

美国教师培养项目增值评价探析^①

——以路易斯安那州为例

冯慧, 饶从满

(东北师范大学国际与比较教育研究所, 吉林长春 130024)

[摘要] 世纪之交以来, 美国教师培养面临内部质量提升和外部质量问责的双重压力。在此背景下, 伴随数据系统和分析方法的发展以及教师增值评价的推广, 教师培养项目的增值评价应运而生。教师培养项目增值评价基于教师培养项目→教师教学质量→学生学业成绩的因果逻辑, 通过统计计量模型来评估教师培养项目对学生成绩的贡献度。其中项目效能的证据选择、效能测量以及效能增值结果的意义阐释构成教师培养项目增值评价的关键内容。教师培养项目增值评价虽然在一定程度上响应了公众对于强化教师教育绩效责任的呼声, 但依然存在着信效度和适用范围等问题。

[关键词] 增值评价; 教师培养项目; 教师教育评价; 路易斯安那州; 美国

中图分类号: G659.10/712 文献标识码: A 文章编号: 1003-7667 (2018) 06-0059-10

教师培养成效评估作为判断教师培养机构是否有效地培养了合格教师的重要手段, 日益成为教育政策制定者和学术界关注的焦点。美国联邦政府对教师培养成效评估的兴趣, 从2008年再授权的《高等教育法》修订案(Higher Education Act)和2009年颁布的《力争上游计划》(Race to the Top)中有关教师培养的章程可见一斑。美国联邦教育部试图以立法和竞争性拨款的形式, 推动建构基于学生学习结果(student-outcome based)的教师培养问责体系。作为教师培养问责体系的重要组成部分, 教师培养项目增值评价基于教师培养项目毕业生^②所教学生的实际成绩与预测成绩之间的差异, 运用统计计量模型来评估教师培养项目对学生成绩的贡献度。目前, 美国已有

路易斯安那州、北卡罗莱纳州、田纳西州、俄亥俄州、得克萨斯州和佛罗里达州六个州将教师培养项目的增值评价纳入教师培养问责体系。

一、实施教师培养项目增值评价的背景

总体而言, 美国实施教师培养项目增值评价源于其教师教育领域存在的问题和教师教育评价改革的需求, 也得益于教育和教师评价领域的相关发展。

(一) 教师培养的内部质量提升和外部问责需求呼唤教师培养评价

近年来, 美国教师培养市场呈现多元化和分权式的特点: 教师培养主体涵盖大学、非营利性和营利性公司; 教师培养项目种类包括传统的大学培养项目和多种替代性培养项目。据

①本文系2016年教育部教师工作司委托项目“卓越中学教师培养计划项目评价指标体系与实施办法研究”的阶段性成果。

②教师培养项目的增值评价的项目毕业生指教龄为两至三年的初任教师。

作者简介: 冯慧, 女, 东北师范大学国际与比较教育研究所硕士研究生;

饶从满, 男, 东北师范大学国际与比较教育研究所教授, 博士生导师。

统计,每年入职的20万新教师中,有近70%~80%接受过传统教师培养项目的培养,而其余新教师均通过全美130多个替代性培养项目进入教师队伍。^[1]公众对于种类繁多的教师培养项目是否为基础教育输送了高质量教师持质疑态度,而决策者则持续质疑当下的教师教育能否回应社会需求,是否已经成为危如累卵的官僚主义体系。^[2]在全美教师质量委员会(National Council on Teacher Quality, NCTQ)2013年针对全国教师培养项目开展的质量评估中,获得4分的中学教师培养项目仅有4个,获得3分及以上的培养项目数量尚且不到10%。^①该委员会发布的评估报告也因此将美国的教师培养形容为“平庸的产业”(industry of mediocrity)。^[3]不仅如此,更有研究者认为当下美国的教师教育处于危险之中。^[4]

美国的教师培养不仅面临内部质量的提升需求,还承受着日益强化的外部问责要求。引发公众对教师和教师教育质量质疑的导火索是美国基础教育质量平庸甚至下降的现状,尤其是美国学生近年来连续在PISA中排名落后、SAT考试分数偏低等现象引发公众将矛头直指教师和教师教育。美国公共政策话语中的教育问责和绩效评估运动,发端于20世纪70年代的最低能力测试运动(minimum competency movement),发力于80年代的《国家处于危机之中》,并在21世纪初布什政府通过的《不让一个孩子掉队法案》中发展至高峰。随着奥巴马政府2009年出台竞争性拨款项目《力争上游计划》,强调教育问责和绩效评估继续成为美国教育政策的时代议题。在这场强调问责与绩效评估的大潮中,学生的学业成绩被视为教育绩效的重要表征,也成为教育问责关注的核心。2002年的《不让一个孩子掉队法案》和2008年

再授权的《高等教育法》修订案均提出以K-12学段学生的学业成绩为标准对教师培养项目领导者和教师教育者实施问责。^[5]

在面临内部质量提升和外部问责要求的双重压力下,各州必须向公众证明本州的教师培养项目能培养出可测量的高效能教师。^[6]加强对教师培养项目的评价势在必行。

(二) 日益强化的培养问责推动教师培养评价方式的改革

美国教师培养项目的传统评价方式以是否达到认证要求为主,即由全美教师教育认证委员会(National Council for Accreditation of Teacher Education, NCATE)^②针对不同教师培养要素制定最低要求作为认证标准,这些要素包括:教师候选人的知识、技能以及专业性向(professional disposition);评价体系和机构评价,实习经历和临床实践;多样性;教师教育者资质;机构管理和资源。^[7]传统的美国教师培养认证是一种过程导向(process-oriented)的教师培养评价,这种过程导向的评价方式难以直接证明教师候选人的课堂教学技能和促进学生学习的的能力,因而无法回应公众的问责。虽然决策者早在1998年再授权的《高等教育法》修订案第二编(Title II)中就意图通过加强教师培养项目评价来回应公众的问责^③,但直至2007年,尚有31个州没有一个教师培养项目被鉴定为薄弱项目。教师培养评估的“良好”结果与基础教育质量的“平庸”现状相差甚远,这引起了各利益相关群体对传统教师培养评价方式的质疑。在此背景下,2013年,教师培养认证委员会(the Council for the Accreditation of Educator Preparation, CAEP)颁布了新的教师培养认证标准,新标准强调结果导向(outcome-oriented),要求教师培养项目证明其毕业生对

①全美教师质量委员会于2013年开展了对美国传统型教师培养项目的质量评估,来自608个教师培养机构的607个中学教师培养项目和593个小学教师培养项目参与了此项评估,依据委员会协商制定的评估指标,采用四分制评估量表对培养项目进行打分。

②2013年国家教师教育认证委员会与教师教育认证协会(Teacher Education Accreditation Council, TEAC)合并为教师培养认证委员会(the Council for the Accreditation of Educator Preparation, CAEP)。

③第二编(Title II)要求各州甄别出本州的薄弱培养项目。

学生成绩的影响。^[8]此外,教师培养项目还需接受社会组织的外部排名评估以及州政府报告卡片的数据汇报要求。教师培养项目的增值评价由于遵循教师培养→教师教学质量→学生学业成绩三者之间的因果逻辑假设,恰好迎合了当下教师培养评价改革的结果导向趋势,因此逐渐在美国教育政策领域备受青睐。

(三) 数据系统与分析方法的发展为实施教师培养项目的增值评价提供了可能

近20年来,美国各州和学区为提高自身评价教师课堂教学质量的能力,不断增加对学生、教师和学校等不同层面的纵向数据系统的投资,^{[9][10]}例如《不让一个孩子掉队法案》要求各州对3~8年级、10~12年级中任一学段学生的数学和英语学科实施统一测验,以及2007-2008学年对3~5年级、6~9年级以及10~12年级学生的科学学科实施统一测验。在《不让一个孩子掉队法案》的规约下,许多州建设了学生评价数据系统,这些数据系统为教师培养项目增值评价的实施提供了基础性的学生学业成绩数据。

此外,教育研究领域量化方法的发展为实施增值评价提供了数据分析的技术支持。近几十年间,教师教育研究者的关注点从描述高效能教师的特点转移至判断教师对学生学业成绩的贡献度。^[11]其中,以教育生产功能为视角评估教师效能的研究者旨在通过控制影响学生学业成绩的教师之外的因素,进而判断教师的教龄或教师资格证书类型等教师的不同特点对学生学业成绩的贡献度。^{[12][13]}在这类研究中,研究者建构不同的统计模型以考察教师对学生学习的独特贡献。这些统计计量模型的开发成为教师培养项目增值评价中的核心内容。

(四) 教师增值评价的广泛推行为教师培养项目的增值评价奠定了重要基础

由于教师被视为与学生学习结果最为密切的校内因素,《不让一个孩子掉队法案》提出美国教育改革的核心目标是让每个教室都拥有一位高效能教师。^[14]“对学生学业成绩及其发展产生积极影响”被视为高效能教师的重要素质之一,而对学生学业成绩没有积极影响的教师素质和能力被认为是不必要的。对于

教师素质的这种认识已被纳入美国教育政策话语中。联邦教育部在《力争上游计划》中明确指出:“高效能教师应该是其所教学生的学业成绩可以获得大幅度提高的教师,更确切地说,就是两次标准化测验分数的改变。”其实早在1984年田纳西大学的统计学家威廉·桑德斯(William·Sanders)和罗伯特·麦克莱恩(Robert Mclean)就提出了以学生成绩的变化为标准评价教师的教师增值评价方法。20世纪90年代,田纳西州建构了田纳西增值评价系统(The Tennessee Value-Added Assessment System, TVAAS),用于评估学校和地区的增值,其中包括对教师增值的报告,但当时尚未应用于教师评价体系。自2011年开始,田纳西州的教师增值分数被正式纳入教师评价体系,占到教师评价总分的35%,并且评价结果可用于人事决策。1992年建立的达拉斯增值问责体系(the Dallas Value-Added Accountability System, DVAAS)将教师增值评价结果作为各学校对本校教师开展内部评价的依据。2015年,达拉斯市提出实施新的教师工资制度,将教师表现作为评判教师工资的标准。田纳西州和达拉斯市是较早运用教师增值评价的地区,此外,纽约、休斯顿、芝加哥、洛杉矶、华盛顿、俄亥俄、科罗拉多、佛罗里达等地区都开始将教师评价体系中纳入教师增值评价。^[15]教师增值评价研究及实践为教师培养项目增值评价提供了一种“概念验证”(proof of concept),使得将学生学业成绩作为教师及教师培养效能的评估指标有据可依。此外,大量研究表明教师培养与教师效能二者之间存在关联,^{[16][17][18][19]}因此在教师培养评价中考虑运用增值评价是对教师培养与教师效能二者之间因果关系的合理逆向映射。

二、教师培养项目增值评价的基本逻辑

虽然美国不同州的教师培养项目增值评价采用的增值模型存在差异,但增值评价遵循的逻辑基本一致。教师培养项目的增值评价以学生学业成绩的变化为标准来判断教师培养项目是否以及在多大程度上为公共教育体系输送了

合格教师。它内含的前提假设是：在控制其他变量的情形下，K-12学段学生的标准化测验分数的变化可以归因于教师培养的成效，其中关涉两层归因：学生成绩变化多大程度上归因于教师的影响；教师效能的高低多大程度上归因于教师培养项目的影响。因此教师培养项目的增值评价遵循图1所示的因果逻辑。^[20]

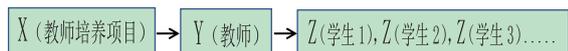


图1. 教师培养项目增值评价因果逻辑

资料来源：Amrein-Beardsley A, Barnett J H, & Ganesh T G. Seven legitimate apprehensions about evaluating teacher education programs and seven “beyond excuses” imperatives[J]. Teachers College Record, 2013, 115(12):1-34.

在具体操作层面，教师培养项目的增值评价采用的增值模型是一种利用随机试验的优势进行因果推断的尝试。^[21]换言之，如果图1所示三层面（即准教师选择教师培养项目、项目毕业生选择从教学学校、项目毕业生选择从教班级的分布均符合随机分配，则可以直接进行双层归因：将学生成绩视为因变量，教师视为自变量；将教师视为因变量，教师培养项目视为自变量。但在教育环境下，随机分配几乎是不可行的。因此在进行双层归因时，非随机分配会产生除自变量外、对因变量产生影响的其他协变量，即无关变量。研究者认为借助回归分析可以控制由诸多协变量导致的可能性偏差。在这一过程中，协变量的选择是剥离其他无关因素的影响，并建立教师培养项目→教师→学生成绩三者之间因果关系的关键。在这三个层面中，自下而上的协变量众多（见图2）。

由图2可见，影响学生学业成绩的协变量包括贫困程度、家庭和社会资源、学校组织和支持，以及住房、健康护理、工作、早期儿童服务等公共服务政策；^[22]影响教师效能的协变量包括学生人口因素、学校课程、班级规模、学校资源、初任教师入门培训及指导等。此外，进入教师培养项目之前，准教师具备的前专业技能和素质、性格、天赋等因素对教师效能也会

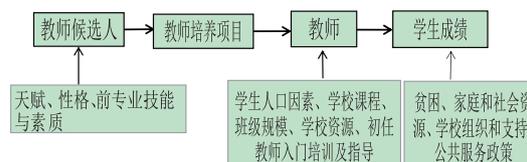


图2. 不同层级的学生学业成绩影响因素

资料来源：作者根据以下文献绘制：Cochran-Smith, et, al. Holding Teacher Preparation Accountable: A Review of Claims and Evidence. Boulder, CO: National Education Policy Center [EB/OL]. [2016-12-25] <http://nepc.colorado.edu/publication/teacher-pre>.

产生影响，但不应归因于教师培养的影响。不同州采用的增值模型的差异主要在于他们选择的协变量类型不同。图3的回归方程显示了基于多种协变量计算项目效能的基本方式。图3中的 A_{ijst} 指基于学生特征 (X)、课堂特征 (C)、教师特征 (T)、培养项目的固定效应 (Π)、随机误差项 (ϵ) 等计算得到的学校 s 和教师 j 的学生 i 在学年 t 的预测成绩。其中学生特征包括学生的种族、性别、经济地位、英语水平、出勤率等；课堂特征包括班级学生特征的平均水平，以及班级规模、年级、上一学年学生考试分数的平均值和标准偏差；教师特征包括年龄、性别、种族、资格考试通过情况及分数。^[23] 尤为值得注意的是，所谓的增值，是指学生成绩的实际值与预测值之差，因此教师培养项目的增值评价属于相对而非绝对评价。

$$A_{ijst} = \beta_0 + \beta_1 A_{ijs(t-1)} + X_{it} \beta_2 + C_{jst} \beta_3 + T_{jst} \beta_4 + \Pi_j + \epsilon_{ijst}$$

图3. 教师培养项目增值评价的回归方程

资料来源：Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. Teacher preparation and student achievement[J]. Educational Evaluation and Policy Analysis, 2009(31):416-440.

三、教师培养项目增值评价的实施操作：以路易斯安那州为例

路易斯安那州是美国首个实施教师培养项目增值评价的州，至今已推行十余年，并且

受到全美瞩目。路易斯安那州的教师培养项目增值评价由州政府授权路易斯安那州立大学心理学院开展项目评价研究,并与州教育部(Louisiana Department of Education, LDOE)、州评议委员会(Louisiana Board of Regents)共同推广实施,于2003年在全州范围内开始试行,并在2006年全面实施。它试图通过对教师培养项目实施增值评价来判断不同培养项目在多大程度上培养了能在3~9年级开展有效教学的合格教师。具体而言,它通过比较教师培养项目的毕业生所教学生的实际成绩与经验型教师所教学生的实际成绩,判断它们达到、未达到或超过基于多种协变量预测的成绩。在教师培养项目的增值评价中,项目效能证据的选择、项目效能的测量和项目效能增值结果的意义阐释构成增值评价的主要环节,并对增值评价结果产生直接影响。

(一) 项目效能证据的选择

教师培养项目效能证据的选择首先涉及如何选择评价对象,即参与评价的学生、教师和项目。在路易斯安那州的培养项目增值评价中,参与增值评价的学生确定为参加英语、阅读、数学、社会科学、科学五门州统考学科的3~9年级学生;参与增值评价的教师确定为获得从教学科教师资格证书,并且从教两年以内的公立学校新教师^①;参与增值评价的教师培养项目是从教于五门州统考学科的各学科公立学校新教师样本量达到25个的教师培养项目。目前,参与增值评价的教师培养项目类型包括传统的本科生教师培养项目和替代性教师培养项目,其中替代性教师培养项目包含从业教师项目(Practitioner Teacher Program)、硕士学位项目(Master's Degree Program)和教师任职资格项目(Certification-Only Program)。

确定评价对象之后,建构学生、教师和教师培养项目的纵向数据库是实施培养项目增值评

价的前提。路易斯安那州教育部已经建构了州标准化考试档案(iLEAP和LEAP-21)、关联学生信息和教师信息的路易斯安那教育问责数据系统(LEADS,曾被命名为“课程数据库”)、教师人口统计数据库、教师资格证书数据库和补充的学生数据库。考试档案和学生数据库提供学生出勤率、入学率、残障学生的情况、享受免费午餐的学生情况以及学生的人口统计变量(种族和性别)等信息。路易斯安那州教育部提供教师出勤率等信息。

建构完整的数据库之后,需进一步整合数据以确定最终可用于增值评价的数据。具体而言,要从不同的数据库提取在学生成绩、教师、教师培养项目三个层面保持一致性和连贯性的数据,剔除无法在三个层面保持一致性的数据。首先,需关联学生成绩、人口统计信息、出勤率、选课记录等学生数据,再将关联之后的学生数据与教师信息进行二次关联。教师的信息包括路易斯安那州教育部下属的计划、分析和信息资源部门提供的教师资格证书数据,路易斯安那州评议委员会提供的教师培养项目信息。其次,依据特定的考试科目(如数学、阅读、英语语言艺术、科学和社会研究)将课程代码进行分组,如关联四年级的阅读与阅读科考试分数,不能与标准化测验进行合理关联的课程代码将被剔除。在数据整合过程中,当1位以上的教师同时教授同一门学科时,学生成绩的比重根据他在这一学科注册的教师数量进行加权。例如,一位学生注册两门数学课,将出现两位教师关联的记录,但与只注册一门数学课的学生相比,这两位数学教师将分别只占50%权重。综上所述,关联之后的数据将保持学科课程、学生成绩、教师以及教师培养项目之间的一致性。

(二) 项目效能的测量

路易斯安那州的教师培养项目增值

^①路易斯安那州立大学增值评价研究小组在2006年之前将可以参与增值评价的初任教师从教年限定为三年。自2006年开始,研究小组以经验型教师(教龄为21~30年)效能为参照组,比较不同教龄的教师效能与经验型教师效能的差异,发现从教两年之后的教师效能逐渐接近经验型教师的平均效能,因此最终确定可以参与增值评价的教师是从教两年以内的初任教师。

评价采用的是州立大学乔治·诺艾尔(George Noell)教授开发的多层线性模型(Hierarchical Linear Modeling)^[24](见图4)。路易斯安那州的多层线性模型运用层层嵌套的模型分层,增值即指多层线性模型中回归方程的残差值,也就是学生成绩的实际值与预测值之差;预测值是依据学生上一学年的成绩、人口信息等协变量计算的成绩;实际值指学生标准化测验的实际成绩。比较初任教师的平均增值与经验型教师的平均增值,即可获得毕业于同一个教师培养项目的初任教师的平均增值,也就是该培养项目的增值评价结果。路易斯安那州教师培养项目增值评价结果表征的是毕业

于同一个教师培养项目的初任教师对学生成绩的平均影响。路易斯安那州的评估团队暂时还没有充分的数据资料考查教师培养的不同要素(如教师招聘、教师培养项目的入学选拔、学科知识培养、教育学知识培养、教育实习、毕业筛选等)对学生学业成绩的影响。

多层线性模型中协变量的选择是剥离教师因素之外对学生成绩产生影响的因素的关键,也是设计回归方程计算学生成绩预测值的前提。路易斯安那州采用的多层线性模型主要涵盖三个层面的协变量:学生、班级和学校,各个层面的协变量如表1所示。^[25]

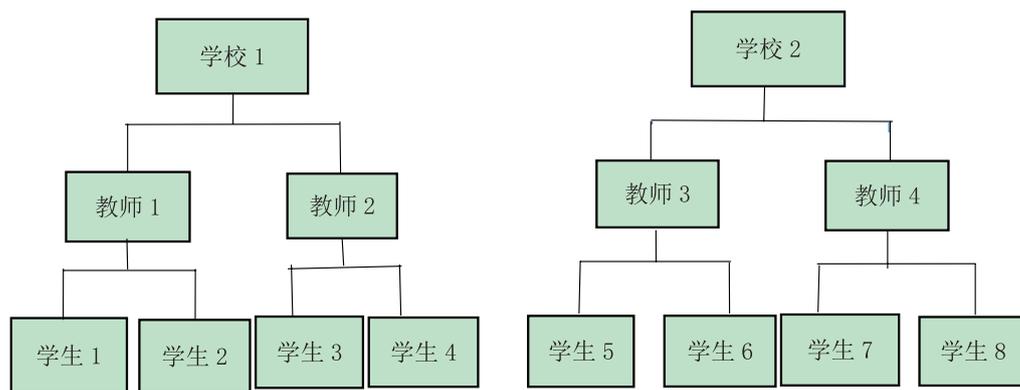


图4. 路易斯安那州教师培养项目增值评价多层线性模型

(三) 项目效能增值结果的意义阐释

路易斯安那州立大学与高等教育委员会、教师教育计划委员会协商将教师培养项目的增值结果划分为五个层级,目的在于通过等级划分赋予增值评价结果以不同的含义,避免公众过度关注差距不大的教师培养项目的增值分数,而这类细微差异更可能归因于测量技术的

误差,并非代表不同培养项目真实的质量差异。路易斯安那州教师培养项目增值评价结果的五个层级代表的项目质量含义如表2所示。^[26]根据现行的教师培养项目干预政策,增值评价结果为层级3以下(即层级4和层级5)的教师培养项目需对自身进行重构和改进,并在下一学年再次接受增值评价。

表1. 路易斯安那州教师培养项目增值评价协变量类型

层级	协变量类型
学生	性别、种族、天赋、母语、英语熟练程度、经济背景(是否享用免费午餐或廉价午餐)、上一学年的各科成绩、心理障碍、智力障碍、特殊教育需求、学习障碍、残疾学生、出勤率
班级	学生性别比例、少数族裔学生比例、享用免费午餐或廉价午餐的学生比例、天赋学生的比例、英语熟练程度低的学生比例、上一学年各学科的班级平均成绩水平、特殊教育需求学生比例、残障学生比例、教师出勤率
学校	学生性别比例、少数族裔学生比例、享用免费或廉价午餐的学生比例、天赋学生的比例、英语熟练程度低的学生比例、上一学年各学科的学校平均成绩水平、特殊教育需求学生比例、残障学生比例

表 2. 路易斯安那州教师培养项目增值评价结果等级

层级1	项目评价效果优于经验型教师的平均效能, 代表其毕业生教学效能高于经验型教师的平均水平
层级2	项目评价效果优于初任教师的平均效能, 代表其毕业生教学效能比初任教师的平均效能更接近于经验型教师的平均水平
层级3	项目评价效果在初任教师平均效能的一个标准误差范围以内, 代表其项目毕业生的教学效能接近于初任教师的平均水平
层级4	项目评价效果在初任教师平均效能的一个标准误差以下甚至更低, 代表其项目毕业生的教学效能未达到初任教师的平均水平
层级5	项目评估效果显著低于初任教师的平均水平

四、教师培养项目增值评价面临的争议

任何评价方法都有基本的出发点, 教师培养项目的增值评价在美国教师教育评价领域的兴起主要因为已有的评价方法无法回应日益强化的教师教育绩效问责。教师培养项目增值评价以学生成绩的变化为标准来判断教师培养项目的成效, 具有区别于其他教师培养评价方法的两个显著特征。首先, 教师培养项目增值评价扮演着过滤器的角色。增值评价甄别出高效能教师培养项目, 可以作为其他培养项目模仿的对象, 而筛选出低效能教师培养项目, 有助于进行有针对性的审查, 从而推动其不断地改进。^[27]其次, 教师培养项目的增值评价试图创造公平的教师培养评价环境, 追求评价的公平性。从评价客体而言, 增值评价关注每一位学生的成绩; 从评价标准而言, 增值评价兼顾学生的原有成绩水平以及影响学生成绩的其他因素, 采用学生成绩的变化作为评估指标, 而非成绩均值或任意标准等绝对值。^[28]但不可否认的是, 由于教师培养项目的增值评价研究和实践均处于起步发展阶段, 因此依然面临一些争议。

(一) 关于增值评价的效度和信度

教师培养项目增值评价面临的最大争议是它的效度, 即增值评价的结果是否准确地表征了教师培养项目对学生学业成绩的贡献度备受质疑, 具体而言主要体现在两个方面: 首先, 如何将教师培养的影响与培养项目入学选拔的影响进行区分。由于各培养项目的入学选拔标准存在差异, 教师培养机构尚未建构完整的准教师入学标准的数据库, 并且如何将入学选拔的影响纳入增值模型, 涉及更加复杂的计量学设

计过程, 目前还未有研究者提出令人信服的用于对教师培养项目影响学生成绩程度进行合理归因的方法。由于不能明确区分教师培养的影响与项目入学选拔的影响, 增值评价缺乏坚实的证据来判断教师培养项目获得高的增值分数应归因于该项目吸引了高质量的教师候选人, 还是该项目培养出了高质量的教师,^[29]因此也就无法准确判断教师培养项目的成效。其次, 将增值评价运用至大规模的教师培养项目评价中需遵循样本随机分配的原则。^[30]具体而言, 也就是准教师进入教师培养项目、项目毕业生进入从教学校、项目毕业生进入从教班级均应符合随机安排。但在实际的教育环境中, 不同培养项目依据不同的选拔标准招募教师候选人; 大多数项目毕业生依据地理位置选择从教学校, 因而教师培养项目类型与毕业生从教学校的类型和地理位置高度相关; 不同班级的学生分配亦不符合随机原则, 更多是由学生自身、家长和当地政策等因素决定。非随机分配产生的影响主要是衍生了诸多影响教师教学质量和学生成绩的无关因素, 如准教师先天固有的能力与动机水平、不同班级学生的差异等, 这些无关因素被视为是增值评价计量过程的系统误差, 它在多大程度上影响增值结果依然是高度不稳定的。^[31]对于如何弱化非随机分配对增值评价产生的影响, 尽管有研究者提出可以采用协变量调整方法控制学生、教师和学校层面的变量, 从而在一定程度上能规避项目毕业生在选择从教学校时的非随机分配产生的影响,^[32]但这种方法依然无法消除政策制定者和公众对增值评价方法效度的质疑。

除了效度问题之外, 相关研究表明教师培

养项目增值评价的信度也具有可推敲之处。美国教育研究协会和国家教育学会指出增值评价的结果具有高度的不稳定性,因为使用不同的增值模型,或使用同一模型,但测量不同学科或不同学年的学生成绩,最终的增值评价结果相差甚远。此外,使用不同的测试卷得出的增值评价结果会存在差异;班级规模也会影响增值评价结果的稳定性,班级规模越小,增值评价结果越不稳定。^[33]

(二) 关于增值评价的适用范围

教师培养项目的增值评价以学生成绩为指标判断教师对学生成绩的贡献度,以及教师培养项目对教师效能的贡献度。这种方法本身存在一个局限:无法将所有教师培养项目和K-12学段学生纳入评价过程。因为增值评价的前提要求是学生至少有一次先前成绩和一次当前成绩,因此不参与州统考的学前阶段、1~2年级以及参与州统考的三年级学生(三年级的成绩作为四年级的先前成绩)均不能包含在增值评价过程中。^[34]由此可见,增值评价的适用范围仅限于特定的年级和学科,如北卡罗来纳州的教师培养项目增值评价的学生成绩样本为3~12年段的数学、阅读,以及中学学段的科学和社会研究的州统考成绩;田纳西州的教师培养项目增值评价以4~12学段学生的数学、阅读/语言艺术、科学和社会研究的州统考成绩为样本。除此之外,增值评价要求教师培养项目在初任教师样本量方面需达到各州的要求,如北卡罗来纳州要求每个学科的项目毕业生数量需达到10个,而路易斯安那州更要求每个学科初任教师数量达到25个的教师培养项目才能参与增值评估。教师培养项目毕业生的就业方向包含多种情况:从毕业生从教地域的角度来看,包括从教于本州的中小学和其他州的中小学;从毕业生从教学学校性质的角度来看,包括从教于公立中小学和私立中小学;从毕业生从教学学科的角度来看,包括从教于州统考的学科和不参与州统考的学科;以及不从事教师职业的情况。而上述各州对教师培养项目毕业生数量的要求均指从教于本州公立中小学的州统考学科教师数量需达到如上要求,因此大量的教师培养项

目不能参与增值评价。同时,增值评价的前提是搜集3年以上的学生资料,并需建构大型的纵向数据库,时间成本和经济成本较高,也影响了增值评价的适用范围。

五、结语

在教师教育评价领域,不存在一种评价方法能全面且综合地考查教师的教学质量或教师培养项目效能的全部内容,^[35]增值评价也不例外。增值评价试图兼顾对学生成绩产生影响的其他因素,为教师培养项目提供公平竞争的环境。但由于增值模型的复杂性以及数据搜集的多样性,技术层面的统计计量可能会导致增值评价结果的偏误,继而使评价结果发生反向倾斜,因此政策制定者和公众需审慎地使用及看待增值评价的结果。此外,增值评价本质上是以衡量为目的的筛选,而非以发展为目的的推动,因此在一定程度上,这种评价方式或许达到了问责或监督的目的,但没有实现诊断并回馈教师培养项目的不足,从而助力项目质量提升的目标。因此如何将研究者设计的增值模型恰当地运用到教师教育问责体系中,需要统计计量学专家和政策制定者的综合考量,需同时权衡教师培养项目的增值评价带来的积极和消极影响。

参考文献:

- [1][7][8][27] Feuer M J, Floden R E, Chudowsky N, & Ahn J. Evaluation of teacher preparation programs: Purposes, methods and policy options[M]. Washington, DC: National Academy of Education, 2013.
- [2][20] Amrein-Beardsley A, Barnett J H, & Ganesh T G. Seven legitimate apprehensions about evaluating teacher education programs and seven “beyond excuses” imperatives[J]. Teachers College Record, 2013, 115(12):1-34.
- [3] NCTQ(National Council on Teacher Quality). 2013 Teacher pre review: a review the national teacher preparation program[EB/OL]. [2016-12-15] http://www.nctq.org/dmsView/Teacher_Prep_Review_2013_Report.
- [4] Zeichner K. Competition, economic rationalization, increased surveillance, and attacks on diversity: Neo-

- liberalism and the transformation of teacher education in the U.S[J]. *Teaching and Teaching Education*, 2010, 26(8):1544-1552.
- [5]Cochran-Smith, M. "Re-culturing" teacher education: Inquiry, evidence, and action[J].*Journal of Teacher Education*, 2009, 60(5):458-468.
- [6]Barnett J H& Amrein-Beadsley A. Actions over credentials: Moving from highly qualified to measurably effective[Commentary].*Teachers College Record*[EB/OL].[2016-12-09] <http://www.tcrecord.org/Content.asp?ContentID=165172011>.
- [9][35]Crowe, E. Measuring what matters: A stronger accountability model for teacher education. Washington, DC: Center for American Progress[EB/OL].[2017-03-19] <http://www.american-progress.org/issues/2010/07/pdf/teacher-accountability.pdf>.
- [10]National Research Council. Preparing teachers: Building evidence for sound policy[M]. Washington, DC: National Academies Press, 2010.
- [11]Nye B, Konstantopoulos S,& Hedges L. How large are teacher effects?[J]. *Educational Evaluation and Policy Analysis*,2004,26(1):237-257.
- [12]Darling-Hammond L, Berry B,& Thoreson A. Does teacher certification matter? Evaluating the evidence[J]. *Educational Evaluation and Policy Analysis*,2001, 23(1):55-77.
- [13]Goldhaber D D,& Brewer D J. Does teacher certification matter? High school teacher certification status and student achievement[J]. *Educational Evaluation and Policy Analysis*, 2000, 22(2):129-145.
- [14]Plecki M L, Elfers A M, & Nakamura Y. Using evidence for teacher education program improvement and accountability: An illustrative case of the role of value-added measures[J]. *Journal of Teacher Education*, 2012,63(5):318-334.
- [15]边玉芳, 孙丽萍. 教师增值性评价的进展及在我国应用的建议 [J]. *教师教育研究*, 2015 (1): 88-95.
- [16]Darling-Hammond L. Assessing teacher education: The usefulness of multiple measures for assessing program outcomes[J]. *Journal of Teacher Education*,2006,57(2):120-138.
- [17]Darling-Hammond L. Constructing 21st-century teacher education[J]. *Journal of Teacher Education*, 2006,57(3):300-314.
- [18]Darling-Hammond L,& Sykes G. Wanted: A national teacher supply policy for education : The right way to meet the "highly qualified teacher " challenge. *Educational Policy Analysis Archives*,11(33)[EB/OL].[2017-01-23] <http://epaa.asu.edu/epaa/v11n33/2003>.
- [19]Wilson S, Floden R,& Ferrini-Mundy J. Teacher preparation research: Current knowledge, gaps and recommendations. Center for the Study of Teaching and Policy, University of Washington.[EB/OL][2016-11-25] <http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf>.
- [21]Organization for Economic Co-operation and Development. Measuring improvements in learning outcomes: Best practices to access the value-added of schools[R]. Paris: Author.2008.
- [22]Cochran-Smith,M.,Stern,R.,Sanchez,J.G.,Miller,A.,Keefe,E. S.,Fernandez,M.B.,Chang,W.,Carney,M.C.,Burton,S.,&Baker,M. Holding Teacher Preparation Accountable: A Review of Claims and Evidence. Boulder, CO: National Education Policy Center[EB/OL].[2017-2-12] <http://nepc.colorado.edu/publication/teacher-pre.2016>.
- [23]Boyd D J, Grossman P L, Lankford H, Loeb S, & Wyckoff J. Teacher preparation and student achievement[J]. *Educational Evaluation and Policy Analysis*, 2009 (31):416-440.
- [24][25][26]Gansle K A, Noell G H, Knox R M,& Schafer M J. Value added assessment of teacher preparation in Louisiana: 2005-2006 to 2008-2009[R].Department of Psychology, Louisianan State University. 2010.
- [28]Gansle K A, Noell G,& Burns J M. Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana[J].*Journal of Teacher Education*,2013,63(5):304-317.
- [29]Noell G H, Brownell M T, Buzick H M, & Jones N D. Using educator effectiveness measures to improve educator preparation programs and student outcomes (Document No. LS-1)[EB/OL].[2016-11-12] <http://ceedar.education.ufl.edu/tools/literature-syntheses/2014>.
- [30][33]Chiang,Y,Cole,C.Delandshere,R.,Kunzman,R.

- ,Guarino,C.,Rutkowski,D.,Rutkowski,L.,Svetina,D.
Yuan,X.,&Zhou,Y. Using Value Added Models to
Evaluate Teacher Preparation Programs (White paper)[R].
Bloomington, IN: Indiana University.2011.
- [31][32]Noell G H, Porter B A, Patt R M,& Dahir A. Value
added assessment of teacher preparation in Louisiana:
2004–2005 to 2006–2007[R].Department of Psychology,
Louisianan State University. 2008.
- [34]Lincove J A, Osborne C, Dillon A, & Mills N. The politics
and statistics of value-added modeling for accountability
of teacher preparation programs[J]. Journal of Teacher
Education, 2014(65):24–38.

An Analysis of Value-Added Assessment of Teacher Preparation Program in United States: The Case of Louisiana State

FENG Hui, RAO Congman

(Institute of International and Comparative Education, Northeast Normal University, Changchun, Jilin, 130024)

Abstract: Since the turn of the century, teacher education in United States is faced with the double pressure from both internal quality improvement and external accountability, and the existing teacher education assessment approaches cannot meet the demand of accountability. In this context, with the development of data systems and data analysis methods as well as the gradual promotion of value-added evaluation of teachers, value-added assessment of teacher preparation program has emerged and gradually gained more and more recognition in United States' education policy area. Based on the causal logic of teacher preparation program, teacher's classroom teaching efficacy and student's learning achievement, value-added assessment of teacher preparation program attempts to judge the unique contribution of teacher preparation program to student achievement via the designed statistical models. The selection of program efficacy evidence, estimation of models and interpretation of assessment results form the three main parts of value-added assessment of teacher preparation program. Value-added assessment of teacher preparation program has responded to the demand of strengthening teacher education's responsibility toward basic education to some extent, but there are still some problems regarding the validity and reliability, as well as the scope of application.

Key words: value-added assessment; teacher preparation program; teacher education evaluation; Louisiana; United States

本文责编: 吴婷