

美国 UTOP 课堂教学质量评估系统的 探索与反思^{*}

曹 慧 毛亚庆

摘要 有效的课堂评价标准,有助于教师发现教学中的薄弱环节,改进课程结构和教学方式,提升教学质量。本文以美国的课堂教学质量评价系统 UTOP(U Teach Observation Protocol)为例,试图探讨建设课堂教学评估工具的必要性。UTOP 系统是美国应对数学与科学(STEM)课程教师质量堪优的系列举措中的重要步骤和工具,本文详细介绍了其理论基础、评分标准、评分步骤,并分析了 UTOP 系统实施效果,探讨了它在教学评估中的优势和局限,以及在我国课堂教育评估中的意义,以期对国内相关研究和实践工作提供参考。

关键词 课堂教学质量; 教学评估; 课堂观察

作者简介 曹 慧/北京师范大学教育学部教育管理学院博士后 (北京 100875)
毛亚庆/北京师范大学教育学部教育管理学院教授 (北京 100875)

课堂教学质量的观察评估方法是建立在教学论的前沿理论和教育实践的提炼总结的基础之上,把有效教学的整个过程拆分为许多不同的细致维度,并提供评分标准,使教研人员和教师都能够了解到课程中的优势与劣势,从而有效改变教学模式和手段,提升教学质量。国际上近 10 余年涌现出多种各有特色的课堂教学质量的观察评估工具,并积累了大量实证研究结果。^[1]其中,美国德克萨斯州立大学 U Teach 教师中心开发的教学观察方案(U Teach Observation Protocol,简称 UTOP)信效度好、易于操作、适用的年级范围广,^[2]有很好的借鉴意义。本文将详细介绍 UTOP 的设计思路、评分标准和实施效果,并探讨它在教学评估中的优势和局限,以期为我国课堂评价系统的开发和使用提供参考。

^{*} 本文系国家自然科学基金项目“中国基础教育内在质量提升与学校管理改进研究”(项目编号:71373028)的阶段性成果。

一、研发背景与基础

(一) 研发背景

1. 美国基础教育发展的大势所趋

课堂的观察评估可以追溯到上世纪 60 年代的弗兰德斯(Flanders)的师生言语互动分析系统(FIAC), 通过将师生课堂上的对话进行编码来评估课堂情况。^[3]随着美国政府“不让一个儿童落后法案”(The No Child Left Behind Act)的实施, 教学质量监测成为基础教育的重要任务, 于是, 教师培训和测评成为公立学校的重要环节。^[4]但是以成绩为导向的教学质量评价方案, 由于不能直接指导教学改进, 受到广泛质疑和反对。^[5-6]因此, 众多基于实证研究的以课堂观察为基础的教学评估工具得到充分发展和应用, UTOP 在这个大背景下快速发展。

2. 教育培训实践工作的自然产物

UTOP 的开发者——Uteach 中心是设立在全美共 22 个州或地区的 44 所大学内的负责培养全国理科教师的培训中心。为应对美国公立学校在理科教学上的落后局面——理科师资缺乏和教学质量不高的困境, 也为了帮助理科教师发展和提升全民的理科教育, 美国国家自然科学基金会(National Science Foundation, 简称 NSF)支助成立 Uteach 中心以培养中学的科学、技术、工程和数学(STEM)课程教师。UTOP 的研发就是为了评价其参训学员的学习效果, 量化教师培训的实际效果, 且有针对性地帮助教师成长。

由于 UTOP 的评估效果良好, 其应用范围也被扩展到其他方面, 如:(1) 一般性的教师的课堂教学质量评价, 在这一方面, UTOP 是美国现有的课堂教学质量评估工具中最常被使用和最受信任的工具之一;(2) 由于其维度开发的科学性和指导性, 现在 Uteach 中心也使用 UTOP 来进行教师课堂教学的培训。^[7]

(二) 理论基础

1. 教学研究理论

UTOP 研究者在研发时对教育教学理论进行了梳理, 以最为前沿的教育理念为参照, 尤其是关于学习机会(opportunity to learn)的研究及实践。^[8]学习机会强调对学校或班级中所有学生都要平等提供促进其学习的机会,^[9] 与此相关的所有影响学生学习条件、课程和教法等的学校内因素都是其研究内容。^[10] 在 UTOP 编制过程中, 尤其重视促进提升学习机会的特定的教学实践方式, 他们认为这可以促进学生的学术成就, 即除了关注教师的课堂驾驭能力之外, 还关注教师是否考虑到学生之前的知识、教学任务的性质和目的, 以及学生的学习投入。^[11] 此外, 受社会建构主义理论影响, UTOP 还强调在数学和科学学习中以概念理解为目的的教学活动的重要性, 尤其是师生对概念进行探讨, 并在不同概念间建立联系, 让学生自发形成概念图的过程。^[12] 总之, UTOP 的观察要点首先建立在促进学习机会的各种有效教学方式上。

2. 教师职业标准

美国科学教育标准(*National Science Education Standards*, 1996)和美国数学教师协会职业标准(*National Council of Teachers of Mathematics Professional Standards*, 1991)颁布之后,对数学和科学教育中的课程、内容和课堂结构都提出了新的要求,从对某一项专业的学科知识和能力的强调转向对丰富的科学知识和适应当代社会的能力的强调,STEM 等学科的课堂教学任务更多地强调对科学和工程学的跨学科理解,强调动手操作的科学实践(而不是科学探究)。UTOP 的设计完全遵照这样的标准,尤其在评分中强调教师对于学生思维能力促进及教师创造学生科学实践机会等方面的能力。

3. 前人的课堂观察工具

UTOP 对当时课堂观察评估工具进行总结和梳理,从中借鉴有效的评价要点。而当时评价要点全面、有学科针对性的评估工具《教室观察法》^[13](*Classroom Observation Protocol*, 简称 COP)被 UTOP 作为主要参考标准。COP 是一个以质性记录为主的课堂观察标准,记录课堂主要的活动、材料和教学目的,包括分维度和整体评分的量化评估条目。UTOP 研究者认为 COP 虽然是当时最为平衡、有实证基础的评价体系,但是仍有两个方面的问题:(1)对教师的学科知识体系和教学知识体系重视不足;(2)COP 编制时的教学观仍然偏向质询和调查,没有充分发挥学生的自主建构能力。因此,UTOP 修改或去掉了 COP 中观念陈旧的条目,补充了与学生自主学习相关的 7 个观察要点,还附加了教师课后访谈来了解教师对于学科知识的看法,最终形成了完整版的 UTOP。

二、UTOP 评估要点^①

UTOP 根据使用情境的不同,分为完整课堂评价版和课堂视频评价版。两个版本都包括四个维度的观察要点,但因课堂视频评价版仅评估授课视频,不进行课后的教师访谈环节,因此相对于完整版,视频版中被删减了涉及课后访谈的观察要点,最后形成的完整版共 32 个题目,主要用于现场听课评估,视频版则包括 22 个题目。^[14]UTOP 评估系统观察要点的四个维度为:课堂环境(*Classroom Environment*)、课程结构(*Lesson Structure*)、执行效果(*Implementation*)和教学内容(*Math/Science Content*)。完整版中每个维度分别包括 6-8 个评分要点,视频版则分别为 4-5 个评分要点(即一个题目)。表 1 中列出了所有的观察要点,其中视频版包含的观察要点用“*”标出。

^① 这里介绍的为理科版的 UTOP,文科版的 UTOP 刚刚有试用版,尚未正式发表,且尚无实证证据支持。

表1 UTOP的测量维度及其课堂观察要点

测量维度	课堂观察要点
1. 课堂环境 Classroom Environment	<p>*(1)课堂参与(Engagement):教师创造良好的课堂氛围,让学生提出想法、问题、猜测和意见。</p> <p>*(2)课堂互动(Interactions):学生之间通过合作解决问题,例如通过对课程进行讨论得出问题答案。</p> <p>*(3)课堂对话(Conversation):通过分析学生的对话,能够显示出学生积极深入地思考课堂问题。</p> <p>*(4)学生专注(On-task):课堂上学生专注于课堂任务。</p> <p>*(5)课堂管理(Management):教师的指令清楚有效,学生能保持良好的课堂纪律。</p> <p>(6)课堂布置(Organization):课堂内的设施和教具满足学生上课需要。</p> <p>*(7)课堂公平性(Equity):课堂环境对不同学生公平一致,不因学生的年龄、性别和身体条件有差异。</p>
2. 课程结构 Lesson Structure	<p>*(1)课程顺序(Sequence):课程的结构明确,学生有清晰的学习目标和路径。</p> <p>*(2)重点突出(Importance):课程关注重要的科学概念,而不是应试为主的解题技巧。</p> <p>(3)即时评价(Assessment):教师能随时评估学生对知识的理解程度。</p> <p>*(4)课程探索(Investigation):教师能够提出探索性问题,帮助学生理解重要概念。</p> <p>*(5)课程资源(Resources):利用视频、音频、模型等工具展示课程内容,有效地帮助学生理解概念。</p> <p>*(6)课程反思(Reflection):教师课后对自己授课的情况复盘,寻求改进。</p>
3. 执行效果 Implementation	<p>*(1)提问(Questioning):教师通过提问的方式促进学生关注和思考问题。</p> <p>*(2)参与(Involvement):教师尽量让所有学生参与到课程中,增加学生之间的交互性。</p> <p>*(3)调整(Modification):教师通过测验了解学生进度,并能随之调整课堂内容。</p> <p>*(4)时间分配(Timing):课程各部分内容安排合理,给学生留下思考和理解的时间。</p> <p>(5)联系(Connections):课程的内容和活动,能够让学生联系到以往的知识 and 经验。</p> <p>(6)安全性(Safety):教师讲授课程时,能够联系到与安全、环境、伦理等相关的问题。</p>
4. 教学内容 Math/Science Content	<p>*(1)意义(Significance):课程内容有意义,符合学生的认知发展水平。</p> <p>*(2)水平流畅性(Fluency):教师对课程理解深刻,因此对学生的指导有效且流畅。</p> <p>*(3)准确性(Accuracy):教师的板书等展示材料内容准确。</p> <p>(4)评估(Assessments):教师的提问、随堂测试、作业等与教学目标紧密结合。</p> <p>*(5)抽象(Abstraction):对概念和知识的抽象表述合理,有助于学生理解。</p> <p>*(6)相关性(Relevance):阐述内容在知识系统中的重要性。</p> <p>*(7)交互性(Interconnections):介绍本学科知识在其它学科中的相关应用。</p> <p>*(8)社会影响(Societal impact):解释课程内容在现实中的作用,以及其历史地位和影响。</p>

注:*代表视频评估版包含的评分要点。

三、UTOP 的评估方法

(一) 计分方法

UTOP 认为通过三种方式的综合应用,观察者能够准确、高效、细致地进行课堂评价。这三种方式包括分维度下的量化评分标准、质性的评分依据记录和整体评价。

首先,每个观察要点都需要按照 5 点里克特量表计分:1 分为最低分,表示在该观察点上表现最不符合期望;5 分为最高分,表示该课堂呈现出最理想的状态。如果课堂上某个观察要点未出现或者无法判断,可以记录为 NA(不适用, Not Applicable)或者 NK(不确定, Not Known)。

第二,质性观察也是不可或缺的评分环节。UTOP 在量化给分的同时还需要观察者记录该评分的判断依据,即每一个评分题目要求评估者用文字简述评分对应的课堂特征,包括学生行为、教师的讲授内容等,促进评分者更细致地观察课堂情况和教学细节。

第三,UTOP 在每个维度的最后,还附加了一个克特量表 5 点评分问题——评分者对该维度的整体性评价(Synthesis Rating)。此处的整体性评价,并不是对本维度每个观察要点的得分进行数字平均,而是观察者对本维度课堂情况的主观感受。同样,观察者也需要记录自己整体性评价的评分依据。该评估与前面的分要点评估的分数将形成交叉验证。设置该总评分的原因是在心理学的研究中发现总评分和分要点评分通常会有不同的预测效力。

(二) 计分步骤

完整版的 UTOP 的实施过程分为三个步骤:课堂评估、教师问卷调查和课后教师回访。课堂评估是 UTOP 最核心的步骤,是由经过专业训练的观察者对教师授课过程(或者授课视频)进行评分,并记录相应的评分依据。在这个步骤中,教师的个人信息,如教师的学历和受训经历等不会告知评分者,以保证授课或者课堂视频的分析过程独立性和客观性。此类信息在教师的问卷调查这个步骤中采集,包括教师的性别、年龄、所在学校和班级、教育背景及培训经历等。第三个步骤是课后教师回访,安排在授课或者课堂视频评估后立即进行,通过面对面访谈、视频会议、电话采访等方式完成(如果没有条件也可以采用问卷的方式收集信息)。访谈的内容包括:课程的教学目的、备课依据、教学材料准备、学生特殊需要和课后的经验总结等。视频版仅完成课堂视频评估步骤,不需要问卷调查和课后访谈。

(三) 效果保证

作为主观性较强的观察评分系统,最易受质疑之处在于如何保证评分的信度和效度。UTOP 的评分过程为此设置了如下保障机制:量化和质性评分方式交叉验证;建立标准化的评分体系;详细的示例;严格的评价者培训。其中,第一点前文已经详细介绍,这里将就后面三点进行述评。

1.标准化的评分体系建设

为了提升评分信度和效度,UTOP 的研究者编制了详细的操作手册,不仅详细介绍了每个维度及具体的观察要点的概念,更是给出了具体、可量化、可操作的给分标准和每个分数点的课堂实例。以其中的观察要点——学生专注为例,其评分标准如下:

1分:课堂内半数以上学生没有认真听课,不听课的行为影响整个班级。

2分:整堂课上,一半到四分之三的学生能够注意听课,课堂上经常有不听讲的行为。

3分:约75%的学生能够在课堂上认真学习,部分同学听课不认真,但是他们不听课的时间很短(几分钟)。

4分:课堂上不认真听课的行为很少,仅有几个学生会出现不专注的情况。

5分:几乎全部学生全程认真听讲,没有不专注的情况。

2.详细的示例

UTOP 的评分标准采用了易于量化的描述方式,观察者可以准确地对课堂情况进行归类评价。为了提高观察者评分效率,UTOP 还提供了不同得分对应的典型课堂行为。仍以学生专注为例:

1分:大部分学生不能专注于课堂学习活动,而且打扰其他人,比如打闹、聊天、睡觉、听音乐等。

2分:很多学生在私下聊天或者不参与课堂活动。

3分:老师在讲解题目的时候,部分同学并没有记录或者聆听;少数人在课堂上翻看其它材料。

4分:学生能够认真做笔记,对老师布置的问题主动寻找答案,很少有参与的学生。

5分:小组讨论的过程中,同学们认真思考、讨论课程相关的概念和问题,没有不认真的行为;其它小组同学发言或者展示的时候,大家耐心聆听思考,并提出有深度的问题,或者仍专注思考自己小组的问题,但整个过程不影响他人。

UTOP 提供的课堂教学中的典型情况,让观察者在评分时能够直接进行归类,提高了评价效率。

3.严格的评价者培训

对UTOP系统来说,其可靠性建立在不同观察者对课堂教学准确一致的评价上。为了让观察者达到统一标准,观察者需要经过严格的培训:首先,受训者集中学习UTOP四个维度的内容,以及不同观察要点的评分标准和课堂实例;然后,受训者分组观看课堂视频,进行小组讨论,基于讨论结果对视频内的教学质量打分;最后,培训专家给出视频的标准评分和打分理由。受训者经过反复比较学习,熟练掌握评分规则后,再正式承担评分工作。此外,观察者也会按照双人组或者三人组的模式进行评分,对得分不同的观察要点进行充分讨论,汇总后给出评价结果。

综上所述,UTOP系统采用了高度结构化的评分流程和严格的观察者培训,

能够保证课堂教学的量化评价稳定有效。其研究结果显示,不同的观察者在对同一教学过程进行评价时,其评分具有良好的一致性。〔15〕

四、UTOP 的实施效果

(一)UTOP 的优势

当前的研究者对课堂评价系统通常关注以下四个方面的功能:指出课堂教学中的优点和不足,给教师发展提供建议;〔16-17〕根据课堂评估的结果,预测学生的学业成绩,评选出对学生成绩帮助大的优秀教师;〔18〕评价系统能够适应不同年级的需要,也适应不同科目的需要;〔19〕课堂评估的人力、时间消耗可控,并具有高度可靠性。〔20-21〕本文也将从以上四个方面介绍 UTOP 的实施效果。

(1)UTOP 找出课堂教学中的薄弱环节,帮助教师提升课堂教学水平。UTOP 对课堂教学的分析,其观察要点源自美国科学教育标准和美国数学教师协会职业标准,覆盖了数学和自然科学的教学要求。因此,通过 UTOP 评分能够找出教师教学过程中存在的问题,提供针对性建议,从而改进教学方法和教学内容。UTOP 的设计初衷,是评价 NSF 资助的教师培训项目是否有效。研究结果显示,经过培训的教师,其 UTOP 评分显著高于未经过培训的教师,而且在持续 3 年的测量中都存在显著优势。〔22〕

(2)UTOP 能预测不同教师对应的学生成绩的提升。美国盖茨基金会开展的有效教学评估项目(Measurements of Effective Teaching,简称 MET),以教师的增值评价(Value-added Measurement,简称 VAM)为基础,分析不同教师对学生成绩提升效应,寻找对学生成绩提升效果最明显的教师。〔23〕增值评价是指教师在学生成绩提升中的贡献,该指标的优势能够避免绝对成绩评价中因地域、生源等非教学因素带来的影响,从而使得跨区域比较成为可能。〔24〕同时,MET 使用包括 UTOP 在内的多种课堂评价系统对教师的课堂教学视频的评分,寻找课堂教学质量最高的教师。通过比较增值评价和课堂评价的结果,研究者发现 UTOP 的结果与增值评价的结果具有良好的一致性,其相关系数优于其它课堂分析系统。〔25〕

(3)UTOP 适用的年级范围广。UTOP 适用的年级范围贯穿从幼儿园到大学本科的各个教育阶段,这是其它课程分析系统所不具备的,使得跨年级比较成为现实。跨年级比较的实现能够追踪教师发展并进行不同年级的教学质量比较,从而为学校内进行教师教学技能发展提供了有益的工具。

(4)UTOP 进行课程评价的效率较高。虽然完整版要求评分者的时间较多,但是现在在非教学指导情境下常使用的视频版对评分时间要求略低,视频版仅包含 22 个观察要点,熟练的评分者能够在 30 分钟左右完成所有评分,效率较高。

(二)UTOP 的局限

在实际使用过程中,UTOP 具有的以下局限性:

(1)适用科目有限。UTOP 设计的初衷是选拔优秀的数学和自然科学教师,其观察要点的设置也是以理学课程标准为基础,适用范围并不包括人文学科。虽然

最新版本的 UTOP 中新增了人文课程评价,但是其效果还有待进一步检验。

(2)初期时间投入较高。合格的 UTOP 系统的课堂观察者,需要经过严格的集中训练,时间投入较大。

(3)对课堂氛围的重视不足。UTOP 作为以数学和科学为主的课堂评价工具,特别在意促进理科思维生成的课堂教学手段的应用,但是对于课堂氛围等软环境的重视程度不足。

基于以上实施效果分析,UTOP 适用于数学和理科课程的课堂评价,有助于指导教师发展、评估教师培训效果、选拔优秀教师。

五、UTOP 对我国教育工作的启示

(一)UTOP 的诞生背景促发的思考

美国自然科学基金会划拨经费研发 UTOP 评价系统,起因是为了检验参加国家支助的 UTeach 中心开展的数学和科学教师培训的教师是否切实符合 STEM 的教学标准,并促进培训改进。我国自 2010 年开始实施“国培计划”,国家及各级政府都投入大量资源用于教师培训,虽然由于教师入职要求的变化,使得职前培训体系较弱。但是在职后培训方面,《教育部关于深化中小学教师培训模式改革全面提升培训质量的指导意见》(教师[2013]6号)中明确规定中小学按照年度公用经费预算总额的 5%安排教师培训,国培、省培计划如火如荼。这种状态其实与 UTOP 的开发背景极其相似。尤其是在帮助确认教师培训是否确实带来教师的专业发展和提升培训质量上,有效的课堂教学评估工具能够带来较大帮助。因此,考虑建立有实证证据的、易于操作的、有着统一标准的课堂评估工具,一方面能够使得培训有方向性,另一方面也能够使得参训者有明确的发展目标,使其发展能够明确指向国家期望的基于教育前沿理论的课堂教学方法。从而能够从根本上保证大量的教师培训投入能够切实带来教师课堂教学质量的提升。同时,通过教学评价发现和了解教师的实际问题,可以使教师培训有的放矢,提高培训效率。

(二)UTOP 的评估要点对课程改革的启示

UTOP 的观察要点主要基于对学习机会的研究和实践,重视建构主义的教学观,以及美国的科学和数学的学科标准,强调教学氛围和能够促进学生学科素养的教学方法。这与新课程标准所倡导的教育理念非常一致,都是指向未来发展的人才培养观,强调高度的科学文化素养和人文素养,着重训练基本的学习功能和基本的知识、技能和正确的价值观和态度。而 UTOP 这样的课堂评估工具的存在,将这样更为上位的课程标准转化为可操作、可检验的具体目标,从而对教师发展和课堂教学质量提升有实质性帮助。例如,在促进学生自主生成的部分,UTOP 强调学生在课堂中的参与、互动和对话,强调的是学生投注在学习任务上的时间比例,这与过去传统的强调教师讲授时间、效率有较大的区别,因为这些指标,让教师更能理解和操作以学生为中心的教育理念。而在课程结

构、执行效果、教学内容三个维度上,UTOP 强调了能够促进学生深度学习的教学结构、内容组织、课堂上的提问、时间分配等,充分展示了以学生为中心,以学科素养为目的的教育理念。因此,在推进新课标的课程改革的大潮中,UTOP 所关注的这些要点或许能给教育研究者、教研员和一线教师一定的启发和帮助。

(三)UTOP 的评估方法对课堂教学评估工具开发的启示

课堂教学质量是基础教育改革的核心,而课堂评估则是分析、评价、促进课堂教学质量的重要手段。目前,评估方式主要有以下三类:以学业成绩为导向的结果性评价、学生对教师的他评问卷测量和对课堂教学质量的观察评估。成绩导向的课堂评价缺乏对教学过程的分析,无法为教师提供有效的反馈和建议,且容易让教师产生不必要的压力,反而不利于提升学生的核心素养;学生视角下的他评问卷则受限于学生的主观感受,既难以形成统一的标准,又不能保证全面和专业。因此,建立以课堂观察为基础、指向提高课堂教学质量的过程性评价体系,是现阶段基础教育质量提升的必行之路。我国的课堂评价系统正处在快速发展中。由于众多学者的努力,国内的听评课方法在学校中的影响力正在不断增长。首先,课堂评价的方式已经从传统的听评课,演变为系统的教师课堂评价方法。华东师范大学提出的 LICC(Learning、Instruction、Curricular、Culture)课堂观察范式,^[26-27]确立了证据、技术与合作的课堂观察理念,同时界定了4维度、20视角、68观察点的问题域,走在了我国课堂评价的前沿。也有研究者尝试从“以学评教”的角度进行课堂教学评价指标设计,主要考察学生在课堂教学中所实际表现出来的学习行为状态,从针对性、能动性、多样性和选择性四个方面进行评价指标设计。^[28]课堂视频的量化分析工具也在我国积累了良好的研究经验,例如通过 Nvivo 软件分析课程教学视频。^[29]总体来说,国内的课堂评价的研究已经越来越接近国际前沿,而 UTOP 的指标体系和评估方法为国内的评估指标建设工作带来了进一步的反思,UTOP 的成功除了应用系统化的指标体系,还在于:(1)指标体系利于量化和标准化,为此,详细的评分标准和示例带来了巨大的帮助。(2)UTOP 的评分者培训也值得借鉴。(3)大量的效果验证和指标优化的实证研究也是 UTOP 保持活力的重要原因。

总之,UTOP 课堂评估体系的经验显示出对课堂评估指标的探索和建立是建设“促进公平,提升质量”的有效手段。而借鉴国外经验,加强实证研究能力,深入探讨“好课”标准,进一步发展中国的评课方式,将能快速而有效促进国内教学质量。

参考文献:

- [1] Kane,T.J. & Staiger, D.O. Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project [R]. Bill & Melinda Gates Foundation, 2012.
- [2] Walkington,C. & Marder,M. Classroom Observation and Value-Added Models Give Complementary

- Information About Quality of Mathematics Teaching [J]. *Designing Teacher Evaluation Systems*, 2013: 234–277.
- [3] Flanders, N.A. *Analyzing Teacher Behavior*[M]. Addison–Wesley P. C, 1970.
- [4] Bates, A. J. & Burbank, M. D. Effective Student Teacher Supervision in the Era of No Child Left Behind [J]. *The Professional Educator*, 2008, 32(2): 1–12.
- [5] Hinchey, P.H. *Getting Teacher Assessment Right: What Policymakers Can Learn from Research*[R]. National Education Policy Center, 2010.
- [6] Baker, E.L. & Barton, P. E., Darling–Hammond, L., et al. Problems with the Use of Student Test Scores to Evaluate Teachers. EPI Briefing Paper# 278[Z]. Economic Policy Institute, 2010: 1–27.
- [7] Walkington, C., Arora, P., Ihorn, S., et al. Development of the UTeach Observation Protocol: A Classroom Observation Instrument to Evaluate Mathematics and Science Teachers from the UTeach Preparation Program[D]. Unpublished thesis. Southern Methodist University, 2012.
- [8] Mo, Y., Singh, K. & Chang, M. Opportunity to Learn and Student Engagement: A HLM Study on Eighth Grade Science Achievement[J]. *Educational Research for Policy and Practice*, 2013, 12(1): 3–19.
- [9] Guiton, G. & Oakes, J. Opportunity to Learn and Conceptions of Educational Equality [J]. *Educational Evaluation and Policy Analysis*, 1995, 17(3): 323–336.
- [10] Hiebert, J. & Grouws, D. A. The Effects of Classroom Mathematics Teaching on Students’ Learning[J]. *Second Handbook of Research on Mathematics Teaching and Learning*, 2007(1): 371–404.
- [11] Gee, J. P. A Sociocultural Perspective on Opportunity to Learn[J]. *Assessment, equity, and opportunity to learn*, 2008: 76–108.
- [12] Smith, M.K., Jones, F.H., Gilbert, S. L., et al. The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices[J]. *CBE–Life Sciences Education*, 2013, 12(4): 618–627.
- [13] Kane, T. J., Mccaffrey, D. F., Miller, T., et al. Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project [Z]. Bill & Melinda Gates Foundation, 2013.
- [14] Goldring, E., Grissom, J. A., Rubin, M., et al. Make Room Value Added Principals’ Human Capital Decisions and the Emergence of Teacher Observation Data[J]. *Educational Researcher*, 2015, 44(2): 96–104.
- [15] Darling–Hammond, L. *Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching*[Z]. Center for American Progress, 2010.
- [16] Steinberg, M. P. & Sartain, L. Does Better Observation Make Better Teachers?[J]. *Education Next*, 2015, 15(1): 70–76.
- [17] Del Schalock, H., Schalock, M. D. & Ayres, R. Scaling up Research in Teacher Education New Demands on Theory, Measurement, and Design[J]. *Journal of Teacher Education*, 2006, 57(2): 102–119.
- [18] Ho, A.D. & Kane, T. J. The Reliability of Classroom Observations by School Personnel. Research Paper. MET Project[Z]. Bill & Melinda Gates Foundation, 2013.
- [19] Casabianca, J .M, Lockwood, J. & Mccaffrey, D.F. Trends in Classroom Observation Scores [J]. *Educational and Psychological Measurement*, 2015, 75(2): 311–337.
- [20] Kane, T., Kerr, K. & Pianta, R. *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*[M]. John Wiley & Sons, 2014.

Reflection of UTOP: A Classroom Observational Assessment Protocol in USA

CAO Hui & MAO Yaqing

(Faculty of Education, Beijing Normal University, Beijing, 100875, China)

Abstract: Effective classroom assessment standards will help teachers to find weaknesses of their lessons, improve the curriculum structure and teaching style, and enhance the quality of education. Here we introduced the classroom assessment system UTOP (UTeach Observation Protocol) from USA as an example to discuss the necessity of establishing classroom assessment tools. UTOP system is an important segment and instrument in a series of steps to solve the problem of lacking high-quality teachers in Mathematics and Sciences (STEM). We explicated UTOP's basic theories, assessment standards and rating process, and analyzed its practical effects. Meanwhile, we discussed its advantages and disadvantages in teaching assessment and its meaning to classroom assessment in China, hoping to provide a reference for theoretical and practical work in China.

Key words: quality of classroom education, teacher assessment, classroom observation

(责任校对:万明明)

(上接第 78 页)

- [28] Pianta, R. C. Classroom Management and Relationships Between Children and Teachers: Implications for Research and Practice[J]. Handbook of Classroom Management: Research, Practice, and Contemporary Issues, 2006, 8: 685-709.
- [29] Hamre, B. K. & Pianta, R. C. Early Teacher-child Relationships and the Trajectory of Children's School Outcomes Through Eighth Grade[J]. Child development, 2001, 72(2): 625-638.

The Development and Application of Value-added Measure in Teacher's Evaluation

ZHANG Yanan & DU Ping

(Faculty of Education, Beijing Normal University, Beijing 100875, China)

Abstract: The using of value-added method to evaluate the teacher effectiveness has drawn the attention of lots of scholars, for its bias and instability. To solve it, the United States has turned to the method of Multiple Measures, including multiple classroom observation. In this paper, firstly there is a discussion about the advantages and disadvantages of value-added measure to evaluate teachers' effectiveness, and then we show an improvement in American teachers' evaluation by introducing the classroom observation evaluation method and give a brief introduction of CLASS developed by the University of Virginia as an example. At last, we points out the enlightenment of reforming for teachers' evaluation system in China.

Key words: Value -added measure; Multiple Measures; Classroom observation; Teachers' evaluation

(责任校对:柯雅梅)