

关注学习效果：

美国大学课程教学评价方法述评

——美国“以学生为中心”的本科教学改革研究之六

赵炬明

【摘要】本文评述了美国大学在以学生为中心的教学改革中,在课程教学评价方面的主要创新,包括形成性评价、评价量表、课堂教学评价方法、真实性学习及效果评价、成果档案袋等五类评价方法,同时讨论了课堂学习效果评价方面的一些概念和理论问题,并以 NILOA 2018 年的调查报告为对象,结合美国在大学教学评价实践方面的经验和教训,给中国大学提供一些建议。

【关键词】以学生为中心 美国大学本科教学改革 课程教学效果评价 课堂学习效果评价 学习效果评价

本系列研究之五介绍了美国大学的教学质量保障系统,这篇文章则侧重介绍在课程学习效果评价(Course-level learning outcome assessment, LOA)领域的方法创新和经验教训。课程是本科教学的基础,如何在课程教学中评价学生学习效果,是教师们最关心的问题。课程教学评价也是美国大学教师做出很多创新的领域之一,值得专门介绍。由于本文只是一个知识地图,只提供基本情况和信息,因此,希望深究者可根据本文提供的信息做进一步的研究。

本文共七部分,首先讨论与课程学习效果评价有关的几个概念和理论问题,随后分五部分介绍形成性评价、评价量表、课堂教学评价方法、真实性学习及效果评价、成果档案袋五类方法。最后是对美国大学当前 LOA 实践情况的介绍和反思。

一、关于 LOA 的几个基本问题

本节讨论几个有关的概念和理论问题。首先是评价和评估。前一篇文章介绍了美国高等教育评价评估界对评价(assessment)和评估(evaluation)这两个术语的区分。评价和评估分别指两类不同的活动。评价指从课程教学角度对学习效果的评估,评估主要指从管理角度对学校教学工作和专业(program)质量的评估。从历史上看,美国是先有从管理角度对学校工作和专业质量的

评估,然后才有从教学角度对学生学习效果的评价。评价和评估这两个词在语义上是近义词甚至是同义词,区别主要在用途方面,因此试图从语义上区别它们常常会导致理解上的混乱。

另一个常见的语义混淆是评价和评分(grading)。实践中很多教师把给学生作业评分当作学习效果评价。确实,有时对学生的评分可以当作学习效果评价,但评分并不等于评价。评价是指教师根据教学目标,检查教学进展情况,以便改进教学。从控制论角度看,评价是根据目标检查行为,以便调整行为,确保最终达到预定目标。在课程教学中,有部分评分活动与此一致,还有部分与此不一致。例如对学生到课情况打分,结果计入总分。这种打分就不属于学习效果评价,而属于课堂管理。此外,不是所有学习效果评价都需要打分,例如课堂随机提问是教师检查学生学习情况的评价方式之一,但这类提问并不打分,也不计入学生成绩。这两种情况都表明,评价不是评分。这里区别是,评价是检查教学活动达标情况,评分是给学生成绩,两者目的和功能并不相同,故不宜混淆。

三是“learning outcomes”究竟应该如何翻译?是“学习效果”还是“学习成果”^①?其实,这不仅是个翻译问题,更是涉及“learning outcomes”究竟指什么的深层问题。对此,在英文文

收稿日期:2019-09-10

作者简介:赵炬明,华中科技大学教育科学研究院教授、博士生导师。

献中也是角度各异,人言人殊,争论不断。例如,从管理评估角度看,学习效果不应该只是学生的数学能力、写作能力、工程能力、编程能力、批判思维能力、创造能力等。这些都只是片段和局部。管理评估要看最后结果,看最后整体效果。也就是说,评估要看最后成型的东西,看学生毕业时是否达到预定培养目标。因此从评估角度看,学习效果评价的标准应该是学生的毕业率、保持率、就业率、专业对口率、起点工资^②等。这些标准是学校教学质量管理的指标。但对教师、对课堂教学来说,这些指标毫无意义,因为教师们无法根据这些指标来设计课程和组织教学。对课程教学来说,学生学习效果的指标应该是学生在学习过程中是否达到教师的预定目标。因此课堂教学评价需要另一类指标。这些指标是根据专业培养目标分解后的课程教学目标发展而来的指标,如计算能力、写作能力、编程能力、工程能力、辩证思维能力、创造能力、问题解决能力、合作交流能力、领导能力等指标。也就是说,关于学生学习效果的评价,确实存在管理和教学两个不同角度,且各自有其合理性,应当区别对待。故本文建议把管理角度的“learning outcomes”译为“学习成果”;而把从课程教学角度看的“learning outcomes”译为“学习效果”。这个区分至少有利于在中文文献中避免出现英文文献中那样的词义混淆,以及伴随而来的误解和无谓辩论。换言之,谈课程教学就用“学习效果”,讲专业及学位质量就用“学习成果”。这个区分也符合中文文献的表达习惯。

那么,如何把课程教学中的“学习效果”转变为质量管理要求的“学习成果”呢?答案是通过专业矩阵。^[1]在制定专业矩阵时,要从知识、技能、价值三个维度来规定每门课程的教学目标。如果制定的专业矩阵是合理的,那么各门课程只要完整实现专业矩阵确定的教学目标,就可以保证学生在毕业时能够达到专业预定的教育目标或“学习成果”。按这个原则,专业评估有两个重点:① 检查专业矩阵设计的合理性;② 检查专业内各门课程的教学是否完整实现其教学目标。抓好这两个关键点,就可以确保课程学习效果转变成专业学习成果。

第四个问题是关于学习效果测量及评价的科学性和可靠性。这涉及三个问题。一是学习效果本身的特点造成的测量和评价困难。学习效果有长期效果和短期效果。有些学习效果不是当下就

显现的,要经过很多年后才可能显现出来。典型的例子是苹果手机。苹果手机的设计以简洁性著称,当人们问乔布斯其设计思想来源时,乔布斯说是由于他早年在大学学过的一门美术字课以及后来他到泰国佛寺里的体验。这些早年学习经验在他后来的产品设计中发挥了作用。这种长期效果不是大学教学评价可以测量的。但这些效果显然是大学学习最宝贵的效果。

其次,即使对短期效果也存在测量困难。因为教学活动及其效果之间的因果性很少是一一对应关系,经常是一因多果和一果多因。造成这种现象的原因很复杂,目前的学习科学只能在一些基本原理的基础上“猜测”大脑中发生的事情,并用这些猜测来解释学习活动的因果关系^③,学习效果测量因此也并没有什么可靠的科学结论可依。目前绝大多数学习效果评价方法来源于教师们的实践经验,而非来自脑科学和学习科学。这种状态使得目前的学习效果评价很难称得上是科学的。^④

三是学习效果评价所用的数学工具的性质。在学习效果测量中广泛使用统计学方法,而统计学只能显示相关性而不能显示因果性。但理论的科学性要求结论建立在因果性基础上。这就严重地削弱了学习效果测量的科学性,更遑论由于多因多果特点而引起的数学建模困难和计算复杂性问题了。^[2]

在我看来,这三方面问题都堪称“阿喀琉斯之踵”(Achilles' Heel),严重削弱了学习效果测量与评价的科学性和可靠性。也正因此,我认为目前在 SC 改革的所有领域中,学习效果测量与评价是最短的短板。更糟糕的是,至少从目前还看不到这个领域会有什么令人期待的突破。

上述问题还导致了学习效果评价的另一个特点:经验性。由于上述种种原因,目前学习效果评价只能主要依赖经验,尤其是长期积累的有效经验,也就是说,目前学习效果评价很像工程与医学,在缺少可靠科学理论的情况下,只能依靠“经验公式”和“临床验方”。话说回来,尽管目前还说不出这些经验公式和验方的科学依据是什么,但我们知道它们在实践中有效。而且相信,凡有效者必有其合理性。从这个意义上讲,目前在 SC 改革中使用的学习效果评价方法很大程度上依赖为实践证明有效的经验。这个特点尤其值得注意,千万不要把它们简单等同于科学。当然,也要

注意,既然有效经验可能包含某种未知的合理性,就不宜简单地将其拒之门外。明智的做法是,在评估实践中首先寻找科学依据;如果不行则可依赖有效经验,以弥补科学理论之不足。这也是当前学习效果评价研究的现实处境。本文下面介绍的绝大多数 LOA 方法都有这个特点,它们基本上是美国大学教师在 SC 改革中积累的有效经验,可供参考借鉴。

第五个问题是 LOA 和教育技术。LOA 涉及两个活动:收集数据和分析数据。如果用传统方式,它们会占用大量时间,让教师们望而生畏。因此,采用传统数据收集模式,LOA 很难展开。目前解决这个问题的主要方法是利用教学技术,如雨课堂、微助教、blackboard 或其他学习管理系统等。这些教学技术中的学习管理功能可以大大减轻教师们收集资料和分析资料的工作负担,从而为推广 LOA 创造了有利条件。

第六个问题是课堂 LOA 需要学校支持系统的帮助,包括学校认可教师们的努力,并给予学校的认可、鼓励和表彰,以及学校为评价工作提供必要的经费、技术环境和物质条件等。学校还应建立专门机构系统收集评价信息。这些数据信息既可为学校教学工作提供决策信息,又可为未来的学校评估和专业评估做好准备。

最后介绍一些重要文献和网站供参考。美国关于课程教学评价的文章和书籍很多,但有三本书值得关注。一是 1993 年出版的《课堂评价技术》(第二版),一是 2016 年出版的《学习评价技术》。^[3,4]这两本都是工具书,特别适合教师们参考。另一本是 2017 年出版的《高等教育测量、评价评估手册》(第二版)^[5],这本书全面总结了该领域的进展情况,适合研究人员参考。关于美国大学的 LOA 实践,建议参考卡内基梅隆大学、北科罗拉多大学、明尼苏达大学、麦吉尔大学等校的教学支持中心网站,这类网站通常都有较丰富的 LOA 方法介绍。希望了解美国整体情况的研究者,建议上美国全国学习效果评价研究所(NI-LOA)网站,该所是美国研究大学学习效果评价的中心,创办于 2009 年。十年来该中心经常发布有关 LOA 的专题研究报告、调查报告和实践研究报告,是了解美国在该领域发展的最佳窗口。下面介绍的情况主要来自这些资源。下面介绍美国学者和教师们在 SC 改革中创造的五类 LOA 方法。

二、形成性评价

形成性评价(formative assessment, FA)是指在教学过程中,教师为了解学生学习情况而进行的学习效果评价。目的是让教师随时了解学生学习情况,及时给学生反馈或做出教学调整,以促使学生更好地学习。这个描述中有三个关键词:①在教学过程中进行;②为了反馈学生和调整教学;③为了促进学生更好地学习。

而传统教学评价的目的是检测学生学习效果,并据此给学生一个成绩。也就是说,教学评价的目的是学生学习成绩认定。因此传统大学教学评价主要是期中和期末考试,并给予分数,并不包括形成性评价的三个特点。以此观之,传统教学评价和形成性评价存在根本区别。

直到今天,教师仍需给学生一个学习成绩,作为对其学习过程和学习结果的评价。为此而进行的 LOA,被称为终结性评价(summative assessment, SA)。也就是说,凡是为最终评价而进行的 LOA,包括小考、中考和期末考,都属于终结性评价。

但是,如果教师根据 SA 反映出来的问题,反馈学生并调整教学,那 SA 就有了 FA 的性质。^⑤这意味着,在 FA 思想启发下,教师可以把 SA 变成 FA。这是以前 LOA 所没有的。也就是说,FA 的核心思想是以 LOA 来促进学生学习,这才是当代教学评价的核心理念。正是这一点,使得 FA 成了 SC 教学改革的主导要素之一。

因此,形成性评价是大学教学评价领域中的一场革命。它改变了对教学评价的传统认识。下面简要介绍这段历史并讨论其对当代大学教学评价的影响。

美国当代 SC 改革深受普通教育领域教学研究的影响,很多理论和方法首先是在普通教育领域中提出来,后来才扩散到高等教育领域中去的,FA 就是一例。

FA 这个概念最初是由美国学者斯克瑞文(Michael Scriven)于 1967 年首次提出来的,为了界定 FA 的性质,他区分了 FA 和 SA。^[6]这个思想首先启发了布鲁姆。布鲁姆提出,每个学习单元之后,教师都应当检查学生学习效果,确信学生已经掌握该单元知识后,才开始下个单元的教学。这就是布鲁姆的“掌握学习”(mastery learning)的思想。

布鲁姆之后,有人不断发掘 FA 概念的潜力。

如果评价不为打分,那 FA 的目的是什么呢?是为了收集学生学习状态数据,以便反馈教学。引进“反馈”这个概念,表明控制论思想进入教学评价研究。控制论的基本思想是,对任何一个控制系统,都可以通过采集系统状态信息来揭示系统状态与目标之间的差距,以便采取调整措施,确保系统保持在预定的目标状态上,从而确保系统稳定和目标的实现。控制者通过采集信息了解系统状态与目标之间的差异的活动叫做反馈。反馈的目的是让控制者可以根据反馈信息采取措施,确保目标的实现。在控制论框架下,采集学生学习状态数据的目的是为了了解当前教学状态,以便反馈教学,保证实现教学目标。从这个视角看,LOA 就是为了反馈教与学,以达到师生主动调整教学和学习的目的。因此,反馈和调整这两个术语的出现,意味着 LOA 变成了师生主动控制教学过程的关键环节,从而赋予 LOA 新的含义:评价不是为了打分,而是为了主动控制教学。这个思想是革命性的。

就像瞄准打靶一样,眼睛要随时观察准心和靶心的差距,并不断调整呼吸和动作,以保证击中靶心。因此,反馈与调整过程必须是连续的,不能只是有小考、中考和大考。信息采集的方式也不能只是考试,而是要针对所有可能影响目标实现的因素,采取多种方式采集信息。与此类似,做教学评价时,要注意所有与学习有关的要素,如学生已有知识准备、学习兴趣、身体状况、情感情绪等,以便采取适当措施,改进教学。同时还要把采集到的信息告诉学生,以帮助他们主动调整学习状态。这样一来,LOA 就变成了教学活动的一部分,而不再只是考试,也不再仅是评价学习,而是为了促进学习。这些思想都极大丰富了人们对教学评价的认识。

根据控制论的思想,FA 包含四个步骤:① 确立评价学生当前表现和预期成就的标准;② 收集信息,了解学生当前学习状态;③ 找到缩小学生当前表现和预期目标之间差距的策略;④ 实施策略缩小差距。

随后人们又根据建构主义心理学,提出学习是学生在自己头脑中建构认知框架的过程,教师的责任则是设计、引导和管理这个过程。根据社会文化建构主义,认知建构是在课堂这个特定社会环境中,是由学生和教师共同完成的。因此课堂氛围、学习目标和期望确定、学习过程等都应当

让学生参与,以提高学生主动参与学习。凡此种,都构成了积极学习理论的基础。人们还认识到,只有当学生处于积极学习状态时,学习才能从“学了就忘”的浅层学习变成“终身难忘”的深度学习。

2005 年奥拉尔和洛佩兹比较了布鲁姆的 FA 与内涵扩大的 FA 概念(见表 1)。^[7]这个比较显示,布鲁姆之后,把学习评价融入整个教学过程和让学生积极参与学习评价是 FA 的两个主要变化。

表 1 布鲁姆的 FA 与扩大的 FA 概念比较

布鲁姆的 FA 概念	内涵扩大后的 FA 概念
在每个教学单元之后插入 FA	把 FA 融入所有学习情境
反馈——纠正——补课	使用多样化方法收集数据
FA 过程由教师掌握	反馈+教学调整——形成规则
学生务必掌握所有学习目标	让学生积极参与 FA
补课使被评估的学生受益	FA 既帮助当前的学生,也帮助未来的学生

2012 年劳尼(Janet Looney)又根据后来的发展,总结了当代 FA 的六个特征:① 建立鼓励互动和使用评价工具的课堂文化;② 建立学习目标,跟踪每个学生的学习进步;③ 使用多样化教学方法来满足学生多样化学习的需求;④ 使用多样化方法检测学生进步情况;⑤ 针对每个学生的学习情况提供针对性反馈,并做相应的教学调整;⑥ 在教学中引导学生积极投入学习。^[8]劳尼的总结显示,除了鼓励学生积极参与、将评价融入整个教学过程之外,加强互动与个性化也是当前 FA 的两个新特征。

这些发展表明,就教学评价来说,FA 更像一种思想而不仅是一种方法,它对当代大学教学评价影响深远。读者还可以根据中国大学的实际和自己的教学体会,从 FA 中发掘出更多意义和方法。因此特别希望读者能从思想这个角度看待 FA 及其对美国大学 LOA 方法与实践的影响,而不是仅仅把 FA 当成一种方法。

三、评价量表(Rubric)

评价量表是美国大学的一个创造,也是运用最广泛的 LOA 工具之一,非常值得注意。关于评价量表的介绍和研究文献很多,很多大学教师发展中心网站上都有,但国内有关介绍和研究并不多。目前比较好的一本介绍性著作是华南理工大学出版社翻译出版的《评价量表:快捷有效的教学评价工具》(第二版)。^[9]此书可为入门书。评价

量表是一种包含评价维度(criteria)、评价等级(grade)和评价标准(standard)的表格。表2是美国一个小型文理学院制定的课程研究论文评价量表。

从表2可以看出,表的第一列是评价维度,表的第一行是评价等级,两者交汇栏中的内容是评价标准。以表格方式呈现评价的维度、等级和标准,这就是评价量表。显然,评价量表是一种可以用于几乎所有课程的评价方法,尤其适合评价主观性较强、且多人参与评价的课程。评价量表可以大大提高这类评价的客观性、统一性和公平性。

其实,评价量表倒也不算新鲜。凡参加过高考改卷的教师都应该见过类似评价量表。高考改卷之所以要制定细致统一的评价量表,目的就是为了保证评分的统一性、客观性和公平性。但美国大学对评价量表的使用颇有创意,常常让人耳目一新。

例如,教师可以在课程开始之初就把制定好

的课程评价量表交给学生,和学生详细讨论评价量表,仔细解释这些评价标准的含义,直到所有学生充分了解为止。如果学生在课程开始前就清晰了解课程的教学目标和评价标准以及教师的期望,他们就会更主动地调整自己的学习,以满足教师的期望。也就是说,课前和学生清晰交流课程教学目标和评价标准,可以促使学生主动学习,减少学生学习中的盲目性。就此而言,评价量表可能是目前所知的最好方法之一。

不仅如此,教师还可以鼓励学生参与评价量表制定。让学生提出他们希望把什么作为课程目标、如何考核以及采用什么考核标准。围绕评价量表的构建,和学生讨论课程的教学目标、学习方式和评价标准。通过讨论,把教师的目标和标准变成学生的目标和标准。这些做法被认为可以更好地调动学生学习的主动性和积极性。在不影响公平的前提下,教师甚至可以允许个别学生制定自己的课程目标、考核方式和评价标准,以满足其特定学习需要。这样一来,评价方法就成了激励

表2 评价量表

	新手	学徒	熟手	专家
知识整合	文章不能显示作者充分理解并能应用课程所学概念。	文章显示作者在一定程度上理解并能应用课程所学概念。	文章显示作者在大多数情况下能理解和应用课程所学概念,但论文有些结论没有得到充分论证。	文章显示作者充分理解并能应用课程所学概念。这些概念被很好地融入作者自己的见解中。结论能显示作者拥有分析和综合这些概念的能力。
主题聚焦	论题没有被很好定义。	就作业要求来说,论题太宽。	论题有所聚焦但缺少方向。论文应该聚焦于一个特别主题,但作者没有确立自己的立场。	根据作业要求,论文论题能很好地聚焦,论题从立场陈述和初始假设两方面为论文提供了明确的方向。
讨论深度	论文结论部分讨论草率,只有少数部分有简要的讨论	讨论各部分深浅不一。过度引用他人观点压倒了自己的看法。	论文讨论有一定深度,大部分论述也比较充分。	论文各部分均有深度讨论和充分论述。
一致性	所有信息均未联系起来,论文不成整体。各部分标题应该联系主概念,文章显示作者没有理解各部分概念之间的关系。	论文偶尔能将各种资源的信息联系起来,但论文明显存在断裂,缺少流畅性。显示作者并没有充分理解来自各种资源的材料之间的关系。	论文能把各种资源的信息联系起来。论述基本流畅,偶有断裂。显示作者对来自各种资源的材料之间的关系有一定的理解。	来自各种资源的信息被很好地联系起来。论述流畅。显示作者对来自各种资源的材料之间的关系有很好的理解。
拼写和语法	太多的拼写与语法错误。	有明显的拼写和语法错误。	偶尔有拼写和语法错误。	没有拼写和语法错误。
资料来源	参考文献不到5种,经同行审查的文献不到2种。引用了不可靠的网络文献,引用的文献不是当前文献。	引用的当前文献不到5种,同行评审的文献不到2种。引用了不可靠的网络文献。	引用了五种当前文献,至少2种以上经同行评议。所有引用的网络文献都是可靠的。	引用了五种以上的当前文献至少3种以上是经同行评议的文献。所有引用的网络文献都是可靠的和有权威性的。
资料引用	没有给出出处。	有些数据给出出处。引用方式要么不一致,要么不规范。	大多数数据引用给出出处。文字和文献引用的方式(APA)统一并且规范。	所有数据引用都给出了出处。文字和文献引用的方式统一并且规范。

资料来源:康奈尔学院(Cornell College)网站。®

学生学习的工具。

又如让学生评阅。由于有了评价量表,就可以让学生自己进行评阅。主要形式是先互评而再自评。经验表明,人发现别人的问题容易而发现自己的问题难。学生互评可以加深学生对课程内容的理解,还可以发展他们的审辨思维能力(critical thinking)。通过互评,学生可以更好理解课程的目标和标准,思索自己作业中的问题和原因,从而更好地完成自己的学习和作业。

学生评阅容易引起的一个疑问是,这会不会影响成绩评价的准确性和公平性?研究表明,如果评价量表制定得足够具体和清晰,学生互评的结果与教师评价的结果之间不会有显著差异。对成人学生而言,学生自评甚至比教师评价更严。^[10]通过学生互评培养学生审辨思维能力,这确实是评价量表方法的独特贡献。此外,对主要依靠助教改作业的大班教学,评价量表还有助于保持评分的统一性和公平性。

三是评价量表的多样性。目前美国大学中,评价量表几乎被使用到所有领域,不仅评价课程、论文、项目,还用于评价实践活动如表演、团队活动等。不仅可以评价整个团队,还可以评价每个成员。不仅可以评价学生表现与成绩,还可以用于评价教师的教学工作如教学大纲、课程计划、教学活动等。总之,评价量表几乎被用在了一切需要评价和可以评价的地方。这就导致了美国大学中的奇特景象之一,各种各样的评价量表有如雨后春笋,到处冒尖。

这种多样性也引起了一个严重问题,即评价量表的可靠性。为了解决这个问题,美国出现了评价量表制定专门化倾向,即由专门组织请专家来制定专门化评价量表,并进行有效性测验。其他人使用这些专门化量表可根据自己的需要略加改进,以保证评价量表的可靠性。例如美国高校联合会(AAC&U)组织专家就16种通识能力制定了16个专门化评价量表,包括研究与分析、审辨思维、创造性思维、写作、口头交流、阅读、定量素质、信息素养、团队合作、问题解决、公民参与、跨文化知识与能力、伦理推理、终身学习基础、全球化学习、整合性学习等^⑦。专业的课程公司如皮尔森公司(Pearson)也为自己出版的教科书制定了专门的评价量表。医学、工程等专业领域的课程教学专家也联合起来,发展专门化评价量表。甚至专门化的评价量表制作公司也应运而生。因

此,今天教师们考虑制作评价量表时,不必从头开始,可先到网上搜寻一下,看有没有现成的或成熟的量表,然后拿来改造使用。这样可以节约很多时间,也会比较可靠。

总之,评价量表是一种非常有效的LOA工具,值得中国大学学习并采用。

四、课堂评价方法

课堂教学(in-classroom)和非课堂教学(out-classroom)是大学教学两种基本形式。顾名思义,课堂教学是指以教室为基本教学场所的教学模式,主要以讲演和讨论为主要教学形式,偶尔也可插入一些室外活动。非课堂教学模式则指把学生带入社会,在真实场景中进行教学,故被称为真实性学习(authentic learning)。本节先介绍课堂评价方法,下一节介绍真实性学习评价方法。关于课堂教学评价方法(in-classroom assessment techniques, CAT),美国有两本比较好的参考书。一本是安吉罗(Thomas Angelo)和克鲁斯(Patricia Cross)编撰的《课堂评价技术》(第二版,1993)^[11],一本是巴克利(Elizabeth Berkeley)和马约尔(Claire Major)编撰的《学习评价技术》(2016)^[12]。这两本书都是工具书,系统总结了美国大学课堂教学评价的实践和方法,而且前后继承,后者基本继承前者的研究成果而且还有所发展。本节简要介绍这两本书的思路和方法。希望深究者建议阅读原书。

此外,美国各高校教学支持中心网站也有很多关于CAT的资料。这些网站上介绍的CAT是为本校教学教师服务的,因此通常具有实用性,而且与学校专业性质有关。因此读者可以在了解通用CAT的基础上,根据自己的专业和课程需要,到这些网站寻找适合自己需要的CAT。

安吉罗和克鲁斯在《课堂评价技术》序言中说,他们的研究始于1986年的课堂教学研究项目,1988年完成该书首版并大获好评,但也收到很多意见和建议,于是他们进一步修改,于1993年出了第二版,此书遂成该领域的经典著作。

该书一共描述了50个CAT方法,并指出这些方法基于七个假设:① 学生学习质量(即使不是完全)与教师教学质量直接相关,因此提高学生质量的途径之一是提高教师教学质量;② 要改进教学有效性,教师首先要制定详尽的教学目标,并能在教学中获得全面的反馈,以显示教学在多大程度上实现了这些目标。③ 要改进学习,学生需

要从教师那里及时获得适当而准确的反馈,同时学生自己也要研究如何评价自己的学习。④最能改进教学的评价是由教师自己回答关于自己教学的问题。⑤系统研究和智力挑战是激发教师动力、促使教师成长的源泉。课堂评价可以提供这种挑战。⑥课堂评价不需要专门训练,所有专注教学的专业教师都可以做到。⑦与同事合作,积极参与课堂教学评价,教师能提高对自己教学的满意度。

该书一个主要贡献是提出了应根据教学目标的性质选择评价方法的思路。由于教学评价的目的是检查教学目标的实现程度,因此评价方法的采用取决于教学目标的性质。这是课堂教学评价的一个重要思路。

该书把教学目标按性质分为6类:高阶思考类、基本学术技能类、专业知识技能类、通识教育类、工作职业相关类、个人发展类。接着列举了各类目标应采用的评价方法。该书作者还提出,为了促使学生全面发展,课程应尽可能采用更多种类的教学目标,而不要仅仅局限于课程本身的目标。例如专业课不仅应该考虑专业教育的目标,还应当考虑通识教育的目标,甚至个人发展的目标等。同理,通识教育课程也应考虑专业教育目标。通过这种方式,打通专业教育和通识教育,给学生更系统全面的训练。

这个“教学目标性质决定评价方法”的思路,启发了巴克利和和马约尔。2016年他们编撰《学习评价技术》一书时沿用了这个思路。但他们没有采用安吉罗和克鲁斯的教学目标分类系统,而是采用了著名课程设计专家芬克(Dee Fink)的教学目标分类系统。

芬克是美国著名大学课程设计专家,俄克拉荷马大学教授。他继布鲁姆课程设计模型后,提出了一个改进版的课程设计模型,他称之为“整合性课程设计模型”。⑧他的“整合性课程模型”是在布鲁姆课程设计三角形基础上,增加深度学习课程目标、积极学习教学活动和教育性评价理论(见图1)。芬克模型代表着布鲁姆模型之后的一个新综合和新发展,非常值得注意。

芬克注意到,有的课程教学能引起“深度学习”(即芬克的“显著学习经验”),有的只能停留在浅层学习层面。芬克研究了能引起深度学习教学的特点,他提出深层学习的教学目标有六个维度:知识基础、知识应用、知识整合、人文、关心和学会

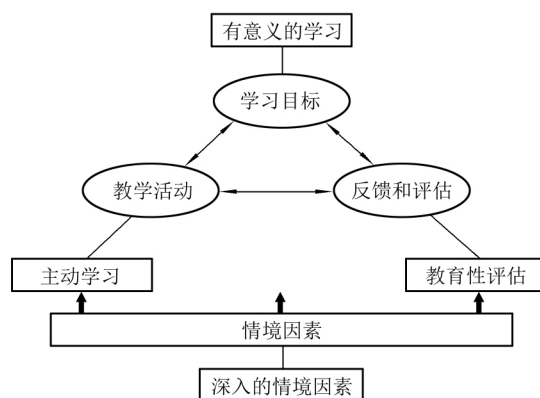


图1 整合性课程设计模型^[13]

学习。知识基础指课程需要掌握的知识;知识应用指应用知识的能力;知识整合指联系其他学科知识的能力;人文指通过学习对自己和对他人形成新认识;关心指在学习中发展出对课程领域的新兴趣和新爱好;学会学习指学习中发展出今后能自主学习该学科的能力。如果一个课程教学能达到这些目标,这显然就属于深度学习,或芬克的“显著意义的学习”。比较布鲁姆和芬克的学习目标分类系统会发现,芬克实际上是把学生在认知、情感、态度等方面的发展目标都整合进去了,而在布鲁姆的教育目标分类中三者是分离的。

当然,芬克也认为,并非所有课程都能达到这些目标,但他认为,课程设计时应尽可能多地包括这些维度。而且涉及维度越多,学习会越深入,效果也越显著。因此他提出了一个包括六个维度的学习目标体系(见图2)。巴克利和和马约尔采用的正是这个学习目标体系。

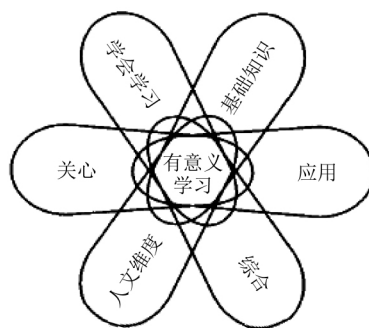


图2 六类学习目标及其互动^[14]

他们为六类学习目标推荐了50个评价方法:知识基础类CAT(10个):旨在检测学生的知识准备及课程知识技能掌握情况。

1. 首日终考:上课第一天给学生一个与期终

考试类似的考试。这样可以和期末考试比较,看学生最终学到多少。

2. 检查背景知识:编制一个问卷让学生填写,检查学生已有基础知识和预备知识的情况。

3. 进出场检查:用三至五个简要问题来检查学生课前准备和课堂学习情况。

4. 总结填空:编写一个课程内容总结,但在关节点上留空白让学生填写,以检查学生掌握情况。

5. 综述要点:在一个单元课程结束之后,让学生简要综述该单元知识要点。

6. 快速写作:老师提出的一个或一组问题,要求学生当堂快速作答。

7. 最佳总结:在一个单元课程结束时让学生写一个简要总结。

8. 快速检查:出多选题让所有学生作答,检查情况后让学生分组讨论他们的答案。

9. 团队测验:让学生以团队方式完成测验。可先个人作答,然后分组讨论,提出小组最佳答案,以促进相互学习。

10. 团队比赛:学生以团队方式参加考试或项目竞赛。

知识应用类 CAT(10个):旨在检测学生思维的批判性、创造性和问题解决能力。

1. 预测指南:在一项学习活动开始前,先让学生预测活动的过程和结果。在活动结束之后,比较实际情况和学生预测。

2. 事实与观点:让学生阅读若干文献,鉴别文献中的哪些是事实,哪些是观点。

3. 文摘评论:从要求阅读的文献中摘出某些部分让学生做评论。

4. 见解/资源/应用:给学生一个作业,要求说明个人见解、观点资料来源、以及可能的应用。

5. 指定主题:给学生一个已经学过的概念或理论,要求学生设想新的应用领域。

6. 问题分析:给学生若干问题情景,要求学生识别问题的类型和性质。

7. 问题听与说:学生两人一组解一个问题。让一个学生讲问题解决思路,另一个倾听,注意思路、步骤和逻辑。然后交换角色。

8. 同伴审查:给每个学生一个问题。做完与身边同学交换过程和答案,彼此相互检查分析过程和答案。

9. 三步法:制定方案、寻找资源、得到最佳效

果是真实问题解决过程的三个步骤。给学生一个问题情景,让他们练习使用三步法。

10. 数字项目:学生一个项目,让他们用数字技术方式呈现。

整合能力 CAT(9个):旨在检测学生联系其他学科知识的能力。

1. 知识网格:给学生一个空格的知识框图或思维导图,让学生填写所有空格。

2. 顺序链:给学生一个事件、活动或决策过程,让学生按活动顺序以步骤图的方式表现出来。

3. 概念图:让学生用思维导图的方式表现出所学课程中所有概念之间的关系。

4. 问题日记:以当代问题为题,让学生结合学习材料,以日记形式,记录事件发展、自己的看法和思考。

5. 二元文章:学生两人一组,读同一篇文章,然后各自提出一个论文主题和论文写作大纲。然后相互交换,比较彼此提出的问题与写作大纲之异同。最后用评价量表来评价彼此的工作。

6. 综合文章:让学生就读过的若干文献,选一个主题,写一篇正式的综述文章。

7. 案例研究:给学生一个真实案例,让学生结合所学知识,撰写一个包括背景、问题、挑战和解决方案的案例研究。

8. 班级文集:让每个学生就课程主题写一篇代表自己最高水平的论文,然后结集发表。

9. 电子档案袋:让学生把自己在课程学习期间提交的所有作业,连同教师对这些作业价值和意义的评价,做成个人电子档案袋。

人文维度(7个):旨在检测学生在学习中对自己和对他人形成新认识的情况。

1. 自由讨论:学生小组就课程相关问题自由交换思想,从准备、倾听、表达、深度四个方面评价他们有效参与讨论的程度。

2. 提名:就本专业领域的某个重要奖项,让学生通过集体研究提出一个候选人,并撰写建议获奖提案,包括其个人成就及支持获奖的理由。

3. 编辑评论:找一篇未发表的文章,让学生以刊物编辑身份给作者写信,告诉作者文章优缺点以及发表与否的决定及理由。

4. 戏剧性对话:设想两个人物(历史、现实、未来均可),让学生就某个问题,创作一个二人对话。

5. 角色扮演:让一组学生在教师指导下通过

独立学习,创造一个戏剧性问题情景,由学生扮演所有角色,展现这个情景。

6. 伦理困境:就一个专业伦理困境,让学生做出选择,并写一篇论文,说明自己决定的理由和思考过程。

7. 数字故事:用数字技术方式,针对特定观众,讲一个个人生活或学术经验的故事。

关心维度(7个):旨在检测学生学习中发展新关切、新兴趣、新爱好的情况。

1. 立场选择:就一个有争议问题,先让学生阅读各方观点文献,再让所有学生独立选择立场,然后按立场分组辩论,各组说明自己的立场和理由。辩论过程中允许个人立场转变,重新选组。

2. 三分钟表达:模仿三分钟学术竞赛。让学生用三分钟和一张PPT,说明自己的立场和理由。其他学生则用讲演评价量表做评价。

3. 公益广告:结合课程内容,让学生就一个相关公益问题,制作一个有说服力的公益广告。

4. 公共讲演:让学生就本地区的一个公共问题,通过调查研究,提出解决方案,并做一个公开讲演,说服当地人重视并解决这个问题。

5. 编辑回信:在教师指导下学生以刊物编辑身份,就文章录用与否给作者写回信。

6. 辩论:就专业领域内有争议的问题组织学生辩论,其他学生用评价量表就各方表现做评价。

7. 简要文章:学生以个人或小组形式,就一个当前问题进行独立研究,综述主要问题,提出解决问题的行动方案。其他学生用评价量表评价。

学会学习(7个):旨在检查学生发展出为今后深入学习该主题的能力。

1. 学习大纲:在教师指导下,学生撰写一份指导他人学习本课程的学习指导大纲,包括课程内容综述、学习要点和诀窍、考试准备等。

2. 制作评价量表:在教师指导下,学生就本课程的一个主要作业制定评价量表。

3. 制定考试试卷:在教师指导下,学生就某一个单元的学习内容,出一份试卷。

4. 罗列学习目标:在教师指导下,学生就某一个单元或主要学习活动,列举学习目标,并评估各目标的学习难度。

5. 课程学习日记:让学生在课程学习过程中,以日记形式撰写自己的学习活动和经验体会,反思自己的学习过程、主要挑战及思考等。教师规定日记要包括的内容和方面。

6. 掌握情况核查:对一项多阶段多步骤的学习任务,制定一个甘特图,反映各阶段各步骤要掌握的知识与技能,伴随学习过程,学生自己根据甘特图,检查自己学习和掌握的情况。

7. 个人学习环境:让学生用关系图方式表现自己在学习时所需各类资源和帮助的来源。

以上是《课堂评价技术》和《学习评价技术》这两本书的简介。作为工作手册,这两本书对每种方法的含义、目的、考察重点、使用方式、工作流程、各种变形、结果评价、注意事项等都做了具体说明。想采用这些方法的教师可参考原著。这两本书都是美国CAT领域中的里程碑式著作,集中体现了30年来来美国大学在CAT领域的主要成果和成就。

我认为美国的实践有三点值得注意:一是对课程教学目标加以分类,提出深度学习与各类目标整合程度密切相关;二是提出评价方法选择与课程教学目标性质相关,不同类型的教学目标采用不同类型的评价方法;三是LOA的教育性,学习效果评价应该有利于促进学生学习。

五、真实性学习及效果评价

真实性学习(authentic learning)是一个很有启示性的学习理念和教学理念。尽管中美两国都有很多此类实践,但作为一类学习理念,还没有得到很好的重视和研究。从文献看,澳大利亚和欧盟比较重视真实性学习,中美的相关文献都不多^⑥。下面会看到,真实性学习对大学教学有重大价值,尤其是对应用型 and 职业型高校。真实性学习作为一种教学理念,提出了一系列与传统课堂教学模式不同的思想,在学习任务、学习环境、教师角色、学习目标、学习过程、效果评价等环节有自己的思考和特色,值得注意。本节将简要介绍真实性学习概念以及相应教学模式,但重点放在学习效果评价上。

伦巴第(Diana Lombardi)在为OECD撰写的研究报告《为了21世纪真实性学习:综述》^[15]中提出,真实性学习是在真实或拟真环境中、以真实问题为任务,在教师示范和帮助下,学生通过真实参与真实问题的解决过程,学习相关知识和技能,获得真实学习体验的学习过程。

伦巴第认为真实性学习有10个特征:①和真实世界相关。真实性学习的学习任务必须是和真实的专业工作相关,甚至就是真实任务。②要解决的问题和真实生活中的问题类似,是非良构

的,而经老师改造过的问题是良构问题。非良构问题没有提供足够解决问题的信息和思路,甚至可有多种解读,需要学生自己去思考并寻找关键信息。③这类问题不是即刻可以解决的,需要通过一定的调查研究。④解决问题需要多种角度思考,而且需要多种资源。⑤解决问题需要合作,与同学及相关工作人员合作。⑥问题解决需要多次决策,需要经常反思自己的工作思路。⑦问题涉及多个学科知识,具有跨学科特点。⑧与对真实工作成果要求一样,最终结果需要抛光打磨,直至完善。⑨工作成果要从多方面多角度评估。⑩学习效果需要整合式评价(integrated assessment)。显然,真实性学习的目的是让学生学到关于真实世界和真实生活的真知识和真本事。因此真实性学习有别于传统大学课堂教学,而更加接近现实生活中与职场工作相关的专业性学习。由此可见,真实性学习特别适合应用型和职业型高校的教学。显然,倡导真实性学习的主要原因是,人们发现传统大学教学模式培养出来的学生对真实生活的认识和就业能力等方面存在不足,因此有必要以真实性学习方式来缩小学校和现实生活之间的距离。

“真实性学习”的倡导者们认为,传统教学研究有三个缺陷,一是过于强调认知而忽视了情感、价值、态度等因素对学习行为的影响,因此主张寻求能激发学生全身心投入学习的教学模式。真实性学习就被认为是这样一种方式。二是忽视学习环境对学习的贡献。真实性学习倡导者认为,真实场景和真实任务对学生的情绪、价值、态度等因素有重大影响,有助于激发学生学习动机,改善学习主动性、提高学习投入。三是忽视学习的建构性。知识不是被老师灌到学生头脑中的,而是学生在真实情景中和与他人的互动中构建的。真实性活动和实践直接塑造学生的知识构建过程。

维拉多尔(Villaruel)等人在文献研究中发现,有研究表明,真实性学习可以提高学生学习的自主性、对学习的责任感和主动性、自律能力、元认知能力、自我反思能力,尤其是处理不确定性和工作压力的能力、审辩性思维和创造性思维、战略性思考和规划能力、与他人合作/互动/交流能力,对与专业相关的学习能力等方面也有提高和促进作用。^[16]

真实性学习设计要注意五个关键要素:环境、目标、任务、活动、评价。学习环境必须是真实和

近似真实的。这里的环境包括物理环境、社会环境、文化环境等,最好就是真实或拟真职场环境。教育目标除了专业学习目标外,还应包括认知发展目标如思考、创造、解决问题能力;情感发展目标如审美、情绪、价值观发展、自我认知等;意志发展目标如决策、行动、承诺、处理不确定性能力等,以及其他心理发展目标如感知、感受、身体动员能力等。也就是说,包括芬克提出的六类学习目标。因此,真实性学习任务应该具有伦巴第归纳的10个特征。真实性学习的活动主要应由学生自己计划和执行,教师只是帮助者和咨询者。

若此,如何评价学生的真实性学习成果就成了一个新挑战,由于和传统课堂教学评价方式有很大不同,以致效果评价成了真实性学习研究中的一个专门问题。文献中研究者们提出了很多真实性学习评价的方法,此处总结为六个点:①强调真实性;②注意整体性;③反思性自我评价为主;④个体评价为主;⑤重视形成性评价;⑥形式多样化。具体解释如下。

一是强调真实性。真实性学习的目的是让学生认识、体验、适应、并参与改变真实世界和真实生活,因此无论最终作业或作品是什么,它们都必须符合行业和专业标准。用行业和专业标准来评价学习最终结果。这些评价标准和要求要事先告诉学生,而且要不断提醒学生注意这些标准,以便学生在整个作业过程中可以根据这些标准调整自己的学习行为。

项目开始时最好邀请专业人士参与评价标准制定,并向学生解释这些标准及理由,务必让学生清楚了解这些标准及其意义。最后评价应邀请专业人士参加学习结果评审。

二是注意整体性(holistic)。真实性学习要能促进学生全面发展,要让学生获得一个有显著意义的学习体验,因此评价要注意评价的整体性。评价标准应包括学生学习的责任感和主动性、自律能力、元认知能力、自我反思能力,处理不确定性和工作压力能力、批判性思维、创造性思维、规划和思考能力、与他人合作/互动/交流能力,以及与专业相关的学习能力等。即包括学生在认知、情感、情绪、意志、责任、价值、自我认知等所有方面的改进与提高境况的评价。

三是反思性自我总结为主。进行整体性评价的一个有效方法是,事先根据任务性质,和学生一起制定整体性评价量表。然后让学生根据评价量

表以日记形式记录自己的提高和改进。最后给自己的工作和表现做出评价,说明理由,并撰写学习总结报告。这就是反思性自我总结。

四是个体评价为主。由于每个学生的项目可能不一样,在项目中担任的角色不同,每个学生的个人情况也不同,因此应根据每个人的发展和进步做出评价,即以个体评价为主。如果是以小组制作为最终考核项目,那么可以按统一标准评价作品,但个人反思报告则采用个体评价。

五是重视形成性评价。教师要积极参与整个过程,以便随时评价学生表现,及时反馈,以帮助学生顺利实现学习目标。也就是说,真实性学习过程中要高度重视形成性评价。

六是形式多样化。最终成果可以有多种形式,如项目作品或作业、学生学习总结报告、学习过程的各种中间产品如计划书、草图、样品、照片、录像。可以把所有这些作品汇集成为学生作品集(portfolio),连同教师的计划书、各项工作进展证据、工作记录、师生之间的反馈与交流记录、最后的作品展记录、录音和录像等。所有这些都是学生学习评价和教师教学评价的证据,也都是今后课程教学评估和专业评估的重要材料,要注意收集和保存。

总之,真实性学习评价和传统课堂教学评价有很大不同,需要进一步研究探索。

六、成果档案袋(achievement portfolio)

与形成性评价类似,成果档案袋既是一种评价思想,也是一种评价工具。说它是评价思想,是因为它体现了形成性评价和成长性评价的思想。说它是一种评价工具,是因为它可以记录学生的学习经历和成长过程。成果档案袋方法可以培养学生主动规划和反思自己学习的能力,可以向学生、家长、政府和社会展现学生的学习成就和教师的教学成果,还可以帮助教师、院系、学校发现教学中的问题。成果档案袋是当代大学教学成果评价的重要方法之一。信息化时代又出现了电子档案袋,大大推动了档案袋方法的使用和普及。目前电子档案袋已成了美国大学主要的学习效果评价工具之一。

成果档案袋是由学生收集最能反映自己学习成果的作品集。作品可以是文章、录音、录像、图纸、照片、制成品(artifacts)等,以及由学生撰写的反映自己学习心路历程的学习心得报告汇总,并配有适当文字说明。成果档案袋方法有四个要

点:① 学生自己收集作品,决定作品呈现方式,撰写学习心得报告,展现自己的学习成果和学习经历;② 只展现学生自己认为最好的作品,展示什么和如何展示,都由学生自己决定。因此反映的是学生自己对自己的看法;③ 教师可以提出建议,但原则上不加干预。成果档案袋可以以课程为单位,让学生随着自己的学习进展,逐渐使个人学习成果档案袋丰富起来。④ 档案袋要求公开展示,以便同学们互相参观来改进学习。目前美国大学流行的做法是,学生一入学就为他们建立个人学习网页,作为展示自己学习成果的平台。然后以课程为单位,记录自己的学习过程和成长经历。教师可以以课程为单位,要求所有学生制作课程成果档案袋。课程成果档案袋可以集中放在教师个人网页中,供未来学生参考。待教师评分完成之后,学生再分别整合到自己的个人网页中去。

成果档案袋的独特优点是学生提供了一个展现自己学习成果和反思自己学习经历的平台。通过收集和选择个人学习成果,学生可以不断明确自己的学习目标,增强学习责任感,反思学习过程,调整学习策略,提高元认知能力。通过不断更新个人学习成果网页,学生可以逐渐更好地认识自己,了解自己的学习兴趣、学习目标甚至个人未来发展目标。从教学评价和评估来看,成果档案袋提供了一个直接展现学生学习效果的平台,是教师了解学生学习情况的重要渠道,为教师研究和改进教学提供了重要证据。也为专业评估和院校评估提供了关于学生学习效果的重要资料。我认为,档案袋方法最宝贵的优点是,它提供了一个从学生角度看待学习效果和学习经验的视角,这与其他评价方法的教师角度不同,增强了教学评价客观化的基础。

成果档案袋方法最早受美术教育启发。美术院校毕业生通常会收集自己的最好作品,毕业时开个人作品展,以这种方法展现自己的学习成果和成就。这启发了教学评价人员,他们把这种方法推广到所有专业和课程教学中去,为教学评价评估获得效果证据、收集问题提供了一个解决方案。近年来教育技术发展大大降低了成果档案袋和个人网站的制作成本,于是成果档案袋方法在美国各大学很快流行起来。

成果档案袋方法的推广也引起了一些讨论,主要是两个问题:一是如何决定成果档案袋的内

容与结构。对此有研究建议,应把编制成果档案袋和课程设计结合起来,尤其是与教学目标和教学评价设计结合起来,使成果档案袋成为 LOA 的一部分。可以通过师生共同制定成果档案袋评价量表的方式,让学生清楚课程要求是什么,教师期望是什么,然后以此引导学生的成果档案袋制作。这种做法,实际上是把教学评价、评价量表和成果档案袋方法结合起来使用。教学评价和评价量表主要反映的是教师角度,而成果档案袋反映的是学生角度。从教师和学生两个角度反映学习效果,结果会比较客观。

二是成果档案袋方法的有效性和可靠性。这主要和成果档案袋的学生角度有关,即我们在多大程度可以根据学生的自我评价来判断真实的学习效果。其实,这和医生在多大程度上能只根据病人陈述来做诊断的问题一样。不必认为学生的自我感受就是客观的和可靠的,但它至少提供一个不同于教师的观察视角。就学习效果评价而言,没有学生视角不行,仅有学生视角也不行。恰恰是多视角观察才可能使评价更加客观可靠。这个问题比较复杂,此处不做深入讨论。

总之,在促进学生学习和学习效果评价来说,成果档案袋方法是一个好方法,建议中国大学尽快采用和推广。

七、实践与反思

之五和之六两篇都讨论美国 SC 改革中的教学评价问题。之五着重介绍美国大学的教学质量保障体系,指出美国大学教学质量保障体系是“一条主线、三级管理”。具体表现为一个使命(学校本科生教育使命)、三个矩阵(通识教育矩阵、专业教育矩阵、课程教学矩阵)和两个辅助机构(教学支持中心、院校研究所)。通过三个矩阵对标本科生教育使命来确保本科教育质量。再用矩阵方法保证所有课程的教学活动都对标本科教育教学总目标。教学支持中心负责教师教学能力培养与提高,院校研究所负责收集数据支持教育教学决策,由此形成一个完整的闭环质量管理体系。

之六则着重介绍了课程教学层面的学习效果评价的理念和方法,包括形成性评价、评价量表、课堂教学评价方法、真实性学习评价、成果档案袋方法等。这些反映了课程层面的学习效果评价。

这些介绍都是理想情况,实际情况如何并未讨论。故本节以美国全国学习效果评价研究所(NILOA)2018年的调查报告为基础,介绍美国

高校教学评价的实际情况,反思其经验教训,并给中国大学提出建议。

NILOA 成立于 2008 年,是美国最重要的学习效果评价研究机构。该所曾于 2009 年、2013 年和 2018 年组织过三次全国性问卷调查。主要调查美国高校教学质量保障和教学评价实际情况,调查对象是大学教务长(provost)。2018 年发表的调查报告的题目是《评价很重要:记录学生的真实学习实践趋势》。^[17]这次调查是 2017 年做的。共发问卷 2781 份,回收 811 份(29%)。其中公立高校为 56%,私立高校为 44%。80%的回复来自教务长,其余的来自学校教学评价负责机构。调查发现:

1) 多数学校做到了专业学习与学校教学使命对标。82%的学校清晰定义了本校的本科教学使命,并要求专业教学对标学校本科教学目标。学校所有专业都根据学校教育使命制定了本专业学习成果指标。只有 36%的博士学位授予大学完成了这项工作。反之,营利性高校全部完成了这项工作。

2) 评价方法多样化。美国高校使用多种方法评价学生的真实学习成果(authentic learning outcomes)。使用的方法包括 NSSE 调查(76%)、评价量表(71%)、课堂表现评价(64%)、毕业项目(61%)、校友调查(55%)、职业资格考试(52%)、雇主反馈(37%)、校外学习场合表现评价(36%)、AACU 通识教育测量(34%)、成果档案袋(32%)等。不同类型学校使用的方法组合不同。

3) 学校支持。学校对评价工作最有力的支持是:让更多教师参与(51%)、为教师员工提供培训机会(46%)、学校更好地评估教辅人员的能力(30%)。

4) 评价资料使用。评价所获资料和数据主要用于改进课程教学与专业教学,也被用于学校教学工作和其他工作改进、学校战略和政策制定、校外评估机构评估、向社会提供问责报告等。如何有效交流并用好评价数据和资料仍是一个巨大挑战。

5) 教育技术。包括两方面。一是教育技术。如何在教学评价和学习效果评价中充分发挥教育技术的巨大潜力仍有待改进。二是学习效果分析技术,如何利用学习管理系统及其数据分析功能来收集和分析评价数据资料,也是一个巨大挑战。

6) 组织与领导。领导是关键。学校董事会

在评价工作方面可以发挥重要作用。由教师领导的教学评价委员会和院校研究所可以为全校的效果评价工作提供必要支持。

7) 学校类型。规模越大、选择性越强的高校越不愿意采用多样化学习效果评价方法,其教师参与教学评价活动的积极性越低、越不愿意做全校性改变。为什么学校规模和选择性水平会导致不同的效果评价态度,原因尚不清楚。

8) 未来哪些方法会比较重要。课程矩阵(77%)、激励教师有效设计学生作业(62%)、发展促使学生完成学业的有效途径(56%)、改进通识教育(52%)、提高高影响力实践质量并扩大采用范围(50%)、采用 AAC&U 通识教育评价量表(44%)、AAC&U 关键学习效果(LEAP Essential Learning Outcomes)指标(28%)、能力发展为专业设计(23%)、发展综合性学习记录(23%)。

以上是 NILOA2018 年调查报告的主要内容。下面做一点分析:

1) 范围。调查发放问卷 2781 份,仅回收 811 份(29%)。由于学习效果评价是一个敏感问题,涉及学校声誉,因此做得好学校应该会比较积极回应,而做得不好的学校则可能采取回避策略。按这个假设,71%的未回复率就有点高。假设其中 30%学校本身做得较好,只是由于种种原因未能回复,70%的学校是因为自己做得不好因而不愿意参加调查。那就意味着在被调查的 2781 个学校中,大约有 50%的学校在有意回避调查,这是个令人担心的比例,即约有 50%的高校在学习效果评价方面做得不够好。所以,该报告以《评价很重要》为题发表是有原因的,它在呼吁更多高校重视学习效果评价工作。2018 年 NILOA 十年回顾报告中总结道,目前“美国高等教育还远不能提供足够的关于学习效果的证据,以指导学习和教学方法改进,从而提高学生和高校的表现”。^[18]这个结论也佐证了上述结论。

2) 研究性大学。在调查中发现,规模越大、选择性越强的学校越不愿意根据学校本科教育使命来要求专业教学。在美国这应该是指研究性大学,即被调查的研究性大学中仅有 36%的学校要求所有专业(program)对标本校本科教育目标。报告说其中原因有待调查。但我的观察是,这是因为研究性大学的教师研究任务较重,因此他们没有把主要精力放到本科教学上。相反,美国在

学习效果方面做得最好的是私立小型文理学院,尽管这些学校的选择性也比较高。因此,如何在研究与教学之间保持平衡,以确保教师用心从事本科教学,这才是美国研究性大学本科教学质量保障的严重问题。中国研究性大学显然有同样的问题。

3) 专业与使命对标。NILOA 把专业教学目标对标本校本科教育使命作为重要评价指标,这个做法值得注意。如果没有各专业教学的支撑,学校本科教学使命就是空的。而要做到这一点,NILOA 认为学校领导至关重要。只有学校领导负起责任,在学校战略、资源、教师绩效评估、教师发展、教学技术提供、教学环境营造、学校政策与管理等方面给予支持,专业与使命对标才能真正兑现,否则确保本科教学质量就是一句空话。

4) 多样化。NILOA 认为各校应该根据自己的情况来开展学习效果评价,不要“一刀切”。课程矩阵方法、评价学生真实学习效果、课程教学设计与作业设计结合、评价量表、创新课堂教学评价方法、真实性学习评价方法、高影响力教学法、电子档案袋方法等,都是被证明是被广泛采用的有效方法。中国大学教师也应该根据自己的实际情况,发展出适合自己的有效的学习效果评价方法。

5) 教育技术。大力倡导教学技术运用,加速学习管理系统、Blackboard、电子档案袋、雨课堂等教学技术的使用,可以有效减少教师在信息收集和分析方面的压力。

6) 教师员工培训。建立学校教学支持中心,为教职员工提供必要培训,为教师提供教学咨询,确保教职员工全面掌握 SC 理念和方法,对建立全校统一的教学质量文化十分重要。

7) 院校研究。建立院校研究办公室,系统收集全校教学效果评价信息,为学校和院系教学改革提供信息,为校内外各类教学评估做好准备。

以上是对美国大学学习效果评价工作的介绍和反思,供中国大学和教师们参考。

至此,SC 改革实践的三个领域——实践与方法、环境与技术、评价与评估——介绍完了,下面转向介绍美国 SC 改革的支持系统,包括研究与培训、组织与管理。这是随后两篇文章的主题。

注 释

① 感谢加州大学院校研究办公室主任常桐善博士,和他的讨论导致了这段文字。

② 就业起点工资指正式入职后的初始工资,不指试用期工资。

③ 对目前学习研究的简要总结参考 Philips, D. & Soltis, J. Perspectives on Learning. (5th edition), 2009, Teachers College Press, Columbia University。更全面深入的介绍参考 Seel, Norbert M. Encyclopedia of the Sciences of Learning[M]. Springer Science & Business Media, 2012。

④ 关于脑科学对学习和记忆的研究,参看 Gluck, Mark et al. Learning and Memory: from Brain to Behavior. 3rd edition, 2016. Worth Publishers。

⑤ 期终考试虽然不能帮助当前学生改进学习,但它可以帮助改进未来学生的学习,因此仍然具有改进学生学习的作用。

⑥ <https://www.cornellcollege.edu/LIBRARY/faculty/focusing-on-assignments/tools-for-assessment/research-paper-rubric.shtml>。2019年8月10日提取。

⑦ <https://www.aacu.org/value-rubrics>。2019年8月10日提取。

⑧ Dee Fink Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses. 2013. Jossey-Bass。该书初版于2003年,2013年又出了修改更新版(revised and updated edition)。浙江大学出版社曾翻译出版了该书第1版。本文主要参考修改更新版。

⑨ 例如 Villarroel 等人用 authentic assessment, authentic instruction, authentic intellectual work 三个关键词搜索 1988—2015 年的英文文献,仅得到 112 篇文献,其中只 40 篇是有关高等教育的,剩余部分为其他教育领域。Villarroel, V. et al (2018) "Authentic assessment: creating a blueprint for course design", Assessment & Evaluation in Higher Education Value 43. No: 5, pp. 840—854。我用 authentic assessment 和 higher education 为关键词搜索了 UCLA 图书馆的数据库中 2000—2019 年文献,仅得到 38 篇,而且主要是澳大利亚学者的作品。

参 考 文 献

[1] 赵炬明,高筱卉. 关注学习效果:建设全校统一的教学质量保障体系——美国“以学生为中心”的本科教学改革研究之五[J]. 高等工程教育研究,2019(3):5-20。

[2][5] SECOLSKY C E, DRNISON D B E. Handbook on Measurement, Assessment, and Evaluation in Higher Education[M]. Routledge, 2012。

[3][11] CROSS K P. Classroom Assessment Techniques. A Handbook for College Teachers[M]. Jossey-Bass, Incorporated, 1993。

[4][12] BARKLEY E F, MAJOR C H. Learning Assessment Techniques: A Handbook for College Faculty[M]. John Wiley & Sons, 2016。

[6][8] JANET LOONEY. Formative Assessment and Improving Learning. In Seel Norbert M. Encyclopedia of the Sciences of Learning[M]. Springer Science & Business Media, 2012: 1318-1320。

[7] ALLAL L, LOPEZ L M. Formative Assessment of Learning: A Review of Publications. In John-Brooks C S. Formative Assessment: Improving Learning in Secondary Classrooms[M]. OECD, 2005: 241-265。

[9] 史蒂文斯·利维. 评价量表:快捷有效的教学评价工具[M]. 陈定刚译. 广州:华南理工大学出版社, 2014。

[10] Sadler P M, Good E. The Impact of Self- and Peer- grading on Student Learning[J]. Educational Assessment, 2006, 11(1): 1-31。

[13][14] L·迪·芬克. 创造有意义的学习经历:综合性大学课程设计原则[M]. 杭州:浙江大学出版社, 2006,90,25。

[15] LOMBARDI M M. Authentic learning for the 21st century: An overview[EB/OL]. [2018-07-18]. http://www.lmi.ub.edu/cursos/s21/REPOSITORIO/documents/Lombardi_2007_Authentic_learning.pdf。

[16] VILLARROEL V, BLOXHANM S, BRUNA D, et al. Authentic Assessment: Creating a Blueprint for Course design [J]. Assessment & Evaluation in Higher Education, 2018, 43(5): 840-854。

[17] JANKOWSKI N A, TIMMER J D, KINZIE J, et al. Assessment that matters: Trending toward practices that document authentic student learning[EB/OL]. [2019-08-06]. <https://www.learningoutcomesassessment.org/documents/NILOA2018SurveyReport.pdf>。

[18] GEORGE D. KUH, STANLEY O. IKENBERRY. NILOA at Ten: A Retrospective[EB/OL]. [2019-08-22]. https://www.learningoutcomesassessment.org/wpcontent/uploads/2019/02/NILOA_at_Ten.pdf。

Align with Learning Outcomes: Highlights of Course-level Learning Outcome Assessments

—Studies of the SC Undergraduate Education Reform in the USA (6)

Zhao Juming

Abstract: This article introduces the major methods of course-level learning outcome assessment inno-

vated by the American universities and colleges in its SC reform, including formative assessment, rubrics, classroom learning outcome assessments, authentic learning assessments, achievement portfolios. Some issues regarding the concept and measurement of learning outcome assessment are also discussed in this article. Based on the NILOA 2018 survey report, the article drew a rough picture of the current practices of learning outcome assessment in American higher education, and offered some suggestions to Chinese higher education institutes.

Key words: Student-centered, American undergraduate education reform, course-level assessment, in-classroom learning outcome assessment, learning outcome assessment. (责任编辑 姜嘉乐)

· 书讯 ·

《新三中心译丛》推动以学生为中心的教育教学改革

由旅美学者、高等教育资深研究者赵炬明教授担任主编,湖南教育出版社历时四年时间精心打造的,全景式展现美国本科教学改革的“新三中心译丛”将于近期出版发行。本套丛书共分6册,分别为:《大学生发展:理论、研究和实践》《学习范式学院》《麦肯齐大学教学精要:大学教师的策略、研究和理论》《设计和实施以学习者为中心的课程体系》《领导以学生为中心的校园:改进学生学习成果的管理者框架》《大学教学学术再思考:学校整合及影响》。本套丛书从多个角度切入,系统介绍了美国本科教学改革,汇集了其最新成果,能够很好地帮助中国高校的管理者和改革者深刻理解美国本科教学改革。

解读美国本科教学改革的钥匙

20世纪80年代,本科生教育质量下降,社会各界对美国高校的本科教育批评不断。从20世纪90年代开始,“以学生为中心”的本科教学理念引起了很多高校的重视,几乎所有高校都成立了专门委员会以提高本科教学质量,最终推动了美国本科教学改革。这场改革极大地提升了美国高校的本科教学质量,为美国社会发展做出了贡献,也为其他国家本科教学改革树立了榜样。丛书展现了美国本科教学改革的本质与精髓,是中国读者理解美国本科教学改革的钥匙。

全景式展现美国本科教学改革

“新三中心译丛”介绍了美国本科教学改革在课程教学、专业课程体系、学校制度、学术研究与教师培训等方面所采取的具体举措,特别强调“以学生为中心”的教学改革是从思想到行为再到组织的从根本上转变传统大学教学模式的“范式变革”。丛书指出,高校用“以学生为中心”的理念设计和规划本科教学时,才能自上而下地推动本科教学改革的顺利进行。这种新思想和新实践将极大影响我国本科教学模式,提升本科教学质量。

推动中国以生为本的本科教学改革

借鉴和学习世界先进大学的教育教学模式,深化本科教学改革,自主探索、构建具有中国特色的本科教学模式成了我国大学改革发展的必然选择。“新三中心译丛”折射的美国本科教学改革的历史发展进程和当今状况,能为中国高校开展教学改革时遇到的一系列问题提供指导和借鉴,从而实现“他山之石,可以攻玉”的目标。