

·域 外·

美国大学生学习成果标准化 评估工具的分析研究

周廷勇¹, 杜瑞军², 张歇雨³

(1. 中国传媒大学 高等教育研究所, 北京 100024;

2. 北京师范大学 教育学部, 北京 100875;

3. 佛罗里达州立大学 教育学院, 美国 佛罗里达州 塔拉哈西, 32306)

摘要:美国大学考试中心的大学学业水平评估、美国教育考试服务处的能力测试以及美国教育资助委员会的大学学习评估是在美国具有广泛影响的学生学习成果标准化评估工具。三者都从大学生通识能力的角度理解大学生学习成果,运用价值增值评估方法,侧重于评估大学生的高阶思维能力和一般综合能力,在评估指标上都重视评估大学生批判性思维能力和写作能力。与此同时,三者在选择评估对象、评估的具体指标、评估形式和评估报告的呈现形式等方面又各具特点。尽管这三个工具受到美国许多高校的信任和欢迎,但它们对学生缺乏吸引力,影响了其评估结果的可靠性。中国今后在应用、引进和开发此类评估工具时,应注意汲取这些工具的正反两方面经验,深入梳理中国大学生学习结果的内涵,着眼于自主开发既具世界视野又切合中国实际的大学生学习成果评估工具,发展更成熟的评估和数据分析方法。

关键词:大学生;学习成果;标准化评估工具

中图分类号:G649.712 **文献标识码:**A **文章编号:**1672-0059(2014)05-0084-07

A Study of Standardized Assessment Instruments of College Student Learning Outcomes in America

ZHOU Ting-yong¹, DU Rui-jun², ZHANG Xie-yu³

(1. Higher Education Research Institute, Communication University of China, Beijing, 100024, China;

2. Faculty of Education, Beijing Normal University, Beijing 100875, China;

3. Educational Leadership & Policy Study Center, College of Education, Florida State University, USA, 32306)

Abstract: Collegiate Assessment of Academic Proficiency (CAAP) from American College Test (ACT), ETS Proficiency Profile created by the Educational Testing Service (ETS) and Collegiate Learning Assessment (CLA) launched by the Council for Aid to Education (CAE) are three broadly influential standardized assessment instruments of student learning outcomes nationwide. The three instruments all employ a “value-added” assessment model in an effort to apprehend students’ learning outcomes through their general abilities in college. They focus on the assessment of students’ higher-order thinking skills and general comprehensive ability, with an emphasis on assessing college students’ development of critical thinking and writing skills. Nevertheless, they vary from each other in their selection of test-takers, specific assessment indicators, test formats and presentations of report. In spite of their reliability and popularity among large numbers of American higher education institutions, their lack of attraction to college students has threatened the validity of their results. When applying, importing and developing

this type of assessment instruments, China should be cautious of both experiences and lessons learned from them and gain an in-depth understanding of the contents of student learning outcomes in college. The ultimate purpose is to develop assessment instruments that both share a global perspective and fit the actual conditions in China. In particular, special attention needs to be paid to the development of mature assessment methods and data analysis approaches.

Key words: College Student; Learning Outcome; Standardized Assessment Instrument

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。^[1]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”奉为众多大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。^[2]

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。^[3]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。^[4]

“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”。^[5]“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)。^[6]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。^[9]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。^[8]

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。^[9]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)。^[10]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。^[11]

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。^[12]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。^[13]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试。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。^[14]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 Postsecondary 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。^[15]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。^[16]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。^[17]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。^[18]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。^[19]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。

参考文献

- [1] BORDEN V M H, KERNEL B. Measuring Quality in Higher Education: An Inventory of Instruments, Tools and Resources [EB/OL]. [2014-06-14]. <http://apps.airweb.org/surveys/>.
- [2] SHAVELSON R J. Assessing Student Learning Responsibly: From History to an Audacious Proposal[J]. *Change*, 2007(1): 26-33.
- [3] LIU O L. Measuring Learning Outcomes in Higher Education [EB/OL]. [2014-06-14]. http://www.ets.org/Media/Research/pdf/RD_Connections10.pdf.
- [4] Association of Public and Land Grant Universities & the American Association of State Colleges and Universities. About the Voluntary System of Accountability [EB/OL]. [2014-06-14]. <http://www.voluntarysystem.org/about>.
- [5] ACT. ACT: The First Fifty Years, 1959-2009 [EB/OL]. [2014-06-14]. http://media.act.org/documents/ACT_History.pdf.
- [6] YOUNG J W. Validity of the Measure of Academic Proficiency and Progress [EB/OL]. [2014-06-14]. <http://www.ets.org/s/mapp/pdf/5018.pdf>.
- [7] KLEIN S, BENJAMIN R, SHAVELSON R, BOLUS R. The Collegiate Learning Assessment: Facts and Fantasies [J]. *Evaluation Review*, 2007, 31: 415-439.
- [8] OECD. AHELO Brochure [EB/OL]. [2014-06-14]. <http://www.oecd.org/dataoecd/37/49/45755875.pdf>.
- [9] MORANTE E A, ULYSKE A. Assessment of Reasoning Abilities [EB/OL]. [2014-06-14]. http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_198409_morante.pdf.
- [10] ROBERT H E. Critical Thinking Assessment [J]. *Theory into Practice*, 1993(3): 179-186.
- [11] CAE.CLA Overview [EB/OL]. [2014-06-14]. <http://cae.org/performance-assessment/category/cla-overview>.
- [12] STEEDLE J, KUGELMASS H, NEMETH A. What Do They Measure? Comparing Three Learning Outcomes Assessments [J]. *Change*, 2010(4): 33-37.
- [13] Educational Testing Service. ETS Proficiency Profile: User's Guide [EB/OL]. (2010-06-21). [2014-06-14]. http://www.ets.org/s/proficiencyprofile/pdf/Users_Guide.pdf.
- [14] ROKSA J, ARUM R. The State of Undergraduate Learning [J]. *Change*, 2011(2): 35-38.
- [15] Fund for the Improvement of Postsecondary Education (FIPSE). Test Validity Report [EB/OL]. [2014-06-14]. https://cp-files.s3.amazonaws.com/26/TVSReport_Final.pdf.
- [16] LIU Q L, BRIDGEMAN L, ADLER R M. Measuring Learning Outcomes in Higher Education: Motivation Matters [J]. *Educational Researcher*, 2012(9): 352-362.
- [17] KUH G. Director's Message in Engaged Learning: Fostering Success for All Students [Z]. Bloomington, Indiana: National Survey of Student Engagement, 2006.
- [18] BANTA T W, PIKE G R. Revisiting the Blind Alley of Value-added [J]. *Assessment Update*, 2007, 19(1): 1-2, 14-15.
- [19] LIU Q L. Value-added Assessment in Higher Education: A Comparison of Two Methods [J]. *Higher Education*, 2011, 61: 445-461.

收稿日期: 2014-04-11

基金项目: 国家社会科学基金"十一五"规划教育学青年项目"高校学生学习质量评价研究"(CIA080222); 2009年教育部人文社会科学青年基金项目"研究型大学中青年教学骨干教师教学发展研究"(09YJC880010); 国家自然科学基金项目"我国本科人才培养质量研究"(71173022)。

作者简介: 周廷勇, 中国传媒大学高等教育研究所副研究员, 从事高等教育评估和大学生发展研究; 杜瑞军, 北京师范大学教育学部讲师, 从事高等教育研究; 张歇雨, 美国佛罗里达州立大学社会文化与国际发展教育专业硕士研究生, 从事跨文化教育和国际发展教育研究。