素养也就有其自身特定的指向性。二是通识能力的视角。该视角认为,不论学生投身于何种学科门类和专业知识的学习,也不论学生将来从事何种专门工作,许多基本的、通识的能力对获得成功是非常重要的。美国兰德公司的研究人员发现,与学科有关的知识能力相比,通识能力恰恰是美国大学教育中普遍强调的目标之一,在大学生个人的职业生涯发展和生活中发挥着更加重要的作用。在高等院校为其所开课程安排的阶段性的考试阶段,美国大学生的学科和专业能力得以评估。CLA、CAAP 和 EPP 这三个评估工具则是通过标准化考试测评大学生的通识能力或一般综合能力。

其二,三者都非常重视评估大学生的批判性思维能力。大学生批判性思维教育自 20 世纪 80 年代以来就受到美国高等教育界的重视,已成为美国各高校教育目标和预期的教育成果的重要方面。美国也因此产生了许多测评学生批判性思维能力的工具。例如,由恩尼斯(R. H. Ennis)和米尔曼(J. Millman)开创的康奈尔批判性思维测试(The Cornell Critical Thinking Test),由斯滕伯格(R. Sternberg)主持编制的三元智能测验(The Triarchic Test of Intellectual Skills),由费星(P. Facione)等人编制的加利福尼亚批判性思维技能测验表(The California Critical Thinking Skills Test)和加利福尼亚批判性思维倾向问卷(The California Critical Thinking Disposition Inventory)。CLA、CAAP 和 EPP 借鉴了这些测评工具的经验。它们专门评估大学生的批判性思维能力,其测评题材主要取自大学课程学习和大学生生活经验。它们对

批判性思维能力的测试主要考察被试是否以及在多大程度上能够辨别一个说法的正确性、合理性并给予有力的阐释和论证。CAAP 的批判性思维评估主要选取大学课程中具有代表性的涉及常见话题的四篇文章,每篇文章都会陈述一个或多个观点。在文章后面的单项选择题中,要求学生阐述、分析、评价和扩展原文中的观点。EPP 也采用单项选择题的形式,测试学生在人文科学、社会科学或自然科学方面的批判性思维能力。CLA 采用主观题测评批判性思维能力,通过情景式问题,要求学生运用已有知识和技能对所给信息进行分析、理解、辨识和遴选。在分析型任务中,CLA 要求学生对所给观点进行判断、评价和论证。

其三,三者都重视测评大学生的写作能力。CAAP 重视全面测评大学生的写作能力,单独开设了两个写作测试模块。一个模块包含 6 个散文段落,反映了学生经常遇到的各种修辞情境,通过单项选择题的形式测试学生对于标点符号、语法、句子结构、文章结构、写作技巧、风格等标准书面英语规范的掌握程度;另一个写作模块是直接测量写作技能,要求学生根据简短的情境提示在规定时间内创作两篇独立的作文。EPP 的写作模块采用单项选择题的方式,测评学生在语法、句型结构、文章连贯性、修辞手法、句意理解等方面的能力。CLA 全部采取主观题:执行型任务试题要求学生在规定的时间内围绕所给问题撰写信件、备忘录或类似题材的文书;分析型任务试题要求学生针对所给主题发表观点或评论所给的观点,通过这些主观题来综合测评学生的书面沟通和写作能力。

其四,三者都采用“价值增值”(Value-added)的评估理念与方法。价值增值的核心思想是:研究者可以根据学生在前一段时期的测试分数预测学生在下一阶段可能取得的成绩。据此,价值增值评估能够表明学生是否取得了预期的进步,抑或超出了预期的进步。通过价值增值的方法还能够测量教师或学校对学生成就的长期影响。尽管美国学术界有不同的价值增值模式,但在高等教育评估领域,主要是以学生进入大学前或入学初和学生离开大学时这两个时间段为参照点,观察和评价学生在这段时期的变化和发展,以此判断学校对学生学习和发展的影响程度。CLA选取学生在高中阶段的学术能力评估考试(SAT)或大学入学考试(ACT)的成绩作为参照控制学生样本的能力基础,对比秋季入学的大一新生的CLA得分和春季毕业的大四学生的得分来测评学生在大学期间的价值增值。CAAP主要将它对大四学生的测评结果与学生在高中时参加大学入学考试的分数进行对比,以评估学生在大学期间的价值增值。EPP选择一个能充分代表全部学生特征的学生群体样本,在大学期间的不同时间点,对他们进行测试。

## (二)CLA、CAAP 和 EPP 的主要不同点

由于主办机构和创办初衷等方面的差异,CLA、CAAP 和 EPP 在评估对象、评估指标、评估内容、评估形式和结果报告等方面都各有不同。

其一,评估样本的选取方法不同。CAAP 建议院校选取那些参与 ACT 组织的大学入学考试的学生群体为评估对象。也有的院校在大一和大四两个阶段分别

用 CAAP 测试学生群体的发展趋势,确定院校有待改进的方面并找到亟待学术指导和干预的学生。EPP 根据测试对象的人数分为标准版和简化版两种:标准版对测试对象人数没有要求;简化版则由于时间短、题量少,为了提供科学的测评结果,仅仅用于测评至少由 50 名学生组成的群体的信息。EPP 对测评对象的年级没有限制,建议院校测试所有学生,或选择一个能充分代表全部学生特征的学生群体样本,在大学期间的不同时间点对他们进行测试。CLA 选择秋季入学的大一新生和春季毕业的大四学生这两个群体进行测评,通过比较两个群体的得分差异来测评一所高校的学生在大学期间的价值增值。

其二,除了测评批判性思维和写作能力外,三个工具各有其独特的评估指标。CAAP 还测评学生在数学、阅读和科学推理方面的能力。在数学测试模块,CAAP 通过 35 道单项选择题测试学生解决基础代数问题和大学代数问题的熟练程度以及大学数学课程中的定量推理能力。阅读模块由 36 道选择题组成,所选文章为大学课程中常见的散文、小说、人文科学、社会科学和自然科学作品中的选段,要求学生概括意义、处理信息、进行比较和归纳并得出结论。科学推理模块考查学生从大学入门科学课程中习得的科学推理能力,共有 45 道单选题,题干通常为研究数据代表、研究摘要或相互矛盾的观点,要求学生理解所提供的资料、检验相互关系、概括和提炼出新的信息,进而得出结论或进行预测。

EPP 还测评学生的阅读能力和数学能力。阅读测试模块要求学生能够理解关键词汇的意义、辨识文章的主旨和大意、

进行合理推断以及把握文章的修辞方法。数学测试模块也通过单选题测评学生在掌握数学术语、理解图表、评价定理和公式、分析百分比率、认读科学单位以及认识和运用数学公式与表述方式等方面的能力。

CLA 的评估指标还包含分析推理、问题解决和书面沟通的能力。就出题方式而言,CLA 并不像 CAAP 和 EPP 那样用独立的试题模块测量写作能力、批判性思维能力等各项评估指标,而是遵循效标抽样(criterion sampling)的测量原则,认为整体大于各个部分相加之和。因此,CLA 的试题是围绕复杂的现实生活情境开展的任务解决题型,要求学生综合运用所学知识和技能去探索解决问题和完成任务的方法,通过分析学生对于一个复杂任务的完成程度,对其关注的能力维度进行判断。

其三,评估的内容架构和试题形式不同。CAAP 提供 6 个独立的测试模块,包括写作能力测试、数学测试、批判性思维测试、作文测试、阅读测试和科学测试。院校可以根据自身的需求自由选择那些最能反映自身总体教育项目的目标和课程的测试模块对学生进行测评。CAAP 的测试模块除了作文测试之外,均为一定数量的单项选择题,采用笔试的形式,每个模块均可在常见的 50 分钟课堂之内完成。作文测试要求学生针对所给的情境,给出一个清晰的观点,提出支持这一论断的理由和证据,并创作一篇流畅的、符合逻辑的作文,以考查在限定的时间内、在不拟草稿的情况下,学生的临场写作能力如何。此外,院校还可以自主地按需增加最多 9 道基于本校的单项选择题。

EPP 包含批判性思维能力、阅读能

力、写作能力和数学能力四个方面的试题。它的标准测试形式包含 108 道单项选择题,每一方面的题目为 27 道,测试时间为两个小时。为了帮助院校节约时间和成本,EPP 还推出了专门测评 50 名以上学生群体的简化版能力测试。简化版能力测试将 108 道标准形式的试题拆分为三个部分,每个部分包含 36 道试题,分别称为“简化形式 1”、“简化形式 2”和“简化形式 3”,分别由三分之一的学生作答。EPP 在答题方式和考查内容上也给予高校个性化的空间:院校可以依据自身情况选择笔试或在线考试;校方最多可以增加 50 道自主编写的选择题和一篇作文,以满足其特定的评估需求。

CLA 的试题全部为主观题,分为执行型任务和分析型写作。执行型任务的题材均来自现实生活的各个领域,要求学生在 90 分钟内根据一个具体的情境撰写信件、备忘录或类似的文书给上司、同事或公司部门。试题包含一个放有若干文档的文件夹,其中有图表和数据,但是可信度和相关程度不等,有些信息甚至可能相互矛盾。学生应先阅读和评估所提供的信息,然后整理出有效的证据,综合归纳出结论,进而给出有说服力的解决方案。分析型写作包括两个题型——讨论型任务和评论型任务。讨论型任务要求学生在 45 分钟内根据题干所给的话题陈述和扩展自己的观点并进行论证;评论型任务让学生在 30 分钟内从所给的一段文字中辨识和描述其中的逻辑错误,并给出自己的理由。由于每个题型的作答时间较长,CLA 采取矩阵抽样(Matrix Sampling)的作答方法。每个学生并不是作答全部题型,而是抽取一道执行型任务试题或两道分析型写



作试题。CLA 的题目尽管是主观题,但采用机考的形式。整个评估过程都通过一个互动的网络平台完成,实现了无纸化的电脑管理。其中,分析型写作任务的答案由自然语言处理软件来评分,执行型任务由训练有素的人工读者进行网络在线评估。

其四,评估的信息和结果的报告形式不同。CAAP 用累积百分比报告学生个人与同一学校同一时期的其他考生相比的成绩水平。它还会为每个测试模块提供一个总分。由于单项得分有助于更准确地反映出教学计划的长处和短处,写作能力测试(用法/结构和修辞技巧)、阅读测试(艺术/文学和社会研究/科学)以及数学测试(基础和大学代数)还提供各能力次级指标的单项得分。EPP 为高校提供多个常模参照量表分数,包括总分、单项技能得分(批判性思维、阅读、写作和数学能力)以及单项领域得分(人文科学、社会科学和自然科学)。此外,EPP 还提供按照能力分类的标准参照分数,测量学生对于数学、阅读和写作三方面技能的熟练掌握程度。分数报告兼具图形和数据,显示学生在各自水平段的百分比。CLA 向院校提供总分和两个题型各自的分数。在计算总分时,执行型任务占 50%的比重,分析型写作中的评论型任务和讨论型任务各占 25%的比重。而分析型写作单独的分数则取评论型任务和讨论型任务的平均得分。CLA 通过比较大一新生和大四毕业生的测试成绩,可以计算出高校学生学习成果的价值增值得分。

### 三、CLA、CAAP 和 EPP 的应用、影响及存在的问题

CAAP、EPP 和 CLA 是当前美国评估

大学生学习成果最有影响力的直接测量工具。目前,全美约有 400 所高校使用 CAAP,其中包括公立和私立的两年制和四年制大学、技校和职业学校。EPP 则拥有全美范围内 500 多所高校、超过 550,000 名学生的测试结果。CLA 的应用范围更为广泛,在全美和国际范围内,有 700 多所高校采用 CLA 来测量大学生的价值增值,并将结果和其他院校的水平进行对比。美国高等教育机构采用 CAAP、CLA 或 EPP 能力测试来满足自身的需要,如了解大学生的学习效果、整体评估本校的通识教育项目、满足认证和绩效拨款的要求以及改进课程大纲和教学等。

这三种评估工具在美国高校的影响主要得益于它们在实践运作中经受住了信度和效度的检验。美国高校教育改进基金会(Fund for the Improvement of Post-secondary Education, FIPSE)于 2009 年组织测评专家对三者进行测试效度研究,得出三点结论:其一,在学生个人层面,CAAP 和 EPP 设置的客观题在同等测试时间内比 CLA 采用的主观题对单个学生能力的测量更加可靠和高效。其二,在院校层面,尽管 CAAP、CLA 和 EPP 的测评方式不同,但均表现出较高的信效度,都能够有效测出大学生在校期间达到的能力水平。其三,高校在选择适合自身的测评工具时,应该综合考虑学生、教职员工、管理者和决策部门对各个工具的接受程度,权衡测评的成本、执行的顺畅程度和测评的具体目的及需求,这样才能有效地达到提高教学质量的最终目标。不过,我们对 CLA、CAAP 和 EPP 的发展历史和运作机制的考察发现,它们的影响力根源于数量统计和计算机技术、大学生学习和发

展的理论研究以及专业组织这三者的有机结合。

然而,这三种评估工具在实际运行的过程中也面临一些障碍。高校普遍反映,学生按自愿原则参与测试往往导致测试学生人数无法满足测试样本的大小。由于这些工具对学生的收益考虑得并不是很多,学生参与这类评估的动机不强、积极性不高,导致无法进行测试或者是测试结果缺乏代表性和可靠性。此外,有的高校在测评报告中反映,本科生和监考老师的日程安排、课外时间和测试考场及设备都需要妥善安排和调度,才能保证测评的顺利开展,但是不少高校的现实条件往往很难充分满足这些需求。除了高校本身在测评的推行过程中遇到的各种现实困难之外,从事大学生学习成果测量的学者对这些评估工具也提出了一些质疑和批评。例如:库(Kuh)对CLA的信度和实用性表示担忧,他认为CLA的执行型任务题没有明确而直接的标准答案和评分基准,缺少衡量学生个人成绩的结构效度。班塔(Banta)和派克(Pike)对价值增值的适当性提出了质疑。他们指出,在一些精英机构中,学生的学习进步受到天花板效应的制约。与那些招收能力较低的学生的院校相比,价值增值的评估方法将令招收精英学生的院校在评估中处于不利地位。2010年出版的《EPP使用指南》指出,虽然EPP具有足够的信度来用于学生的咨询服务和辨识在个别能力方面存在困难的学生,但若基于这些分数做出高风险的决定(如分配奖学金),则不够可靠。

#### 四、对中国开展大学生学习成果评估的启示

随着本科教育质量意识的凸显,对大学生学习经验、学习投入和学习成果的调查与研究成为中国高等教育界的焦点和热点。只有对大学生的学习经验、投入和成果进行系统的调查研究,才能充分发挥它们对于本科教育质量的诊断作用。其中,评估大学生的学习成果尤为关键。通过对美国大学生学习成果评估的三种标准化测评工具的分析,我们得到如下启示:

其一,构建中国大学生学习成果的内涵是开展学习成果评估的首要前提。怎样才算是一个受过高等教育的人?中国高等教育体系对于大学生的学习成果有什么样的预期,又应该有什么样的预期?这些都是很难找到标准答案却又不得不去探索的问题。否则,提高本科教育质量就失去了方向。实际上,自有教育始,这些问题就被纳入到哲学家、思想家和民众的头脑中去了。“培养完整的人”,“培养德、智、体、美全面发展的人”,“培养创新型人才”,诸如此类的关于高等教育目的的表述见诸于学者的论著和演讲以及国家的教育方针与政策之中。然而,这些表述还过于凝练和抽象。美国高等教育评估界主要从认知、情感和技能三个方面去理解大学生的学习成果。虽然他们对这三个方面学习成果的描述有不同的观点,但是对于学生认知、情感和技能都涉及哪些方面的发展的描述都是非常具体的。CLA、CAAP和EPP这三种测量工具从个体通识能力的角度理解大学生学习成果的内涵,抓住批判性思维能力和写作能力这两个美国社会对高等教育以及美国高等教育对学生共同的、最重要的期待;与此同时,三者又各有自己对大学生学习成果的独特理

解之处。当然,这三种工具对大学生学习成果的理解并不完整,它们主要从认知和技能的角度理解大学生学习成果,而忽略了情感和价值观的角度。作为旨在培养“完整的人”、“全面发展的人”的高等教育,不应忽略学生的情感和价值观发展。

其二,自主开发既具有世界视野也切合中国高等教育实际和中国大学生特点的多样化的学习成果评估工具。学生的学习成果主要分为认知、技能和情感三个主要方面,它们分别表现在学生的心理和行为之中。学生在认知和技能方面的成果相对容易被观察,在情感上的发展则相对难以判定。学生的行为变化相对容易被捕捉,但追踪学生的心理发展却并非易事。在这种背景下,设计出一个或多个适当的工具来科学地搜集学生学习成果的数据和资料显得相当重要。评估工具的开发需要充足的经费、人员和机构等方面的支持和保障。CLA、CAAP和EPP这三个测量工具能够发展起来,与它们背后的机构、经费和专业研究团队的保障不无关系。尽管中国当前在引进国外相关评估工具时进行了本土化的改进,但工具隐含的理论和文化之根仍然属于他国,这些工具对于提出创新人才培养目标的中国高等教育体系和实践的适应性仍需不断调整和改进。在美国,对大学生学习成果的标准化测试工具和问卷调查形式的间接评估工具各有其优势和劣势。正是由于这些评估工具的多样化存在,它们之间能够相互取长补短。不过,评估工具的多样化也增加了美

国高校在工具选择过程中的时间、人力和物力成本,给高校的管理带来一定程度的困扰。在中国开展大学生学习成果评估,既应提倡多样化的评估工具,也应避免美国出现过的问题,使多样化的评估工具能够和谐共生。

其三,发展成熟的评估和数据分析方法。数据能否得以客观、科学地呈现,从而服务于学生学习成果评估的目的,关键在于以何种评估方法为指导来进行数据分析。CLA、CAAP和EPP这三种测量工具都是以“价值增值”的评估理念和方法进行数据分析。国内学界对“价值增值”这一概念早有所闻,但还很少有人运用这一评估理念和方法来开展大学生学习成果的评估研究和实践。在美国高等教育评估领域,存在不同的价值增值模式。不论何种模式,都涉及如何计算价值增值的系数问题。本文研究的三个评估工具价值增值系数的计算方法对中国开展大学生学习成果评估具有重要的参考价值。

综上所述,当前在中国推动大学生学习成果评估,除了要调整评估的制度和政策导向等问题外,还要根据中国社会经济文化背景明确大学生学习成果的内涵,设计开发本土化的评估工具,发展更成熟的评估和数据分析方法。在引进和应用其他国家的评估工具时,尤应注意这些工具的文化背景和理论基础,汲取它们在实践中运用中的正反两方面经验,如此方能为提升中国本科教育质量做出切实贡献。

# 经合组织国际高等教育学习成果评价项目——AHELO 解析

廖 青

对于世界上所有的国家和地区来说,明日劳动力大军的素质将是决定其发展快慢和财富多寡的关键。而在当今瞬息万变的时代,高质量的高等教育已经成为保证经济、科技和社会进步的重要手段。正是基于上述考虑,政府和个人都对高等教育投入了巨大的资源。然而,由于高等教育的复杂性和多样性,对高校的评价方式不一而足,高等教育的国际评价更是众说纷纭。

在此背景之下,经济合作与发展组织(Organisation for Economic Co-operation and Development, OECD)将高等教育评价的重点放在高校的教与学上,并力图找出一系列影响高等教育发展的因素。其所倡导的“高等教育学习成果评价”(Assessment of Higher Education Learning Outcomes, AHELO),正是这一理念的产物。

2008年,经合组织教育部长会议在东京举行,其中高等教育质量评估是会议的重要议题。在咨询国际专家的意见之后,经合组织最终做出决定,即组织“高等教育学习成果评价”项目的研发工作,并将其作为 OECD 在高等教育领域的最大项目。

AHELO 的测评对象是即将获得学士学位的大学生,参与其中的准毕业生们将面临两个方面的测评,即所有大学生都应该具备的“通用技能”(Generic Skills)和不同专业学生所必须掌握的“学科专业技能”(Discipline-Specific Skills)。按照时间计划表,AHELO 项目在 2010 年 1 月至 2011 年 6 月完成测评工具的开发并对其进行小规模的有效性检测,2011 年 6 月至 2012 年 12 月在参与此项目的高校进行测评,2012 年底讨论可行性研究结果。目前,已有 16 个国家参与了 AHELO 项目的可行性研究。

## 一、高等教育学习成果评价项目的管理

经合组织秘书处负责 AHELO 项目的前期筹备、监督、管理和协调工作,具体负责现阶段可行性研究的则是高等教育机构管理(Institutional Management in Higher Education, IMHE)。澳大利亚教育研究委员会(Australian Council for Educational Research, ACER)领导的国际化专家团体——教育援助委员会(the Council for Aid to Education, CAE)、OECD 和参与国的国家团队协作管理这项工作的进行。

### 1.AHELO 项目的资金来源

目前,可行性研究主要由参与国提供经费支持,这笔经费大多来自相关的基金会,包括美国路明娜教育基金会(Lumina Foundation for Education)、意大利圣保罗银行基金会(Compagnia di San Paolo)、美国休利特基金会(Hewlett Foundation)、葡萄牙古尔本金安基金会(Calouste Gulbenkian Foundation)、瑞典中央银行基金(Riksbankens Jubileums fund)和美国斯宾塞基金会(Spencer Foundation)。此外,英国高等教育基金委员会(the Higher Education Founding Council for England, HEFCE)和爱尔兰高等教育局(Higher Education Authority of Ireland, HEA)也是 AHELO 项目经费的资助者。

### 2.AHELO 测评工具及背景问卷的开发

目前的通用技能测评工具和经济学、工程学评价工具分别由不同的专家小组领衔完成。

通用技能测评工具改编自美国大学学习评价(Collegiate Learning Assessment, CLA);美国教育考试服务中心(Educational Testing Services, ETS)领导了 AHELO 经济学框架和经济学评价工具的开发;工程学框架和评价工具的开发由澳大利亚教育研究委员会(Australian Council for Educational Research, ACER)牵头,由一些国际组织参与共同完成,其中包括日本国家教育政策研究所和意大利佛罗伦萨大学组织的欧洲工程师网络。目前,澳大利亚教育研究委员会领导的联合会、高等教育政策研究中心(Centre for Higher Education Policy Studies, CHEPS)和中学后教育研究中心(the Center for Postsecondary Re-

search, CPR) 已经开始着手背景问卷的开发工作。

### 3.AHELO 可行性研究的实施者

除了经合组织之外,每个项目的参与国家也分别成立了两个机构,以便更好地完成测评工作。其中,国家专家组(Group of National, GNP)负责测评技术方面的工作,国家项目负责人(National Project Managers, NPMs)则负责项目的具体实施。

此外,参与可行性研究的大学或机构在实施评价之前,也都要找小部分学生对相关评价工具进行预调查。预调查包括两个部分,一是参与者在认知实验室鉴定翻译的准确性,二是在座谈会中进行初期反馈。

## 二、现阶段高等教育学习成果评价项目的运行

目前,OECD 正在通过小规模试验确认适合在全世界范围内使用的工具,这表明,测试工具的跨文化有效性将成为一个十分重要的标准。AHELO 项目已经成功地研制出通用技能问卷、经济学和工程学问卷,由于 OECD 还立志于找出影响高等教育学习成果的因素,因此,学生的背景问卷亦有着十分重要的意义。图 1 为 AHELO 测评工具的简单流程。

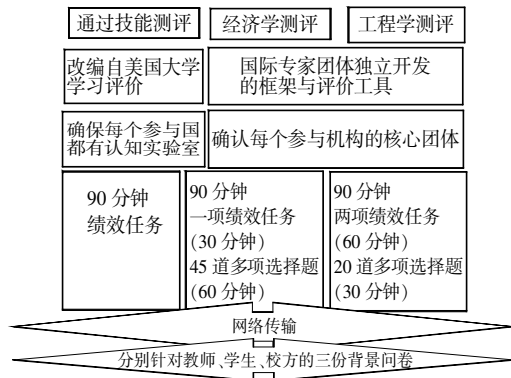


图 1 AHELO 测评工具流程图

## 1.通用技能测评工具的开发与实施

由于通用技能测试工具改编自CLA,为了保证该测试的跨文化有效性,参与国纷纷对其进行了必要的翻译和修改。通用技能测评工具的翻译过程包括如下步骤:由资深翻译者进行双向翻译;调整和复查双向翻译;在认知实验室检测经过调整的双向翻译;基于认知实验室的发现调整各国的翻译。

其中,认知实验室在测评工具的检测过程中发挥了重要作用。通常情况下,认知实验室需要找到几名符合条件的大学学生做测试,同时用语言表达他的想法。而绩效任务是通用技能测试的重要部分,认知实验室旨在确保绩效任务在从英语翻译成其他语言时达到以下要求:测量的概念不发生变化;翻译内容符合最初的设计;难度保持不变;每个参与通用技能测验的国家都要指导认知实验室对6至13名大学生施测。事实证明,认知实验室在保证翻译的准确性上发挥了重要作用。

接受通用技能测试的学生将被要求用90分钟时间完成一个或者两个呈现给他的绩效任务。该任务要求学生应用一系列综合技能,包括批判性思维、分析推理、问题解决和协作交流,来回答大量的开放性问题。这些开放性问题都假设了一个现实的情境,需要学生从不同来源(书信、地图、备忘录等)来收集必要的证据。

## 2.学科专业技能测评工具的开发与实施

学科专业技能测试旨在考察学生在各自专业领域的能力。AHELO项目组认

为,学科专业技能的测评不应局限于事实性知识的考察,而应将学生置于新的情境之下,创造性地应用知识。鉴于不同国家的高等教育有着巨大的差异,AHELO聚焦于受文化差异影响最小的学科——经济学和工程学。如果针对上述学科的测评工具取得了成功,OECD将着手其他学科测评工具的开发。与通用技能测评工具的开发不同,经济学和工程学测评工具由OECD和相关领域的专家共同研制开发。

### (1)经济学测评工具的开发与实施

AHELO项目组认为,尽管世界各国对于经济学学科专业技能应该包括什么仍存在很大的争议,但有一些是大家共同认可的。例如:解决经济问题的方式,当代经济学家所要达到的目标,以及达到这些目标所需要克服的约束和限制。基于此,AHELO将重点考察学生用“经济学语言”解决实际问题的能力,包括知道并且理解基本的经济学概念,会应用这些概念,会使用恰当的工具来评估某个经济学问题,会分析数据并表达其分析结果。

经济学测评工具的开发分为两个步骤:一是经济学框架的制定;二是基于框架的经济学测评问卷的制定。美国教育考试服务中心(Educational Testing Services, ETS)引导了经济学框架的开发以及经济学测评工具的评估。这一框架注重五项学习目标,如表1所示:

五项学习成果	子目标
学科知识和理解	清晰连贯地掌握经济学语言。包括清晰地定义宏观和微观经济学领域中的标准术语,解释基本概念的能力
	清晰连贯地掌握宏观和微观经济学的原理,同时在经济学论辩中具备用清晰且让人心信服的语言阐释观点的能力
	解释经济主体(个人、家庭、公司、政府等)如何做出选择,并解决相关的经济决策问题
	有能力解释经济系统的基本运作方式,并分析政策在这个系统中扮演的角色
	有能力找出模型和分析方法的缺点与不足
学科知识及应用	在特定研究领域中有应用经济学推理和方法的能力(如市场、财政、环境、健康、劳动力市场、国际贸易等)
	运用经济学推理形成和评估经济方面的建设和政策,上述建设既包括针对个人的,也包括对公共部门
有效运用相关数据和定量分析方法	对经济学和社会数据信息来源有足够的了解,包括如何找到和在哪里找到这些数据,以及分析处理数据的能力
有效交流	向专业和非专业人士传达和解释经济学参数的能力,既包括口头表达也包括书面表达,可能涉及运用网络、电脑和投影设备的能力
学得独立学习技能	就某一系列经济学问题进行战略性和批判性思考
	在经济学领域提出和探索问题的能力。包括判断问题的研究价值,知道运用合适的方法进行研究并得出结论的能力。

表 1 经济学测评框架所重视的五项学习成果

经济学测评时长 90 分钟,包括一个复杂的绩效任务和大约 45 道多项选择题。这两项任务都避免引导学生回忆简单的知识,而是将重点放在超越内容的技能上,包括概念的应用以及应用适当的统计学和非统计学工具得出结论并对政策提出建议。经济学评价最初是用英语开发的,现在已经被翻译成意大利语、俄语和荷兰语。

## (2) 工程学测评工具的开发与实施

与经济学不同,世界各国专业人士在工程学及其人才培养方面所达成的共识要广泛的多,许多国家都有工程学的学习

框架。AHELO 工程学框架和工程学测评问卷旨在考察学生应用工程学语言的能力以及专业工程师所必须掌握的非技术能力。

工程学框架和评价工具的开发由澳大利亚教育研究委员会 (Australian Council for Educational Research, ACER) 领衔,由一些国际组织参与共同完成。工程学测评框架认为该学科的学习成果应该包括一般技能、基本技能和工程科学、工程分析、工程设计、工程实践等五项学习成果。具体如表 2 所示。

五项学习成果	子目标
一般技能	在团队中有效地行使自己职权的能力
	在工程师团体及其他社团进行有效交流能力
	在今后的生涯中意识到并切实参与终身学习
	对与工程学相关的多学科背景有一定了解
基本技能和工程科学	在所在分支工程学的科学和数学规则有一定的知识和理解
	在所在分支工程学的关键方面和概念有系统的理解
	对所在分支工程学有全面的了解,包括最新的热点话题
工程分析	应用所掌握的知识,通过已有的方法来鉴别、表达和解决工程学问题
	应用所掌握的知识来分析工程学产品、过程和方法
	选择和应用相关的分析与模型方法
	有能力进行文献检索,会使用数据库和其他来源的信息
	设计并进行恰当的实验、解释数据并得出结论的能力
工程设计	为满足特定要求,应用所掌握的知识来设计恰当的方案
	理解设计方法学,并知道如何将其用于实践当中
工程实践	选择恰当的工具、设备和方法,并将理论与实践相结合以解决工程学问题
	理解技术和方法的应用以及他们的局限
	理解工程实践中的非技术含意
	有能力掌握车间和实验室的各项技能
	理解工程实践中的健康、安全、法律以及责任问题,理解工程解决方案对社会和环境的影响,并对专业道德、责任以及规范恪守承诺
	了解项目管理和商业惯例,如危机管理等

表 2 工程学测评框架所重视的学习成果

工程学测试时长 90 分钟,包括两个复杂的绩效任务和大约 20 道多项选择题。上述任务将重点放在超越内容的技能上,包括概念的应用以及恰当运用统计学和非统计学工具得出结论和政策建议的能力。绩效任务给学生呈现特定环境下的一个真实的工程学方案,并介绍一系列相关的背景内容。这一模式旨在将学生置于真实情景中,允许学生发挥他们的创造性思维。多项选择题重在考察工程科学,是考察学生对于工程学知识和技能掌握情况的一种快捷有效的方式。

工程学评价工具已经从英语翻译成了日语和西班牙语,译成阿拉伯语和斯洛伐克语的工作正在进行中。参与评估的国

家大都进行了分组座谈会,以确保学生了解评价任务的进展,并保证该评价在不同文化背景下的有效性。从学生最初的反馈来看,真实情境的使用的确激发了学生对该任务的兴趣。

### 3.背景问卷的开发

AHELO 项目不仅对高等教育学习成果进行评估,还试图找出影响学习成果的各项因素。

受测学生的测试结果可能受许多变量的影响,包括学生的家庭环境、最初的知识 and 技能,学生的努力程度、教学资源 and 教学质量等等。通过分析背景数据和学习成果表现,可以分析导致大学取得高成就的原因,从而判断并帮助学校应该采取



何种行动以使学生的学习成果最优化。通过这种做法,AHELO 将会帮助鉴别对于不同背景不同学生而言,到底是什么因素在起作用,这在减少辍学率和加强学习成果方面很有潜力。

基于此,所有参与 AHELO 的学生和老师以及大学机构都要完成一个十分钟的背景问卷,以便收集上述投入和过程变量的背景信息。AHELO 十分关注学生学习成果,并检测所有起作用变量的学习效果,评价包括学生取得成就的水平以及教学系统的运转效率如何。

目前,澳大利亚教育研究委员会领导的联合会、高等教育政策研究中心(Centre for Higher Education Policy Studies, CHEPS)和后中学教育研究中心(the Center for Postsecondary Research, CPR)已经着手开发这项评价的问卷。

### 三、对高等教育学习成果评价项目的思考

#### 1.明确的学习成果评价是否有助于提升学习质量

《AHELO 经济学预期学习结果概念框架》指出,概念明确、能被师生共同接受的目标有助于提高课程的质量,为达成学习目标而制定的清晰的过程和方法有助于提高教学质量。在学习成果背景之下的学生评价要求清晰地定位目标、测量成就,更重要的是找到实现目标还缺少哪些因素。此外,AHELO 测评工具的开发者还致力于使毕业生们所到达的学习成果与社会需要、雇主要求以及国际社会的学术要求保持一致。AHELO 项目组相信,明确的学习成果评价有助于提升学习质量。

OECD 组织世界高等教育评估项目的

初衷并不是制造新的世界大学排行榜,而是聚焦于学生学习成果,希望通过这种评价来促进高校教学质量的提升,进而提高学生的成果。然而,美国著名高等教育学家阿特巴赫(Philip G. Althach)指出:“实际上,没有广泛接受的测量教学质量的方法,评价教育对于学生的影响是一个远未得到探讨的领域。”究竟国际性的高等教育评估是否有助于高校教学质量的提升,这是一个值得深思的问题,也是影响世界各国参与 AHELO 项目的重要因素,尽管 OECD 在国际性教育评估领域有着丰富的经验,如影响广泛的国际学生评价项目(Programme of International Student Assessment, PISA)。

#### 2.分析学生的背景因素是否有助于提升学习质量

AHELO 项目组认为,诊断高等教育质量的工具很多,排行榜即为其中之一,尽管他们一般基于可获得的数据,但却忽略了教与学。而 AHELO 则致力于测量不同文化和语言下,不同设置、不同使命高校的学习成果。从这种意义上讲,AHELO 将通过丰富的背景资料来诊断教与学的质量。基于此,项目组认为,分析结果有助于在诊断高校的问题之后进一步提出解决方案。

影响学生表现的变量很多,包括学生的背景、最初的知识储备和技能、课程设置、学生的努力程度、教学资源、实践和教学质量。所有参与 AHELO 的学生、教师 and 高校都要完成一个十分钟的背景问卷,项目组将通过这一措施收集有关投入和产出变量的背景信息。基于此,该项目将评价学生取得成就的水平以及教学系统的运转效率如何。

那么,在高等教育领域,分析不同成就水平学生的背景因素,是否就能提出可行的方案,并将之付诸实践,从而提升教学效果呢?或许 AHELO 项目在未来的运行中能够回答这一问题。

### 3.国际性高等教育评估是否具有可行性

AHELO 是一项国际性的高等教育评估,难免被人质疑为新的世界大学排行榜,OECD 对此进行了坚决否认。首先,AHELO 并无意提供一份世界大学的排行榜,其次 AHELO 的评价指标只是学生的

学习成果。

从目前来看,大多数世界大学排行榜将评价的指标给了科研,上海交通大学的世界大学排行赋予科研以 90%的权重,而学习输出只占 10%;英国《泰晤士报》所提供的世界大学排行榜则将权重分别赋予了新生成绩、输入-教师、科研和声誉。尽管各式各样的高等教育评估有着不同的评价指标,但都饱受诟病。那么,完全聚焦于学生学习成果的 AHELO 项目能够提供什么新的思路呢?让我们拭目以待。

