

## 教师增值性评价的进展及在我国应用的建议

边玉芳, 孙丽萍

(北京师范大学认知神经科学与学习国家重点实验室, 北京 100875)

**[摘要]** 教师增值性评价是一种起源于英美、目前已在一些国家和地区应用的教师评价新模式。这种评价模式借助统计模型, 将教师对学生成绩进步的单独贡献分离出来, 更加公平、精确, 具有引导教师关注全体学生、促进教师专业化发展、缓和生源大战及促进教师资源合理配置等方面的优势, 对促进我国教师评价改革有重要的参考价值。尽管增值性评价在理论研究和实践应用方面都尚存在一些问题, 但美国一些地区已开始将其应用于教师评价中, 教师增值性评价的应用价值正受到越来越多的关注。对我国来说, 应加强教师增值性评价的相关研究, 推动教师增值性评价在我国教师评价中的应用。当然, 教师增值性评价不能一步到位, 将教师效能与其他教师评价指标结合起来, 取长补短, 建立更为科学、合理、公平的教师评价体系, 是我国实施教师效能增值性评价的可取途径。

**[关键词]** 教师评价; 增值性评价; 教师效能

**[中图分类号]** G40-058.1 **[文献标识码]** A **[文章编号]** 1672-5905 (2015) 01-0000-00

### The Research Development of Teacher Value-added Evaluation and Its Practical Recommendation in Our Country

Bian Yufang, Sun Liping

(National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875, China)

**Abstract:** Teacher Value-added Assessment is a new pattern of teacher evaluation which was originated in American and England and now is being applied in some countries and areas. This evaluation method separates out the contribution of individual teachers on student achievement with the aid of some certain statistical models so that it makes evaluation more fair and accurate. Besides, it has superiority over the traditional form of teacher evaluation in guiding teachers to pay attention to all the students, promoting teachers' professional development, alleviating war of competing students between schools and allocating teacher resources rationally. Therefore, it has significant referential value in promoting teacher evaluation reform in our country. Even though value-added assessment has some problems in both theoretical research and practical application, some districts in the United States have begun to apply value-added assessment in teacher evaluation. The worth of teacher value-added assessment has drawn more and more attention. Hence, for the present stage in our country, in order to accelerating the pace of practical application, related research should be emphasized. Teacher value-added assessment can not be accomplished in one step. Combination teacher effect with other teacher evaluation index in order to establish a more scientific, reasonable, and fair teacher evaluation system is a practical approach of implementing teacher value-added assessment in China.

**Key Words:** teacher evaluation; value-added assessment; teacher effect

---

**[收稿日期]** 2014-10-10

**[基金项目]** 2010 年度教育部人文社会科学研究规划基金项目 (10YJAZH003) 研究成果之一

**[作者简介]** 边玉芳 (1967—), 女, 浙江诸暨人, 北京师范大学教授, 博士, 主要研究方向为教育评价、心理测量、心理健康。

百年大计，教育为本；教育大计，教师为本。教师是教育教学的实施者，是教育活动的直接参与者，对学生的发展有非常重要的影响。科学有效的教师评价可合理引导教师工作方向，提高教师工作积极性、工作满意度和职业幸福感，从而间接促进学生发展和教育质量提升。建立科学合理的教师评价体系无论对于教师群体本身还是学生、乃至整个国家教育质量都至关重要。

《国务院关于加强教师队伍建设的意见》要求“健全教师考核评价制度……严禁简单用升学率和考试成绩评价中小学教师”。我国传统一般以升学率或平均分对教师进行评价，这种评价方式存在很多弊端。首先，这种评价方式是不公平的。若教师所教的学生入学成绩较好、能力较强，即使教师的教学工作不是很突出，也能取得相对高的升学率；而若教师所教学生能力水平较差，即使教师本身的教学工作很出色，学生虽然取得了进步但可能仍然无法达到升学率的要求。由此，教师评价往往不能起到引导教师改进教学的作用，反而使得学校千方百计争取好学生、教师想方设法进入好学校，加剧了生源大战和师资分布不均问题。其次，这种评价方式也是不准确的。学生的学业成绩受到诸多因素的影响，如学生自身能力水平、家庭环境、班级环境、甚至社区环境等，升学率、平均分是这些影响因素共同作用的结果，无法准确衡量教师的单独贡献。可见，传统教师评价模式相对片面，有失公允，也不够准确，无法科学评估教师的教学质量。

在英美等国家，一种有别于传统教育评价的评价模式——“增值性评价”(value-added evaluation)受到很多研究者和教育实践者的关注，教师增值性评价的理念开始被一些国家和地区应用到教师评价中。

## 一、 教师增值性评价的内涵与优势

### (一) 内涵

增值性评价最早起源于经济学中增值(value-added)的概念，它要求在评估“产出”(output)时考虑“投入”(input)的多少，降低成本，提高收益，追求收益即增值最大化。

教师增值性评价是基于一个特殊指标——教师效能(teacher effect)，对教师进行评价的一种评价模式。所谓教师效能是指在对教师进行评估时，通过追踪教师所教学生在一段时间内学业成绩的变化，运用科学的统计模型和分析方法排除对学生成绩有影响但不受教师控制的因素（如学生的人口学因素、原有成绩水平、家庭背景等），教师因素所带来的学生学业成绩的变化，即教师对学生成绩增值的“净效应”。<sup>[1-3]</sup>

教师增值性评价有别于传统教师评价主要在于两点：一是“增值”，即关注学生成绩的进步，而不仅仅关注学生的最终成绩；二是“净效应”，即将教师的贡献与其他影响学生学业成绩的因素区分开，对教师单纯的贡献加以评价。这两点正是教师增值性评价的优势所在，保证了评价的公平性和科学性。

### (二) 优势

教师增值性评价内涵的特殊性决定了教师增值性评价相比于传统教师评价的优越性，具体来说，主要体现在以下几个方面。

#### 1. 促进教师评价的公平性

从理论上讲，若要对教师进行绝对公平的评价，需将所有学生随机分配，保证每个教师所教学生的能力是完全一样的，并且所有学生没有接受家长辅导、课外学习等其他非教学因

素的影响，此时学生的学业成绩完全是教师教学的结果，代表了教师真实的教学水平。在这种理想情况下，通过学生的平均分、升学率等对教师进行评价才是公平的。而实际上，学生不可能完全随机分配，由于学生择校、智力能力差异等原因，不可避免地存在生源质量差异的问题；且家庭条件、父母教育观念等的差异会使得学生在学习之外获得不同的学习、教育机会。所以，学生的学业成绩是教师、学生自身、家庭、社会等诸多因素综合作用的结果，基于此对教师进行评价是不准确的，也是不公平的。

教师增值性评价考虑到学生的起始能力水平的差异，基于学生成绩的进步对教师进行评价，这就在一定程度上解决了学生非随机分配的问题，弱化生源质量差异带来的影响。同时，在模型设计合理的情况下，可以将家庭、社会等因素的影响加以排除，分离出教师对学生学业进步的单独贡献，获得教师的“净”效应。

综上所述，教师增值性评价在促进教师评价的公平性方面有着传统教师评价所无法比拟的优越性。

## 2. 引导教师关注所有学生

升学率、合格率等指标关注的是学生的最终学业成绩，不考虑学生最初能力水平的差异，采取“一刀切”的统一划线方式。在这种评价模式下，教师容易将教学关注点放在成绩中等及以上的学生身上，因为这部分学生更容易取得升学的资格或达到合格的标准，而忽视对后进生的辅导和帮助。

教师增值性评价基于学生成绩的增值来对教师进行考评，不设置统一的划线标准，不关心学生的起始能力水平如何，只关注学生经过教师教学之后的进步情况。这就给教师一个更为科学的引导——关注所有学生，因为不仅好学生可以取得进步，后进生经过有效教学也可以取得进步，甚至进步空间更大。

可见，教师增值性评价在引导教师关注所有学生、激励教师根据不同学生的需求采取相应的教学措施、促进每个学生的进步等方面具有传统教师评价所不具备的优势。

## 3. 有利于学校教师队伍建设和教师自身专业化发展

教师增值性评价基于教师效能对教师进行评价，通过统计分析可获得每一个教师的效能值，进而对不同教师的效能值加以比较，区分出高效能教师与低效能教师，这在实践上具有重要价值。对学校而言，通过探讨高、低效能教师的特征差异，可帮助学校在教师招聘中有针对性地选择有潜能的高效能教师，同时学校可更有针对性、更有效地开展教师培训，对已入职的教师进行专业培训，提高所有教师的效能，建设一只高水平的教师队伍。对教师自身而言，可帮助教师以高效能教师为学习对象，督促教师进行自我反思、自我改进，帮助教师明确个人专业发展需要，制定未来发展目标，促进教师自身的专业化发展。

## 4. 缓和生源大战，促进区域内师资合理分布和区域教育质量均衡发展

教师评价的根本目的，简单说来就是通过科学评估教师工作业绩，让教师明确自己工作的利弊得失，进而有针对性地进行教学改进，从而更好地发挥教书育人的作用。然而，在以升学率、平均分为指标的教师评价模式下，抢得好生源往往就意味着评价的胜出。为了在最后的评估中占据优势，学校之间甚至教师之间展开激烈的生源大战，教师评价的改进目的被忽视。不宁唯是，传统的平均分、升学率评价模式还会导致师资分布不均衡。在同等教学的情况下，生源质量较好的学校往往升学率更高，故好学校的教师在评价中更占优势，教师在应聘时会更倾向于选择生源质量好的学校，于是好学校师资质量更优、数量更多，而师资质量和数量的差异又进一步拉大了学校间的升学率差异，升学率差异加大又带来新一轮师资不均衡，形成恶性循环。

相比之下，教师增值性评价关注学生的进步情况，生源质量与教师评价结果并无直接关系。无论生源好坏，只要教师用心教学、采取科学有效的教学方法，就能使学生取得进步，

进而在评价中取得好成绩。这就在一定程度上缓和了生源大战，促进了区域内师资的合理分布，师资合理分布又在一定程度上促进了区域教育质量的均衡发展。

## 二、 教师增值性评价的发展现状与面临挑战

### （一）发展现状

#### 1. 国际研究现状

国际上关于教师增值性评价的研究可归纳为三个方面：一是对教师增值性评价纯理论的探讨，主要是模型研究，目的是实现对教师效能的更准确、有效估计，为应用奠定理论基础；二是对教师增值性评价价值的探讨，一般是实证研究，目的是通过实证数据证明教师增值性评价在应用上的价值及可行性；三是教师增值性评价的应用实践，为教师增值性评价的广泛应用提供经验。实际上，这三方面正是教师增值性评价从理论到实践的发展步骤，只有这三步都走好，教师增值性评价才能真正地应用于教师评价的实践中。

##### （1）理论模型

教师增值性评价依赖于一类特定的模型——增值模型(Value-Added Models)，理论研究主要是对模型的探索。早期主要的增值模型有：获得分数模型(Gain Score Model)、协变量校正模型(Covariate Adjustment Model)、田纳西模型(Tennessee Value-Added Assessment Model)、交叉分类模型(Cross-classified Model)等。这几种模型是比较成熟的、认可度较高的模型，且有些模型已被一些地区应用于评价中，如达拉斯市采用协变量校正模型对学校效能加以评估、<sup>[4]</sup>田纳西州建立了基于田纳西模型的教育增值评价系统。<sup>[5]</sup>

近年来，随着增值性评价实证研究的增多，模型不能很好地拟合真实数据成为研究中的一个突出问题，经典模型开始受到越来越多的攻击和批判，探索更优、更拟合真实数据、估计结果更精准的新模型成为一些研究者的关注重点。在典型模型的基础上，出现了很多模型变式，<sup>[6-8]</sup>教师效能在模型中从一个单维的值发展到多维的几个值，从一年效能发展到可实现对教师连续几年效能值的估计，<sup>[8, 9]</sup>模型估计方法也从早期的最小二乘估计发展到贝叶斯估计。<sup>[10]</sup>总之，模型越来越灵活，越来越能处理真实、复杂数据。

但这些新模型作为新的尝试和探索，还尚未像经典模型那样得到广泛认可，各个新模型的优劣也大多是百家之言、各执一词。由此，对不同模型之间关系的探讨成为研究的热点之一，不同的研究者得出了不同的研究结论。有研究发现采用不同模型对估计结果的影响不大，模型选择、控制变量的选择对教师效能估计结果的影响很小。<sup>[11, 12]</sup>但也有研究者持相反意见，认为教师效能估计结果对模型非常敏感，采用不同模型可能导致不同的估计结果，但并未证明哪种模型更优。<sup>[7, 13]</sup>

##### （2）价值探讨

早期教师增值性评价研究证明了教师对学生学业表现的重要影响。Sanders 等人通过一系列研究，指出教师是影响学生学习最重要的因素，<sup>[2, 3, 14]</sup>此后很多研究者均使用增值方法证明了教师在促进学生学业进步方面的作用。<sup>[15-19]</sup>早期的这些研究推动了美国一些州教育评价体系的改革，拉开了教师增值性评价的应用序幕。

探讨高效能教师的特征曾一度甚至仍是当前的研究热点之一，为教师增值性评价如何指导雇佣、报偿等实践工作提供了一定的启发。有研究表明，教龄高的教师能有效提高学生的阅读成绩，十年的教学经验可相应提高学生词汇和阅读理解成绩大约0.15和0.18个标准差，这提示学校在师资配置时，应分配教龄高的教师进行阅读教学。<sup>[20]</sup>还有研究者通过对10000名澳大利亚教师的教师效能探讨也发现类似的结论，高效能的教师拥有更多教学经验，且在

阅读表现高效能的教师往往是女教师。<sup>[21]</sup>一项对芝加哥公立高中教师效能研究发现,拥有硕士学历的教师比拥有博士学历的教师效能更高,这提示学校在招聘新教师时或许可以多考虑硕士学历的教师。<sup>[22]</sup>

近年来,关于教师增值性评价的有效性研究成为一个研究热点,这是对教师增值性评价能否用于评价实践的直接探讨。研究者分为明显的两派。反对派认为,教师效能估计存在偏差,教师增值性评价研究还有很多问题尚未解决,不应将其应用于教师评价及相关的教师雇佣、教师工资制定等政策中。<sup>[23, 24]</sup>而支持派认为虽然教师效能估计存在一定的偏差,但这种偏差并不大,总体上教师增值性评价的结果是可信的,相较于其他教师评价指标而言,教师增值性评价更为科学、合理。<sup>[25, 26]</sup>研究者使用实证数据证明了教师增值性评价结果的可信性,支持美国佛罗里达州基于教师增值性评价的教师解聘政策。<sup>[27]</sup>

### (3) 应用实例

在应用实践方面,最早将教师增值性评价应用于教师评价实践中的是美国。美国一些地区包括田纳西州、达拉斯市等都开始使用教师增值性评价对教师进行评估,并把评估结果应用于教师工资制定、促进教师专业化发展等方面。

田纳西州增值评价系统(the Tennessee Value-Added Assessment System, TVAAS)是发展最早也是目前最完善的教育增值评价体系。早在 20 世纪 90 年代, TVAAS 就开始对学校及地区进行评估,其中也会涉及教师评价,但仅仅是简单地报告教师效能值,尚未真正应用于教师评价。后来,随着 TVAAS 的进一步完善,田纳西州通过新的教师评价立法,建立了新的教师评价体系,并于 2011-2012 学年投入使用。新教师评价体系规定,田纳西州每年对每个教师都要进行评估,评估包括三个方面——班级观察、学生学业表现以及教师增值分,其中,教师增值分数占到教师评估总分的 35%。基于评估分数将教师分类(五类——显著低于期望值、低于期望值、符合期望值、高于期望值、显著高于期望值),并提供详细的用于指导改进的反馈报告,评估结果可用于人事决策。<sup>[28]</sup>

达拉斯增值问责体系(the Dallas Value-Added Accountability System, DVAAS)建立于 1992 年。在发展早期, DVAAS 使用教师效能指数(Teacher Effectiveness Indices)界定高效能教师及需要帮助的低效能教师,虽然该指数仅用于学校对本校教师的内部评价,并不用于全市范围的教师评估及相关决策,但推动了基于教师效能对教师进行评价的应用实践。<sup>[4]</sup>今年达拉斯市宣布将在 2014-2015 学年启动教师卓越计划(Teacher Excellence Initiative),更好地推动教师专业化发展,并计划在 2015-2016 学年启动新的教师工资系统,教师工资不再基于传统的教龄等指标,而是更多地与教师表现挂钩。<sup>[29]</sup>

田纳西州和达拉斯市是增值性评价开始较早的地区,也是教师增值性评价发展较快的地区。除此之外,目前纽约、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增值性评价应用于教师评价中去。

## 2. 国内研究现状

国内关于教师增值性评价的研究进展较缓慢,理论研究不多,大多是对国外教师增值性评价研究现状的述评。周燕、边玉芳对美国教师增值性评价的起源、发展、实施效果及争议进行了论述。<sup>[30]</sup>徐士强和周燕等人分别对美国田纳西州教育增值评价体系及对我国的启示进行了述评。<sup>[31, 32]</sup>邓森碧、边玉芳基于模型选择对教师增值性评价结果的重要意义,对主要增值模型进行了比较,为实际应用中增值模型的选择提供了一定的依据。<sup>[33]</sup>国内也有研究者使用实证数据进行了相关研究。张文静、辛涛使用增值模型对房山区小学四年级学生进行了研究,探讨了对学生数学成绩有显著影响和无显著影响的教师特征变量。<sup>[34]</sup>凡细珍、任杰使用增值模型对新疆双语教学小学五年级汉语学科进行了增值性评价,报告了汉语教师在汉语全卷、听力和阅读上的增值分。<sup>[35]</sup>

## （二）面临的挑战

虽然相比于平均分、升学率等绝对指标，教师增值性评价更能体现公平性、科学性，但教师增值性评价还有一些尚存争议之处，使得教师增值性的实践应用面临挑战和质疑。

### 1. 评价指标是否有效且可信

教师增值性评价研究的一个争议焦点是评价指标是否有效且可信，即基于增值模型的教师效能值是否真实代表了教师对学生学业进步的贡献。

有研究者从理论的角度出发，认为虽然增值模型号称可区分出教师的单独贡献，但学校教育作为一个系统工程，是所有教育资源整合、共同发挥作用的结果，学生学业的进步到底是归于教师还是归于学校、这二者之间如何区分是无法完全划分清楚的，因此建议决策者在使用教师效能指标对教师进行评价时应慎重考虑这个问题。<sup>[36]</sup>

此外，教师效能这个评价指标本身的一些统计特性也存在争议。有研究者通过实证研究证明教师效能是对教师贡献的无偏估计，可以真实代表教师对学生学业进步的作用，<sup>[37]</sup>认为采用教师效能对教师进行评价可促进学生学业的进步，并指出华盛顿、洛杉矶等一些地区的学校都开始使用教师增值性评价。<sup>[38]</sup>但反对者认为教师效能这个指标存在偏差，在应用中应慎重，<sup>[23, 24]</sup>采用不同模型将导致对教师效能的高估或者低估。<sup>[39]</sup>有研究者对斯坦福数学评定测验(the Stanford 9 mathematics assessment)的两个维度程序(procedures)和问题解决(problem-solving)分别赋予不同的权重，得到构念不同的多组测验，发现教师效能在不同构念的测验之间存在一定差异。<sup>[11]</sup>另有研究者对同一批学生在同一时间、同一学科进行测验也得到类似的结论，发现采用不同的测验会导致同一个教师的效能值发生改变。<sup>[40]</sup>反对者认为既然采用不同测验、不同模型将导致不同的教师效能估计结果，那么教师效能这个评价指标是否可信便值得怀疑。

### 2. 评价结果是否被公众认可

由于教师效能的估计基于复杂的统计模型，它不像平均分、升学率那样简单、清楚、明了，教师效能的含义是什么、如何得来的、是否可信，这对公众来说是很难理解的。以获得分数模型为例，教师效能是教师层的残差，为什么教师层残差代表教师效能，为什么这个残差大代表教师是高效能的、残差值小代表教师是低效能的，公众不得而知，所以公众对这种评价方法不容易建立信任感。因此，如何增加教师效能对公众的透明度、使得评价结果更易于公众理解和接受，是教师增值性评价应用的一个重要问题。但有研究者认为这个问题是可以解决的，从项目反应理论(Item Response Theory, IRT)的发展来看，IRT 在发展初期也遇到了类似的问题，人们不理解这个能力 $\theta$ 是如何得出的、代表什么意义，但随着理论的发展，虽然 $\theta$ 的获得对公众来说仍是“黑箱”，但是，越来越多的人开始认可 IRT 对考生能力估计的准确性，IRT 在实际中尤其是大规模测验中的应用越来越多。<sup>[39]</sup>或许，教师效能在未来会像 IRT 一样推广起来，但在现阶段，评价结果的透明度问题仍是教师增值性评价应用于实践的一个阻碍。

### 3. 基于学业测验的评价方式是否有利于学生的全面发展

教师增值性评价是基于学生的学业成绩进行的评价，学业测验的测验构念(construct)问题引起了一些研究者的关注。<sup>[6, 19, 41]</sup>有研究者认为，为提高估计精度，在测验设计时应偏重客观题，但一些学科如历史、语文等采用客观题可能无法测得学生在该学科的关键能力，这就带来了一定的问题。<sup>[40]</sup>还有研究者指出，为提高估计精度、确保不同年级的测验成绩可比（垂直等值），在测验设计时应侧重在同一个垂直量尺上的内容即在年级间连续教授的内容，<sup>[41]</sup>同样地，这些内容可能并不能有效考察学生的关键能力。也就是说，为保证模型估计结果的准确性，教师增值性评价势必会牺牲测验内容的合理性和有效性。倘若测验测的并

非学生需要掌握的重点知识和需要培养的关键能力，那以此评价教师又有何价值？

退一步讲，即使测验构念合理，测验可有效考察学生的学业表现。但是，对学生来说，除了掌握知识和技能，更重要的是发展各方面的能力。教育的目标不仅仅是让学生获得一个好成绩，教育的最终目标是培养一个全方位发展的人才。而学业测验仅能测查学生的学业水平，许多对学生个人发展非常重要的、学校教育应重视培养的能力无法通过学业测验来测得，<sup>[39]</sup>如学生的领导力、人际交往能力、团队合作精神等。因此，有研究者质疑：基于学业测验的教师增值性评价真的有利于促进学生的全面发展还是仅仅能帮助学生提高学业成绩？在这个问题上，也许将学业测验扩展到其他能力测验会是一条解决之路，但在现阶段可能尚无法给出满意的回答。

### 三、 对我国应用教师增值性评价的建议

#### （一）增值性评价是一种相对客观、公正、科学的评价模式，应该成为新时期我国教师评价的重要方式

虽然教师增值性评价无论在理论研究还是实际应用中都存在一些问题，但毋庸置疑，相比于传统教师评价，教师增值性评价更加科学、公平、合理。教师增值性评价正受到越来越多研究者及政策制定者的关注。美国一些地区（如田纳西、俄亥俄、科罗拉多、佛罗里达等）及英国、荷兰等已开始尝试将增值性评价应用于教师评价中，并逐渐与教师雇佣、报偿、晋升、分配等政策法规相挂钩。

在我国，教师评价尤其是中小学教师评价以升学率、平均分等指标为主，既缺乏公平合理性，又易带来生源大战、师资分布不均等问题，受到越来越多的质疑和批判。为解决我国当前教师评价存在的固有问题，建议将教师增值性评价引入到我国教师评价体系中去，这是解决传统教师评价弊端的一个重要途径，也是我国未来教师评价发展的有效之路。

#### （二）加强教师增值性评价的理论研究与应用探索，推动我国教师评价体系改革

目前我国国内教师增值性评价研究进展相对缓慢，这直接制约了教师增值性评价在我国的应用。为推动教师增值性评价在实践中的应用，应切实加强相关研究，为教师增值性评价应用奠定基础。

首先，应加强教师增值性评价的理论研究，包括前面所讲到的模型问题、估计偏差问题等，还有如何设计内容上符合中国课标要求、性质上满足增值模型需要的学生学业测验，以及对我国学生学业成绩影响较大的学生自身、教师、家庭等诸多因素有哪些，如何在模型中加入这些影响因素，都是值得探讨、解决的问题。

除了理论研究，实践探索也是必不可少的。由于目前我国教师增值性评价的应用经验尚属空白，建议首先在条件允许的地区（如北京、上海）进行小范围应用尝试，探索、总结出我国区域实施教师增值性评价的实践经验。在小范围试验成熟的条件下，可进一步推广应用，进而在全国范围内实施教师增值性评价，在此基础上推动我国教师评价体系改革。

### （三）借助科研机构的力量，推动教师增值性评价的实施

实施教师增值性评价对技术的要求较高，建议借助大学、研究所等专业研究机构的力量，加速推动教师增值性评价的应用实践。

应用教师增值性评价的技术难点主要有以下几个方面。首先，增值性评价依赖于多次测验成绩的有效测量。Young 指出，大多数增值性评价需要一个垂直化的分数量尺，以便将学生在连续几个学年的成绩加以比较，故用于增值性评价的测验需要经过很好的垂直等值设计。如何进行垂直等值、如何在垂直等值的基础上合理设计测验是一个主要的技术难题。<sup>[42]</sup>其次，增值性评价需将学生自身、家庭因素等其他环境因素与教师区分开来，那么在对环境因素的考察中，考察哪些因素、如何设计测评工具也是相对专业且难操作的问题。此外，像增值模型选择、模型构建、模型估计、结果分析等都对技术的要求较高。

因此，建议各地区在实施教师增值性评价时与大学或其他科研机构建立合作关系，借助专业机构的力量完成教师增值性评价的前期设计与后期结果分析等相关工作。当然，在合作一段时间以后，各地区已具备独立实施条件和能力的情况下，各地区可逐步脱离科研机构的帮助，独立实施教师增值性评价。

### （四）将教师效能指标与传统教师评价指标结合，逐步推进教师增值性评价

教师增值性评价还有一些问题需要澄清和解决，在现阶段，教师评价尚不能完全背离传统的评价模式而完全采用教师增值性评价。

关于教师增值性评价的应用，目前一个比较统一的观点是，将教师增值性评价作为教师评价的指标之一，纳入到教师评价体系中去。美国俄亥俄州在学校评价中采用了这种方式，评价指标包括毕业率、出勤率、NCLB(No Child Left Behind)年度进步情况、基于所有学科测验的学业表现指数、以及学校效能值。<sup>[39]</sup>

建议我国各地区在实施教师增值性评价时，把教师增值性评价引入到当前教师评价体系中去，将教师效能与其他教师评价指标结合起来，取长补短，建立更为科学、合理、公平的教师评价体系。

对于一些已经具备实施教师增值性评价的地区，应在教师评价中积极引入教师效能指标，探索包含教师增值性评价的教师评价新模式，合理确定教师增值性评价在教师评价中的权重，为其他地区工作提供借鉴和指导。而对那些尚不具备实施条件的地区，可将教师增值性评价列入发展目标，借鉴其他地区的工作经验，逐步推进，慢慢摸索本地区的教师评价模式。

## [参考文献]

- [1] Rivkin S.G., E.A. Hanushek, and J.F. Kain. Teachers, schools, and academic achievement[J]. *Econometrica*, 2005, **73**(2): 417-458.
- [2] University of Tennessee Value-Added Research and Assessment Center. Cumulative and residual effects of teachers on future student academic achievement[R]. Knoxville, TN, 1996.
- [3] Sanders W.L., S.P. Wright, and S.P. Horn. Teacher and classroom context effects on student achievement: Implications for teacher evaluation[J]. *Journal of Personnel Evaluation in Education*, 1997, **11**(1): 57-67.
- [4] Jason Millman. Grading teachers, grading schools: Is student achievement a valid evaluation measure[M]. Corwin Press Inc, 1997.
- [5] Sanders, W.L. and S.P. Horn. The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment[J]. *Journal of Personnel Evaluation in Education*, 1994, **8**(3): 299-311.
- [6] Martineau J.A. Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability[J]. *Journal of Educational and Behavioral Statistics*, 2006, **31**(1): 35-62.
- [7] Sass T.R., A. Semykina, and D.N. Harris. Value-added models and the measurement of teacher productivity[J]. *Economics of Education Review*, 2014, **38**: 9-23.
- [8] Mariano L.T., D.F. McCaffrey, and J. Lockwood. A model for teacher effects from longitudinal data without assuming vertical scaling[J]. *Journal of Educational and Behavioral Statistics*, 2010, **35**(3): 253-279.
- [9] McCaffrey D.F., et al. Models for value-added modeling of teacher effects[J]. *Journal of Educational and Behavioral Statistics*, 2004, **29**(1): 67-101.
- [10] Lockwood J., et al. Bayesian methods for scalable multivariate value-added assessment[J]. *Journal of Educational and Behavioral Statistics*, 2007, **32**(2): 125-150.
- [11] Lockwood J., et al. The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures[J]. *Journal of Educational Measurement*, 2007, **44**(1): 47-67.
- [12] Kersting N.B., M.-K. Chen, and J.W. Stigler. Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores[J]. *Education Policy Analysis Archives*, 2013, **21**: 7.
- [13] Newton X.A., et al. Value-Added Modeling of Teacher Effectiveness: An Exploration of Stability across Models and Contexts[J]. *Education Policy Analysis Archives*, 2010, **18**(23): 23.
- [14] Sanders W.L. and S.P. Horn. Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research[J]. *Journal of Personnel Evaluation in Education*, 1998, **12**(3): 247-256.
- [15] Rowan B., R. Correnti, and R. Miller. What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools[J]. *The Teachers College Record*, 2002, **104**(8): 1525-1567.
- [16] Rothstein J. Teacher quality in educational production: Tracking, decay, and student achievement[J]. *The Quarterly Journal of Economics*, 2010, **125**(1): 175-214.
- [17] Lefgren L. and D. Sims. Using subject test scores efficiently to predict teacher value-added[J]. *Educational Evaluation and Policy Analysis*, 2012, **34**(1): 109-121.
- [18] Mendro R., et al. An application of multiple linear regression in determining longitudinal teacher effectiveness. in Annual Meeting of the AERA, San Diego, CA. 1998.
- [19] McCaffrey D.F., et al. Evaluating Value-Added Models for Teacher Accountability[M]. California : RAND Corporation, 2003.
- [20] Rockoff J.E. The impact of individual teachers on student achievement: Evidence from panel data[J]. *The*

- American Economic Review, 2004, **94**(2): 247-252.
- [21]Leigh A. Estimating teacher effectiveness from two-year changes in students' test scores[J]. Economics of Education Review, 2010, **29**(3): 480-488.
- [22] Aaronson D., L. Barrow, and W. Sander. Teachers and student achievement in the Chicago public high schools[J]. Journal of Labor Economics, 2007, **25**(1): 95-135.
- [23]Rothstein J. Student sorting and bias in value-added estimation: Selection on observables and unobservables[J]. Education, 2009, **4**(4): 537-571.
- [24]National Bureau of Economic Research. Teacher quality in educational production: Tracking, decay, and student achievement[R]. North Carolina, 2008.
- [25]Center for Education Data & Research. Assessing the "Rothstein Falsification Test" Does it Really Show Teacher Value-Added Models Are Biased?[R]. Seattle, 2012.
- [26]Koedel C. and J.R. Betts. Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique[J]. Education Finance and Policy, 2011, **6**(1): 18-42.
- [27]Winters M.A. and J.M. Cowen. Who would stay, who would be dismissed? An empirical consideration of value-added teacher retention policies[J]. Educational Researcher, 2013, **20**: 1-8.
- [28]Tennessee Department of Education. Teacher Evaluation in Tennessee: A Report on Year 1 Implementation 2012[R]. Tennessee, 2012
- [29]TAWNELL D. HOBBS. Proposed Dallas ISD teacher evaluation system based on merit[N]. The Dallas Morning News, 2014-5-19.
- [30]周燕, 边玉芳. 美国教师效能增值评价研究与应用进展[J]. 全球教育展望, 2011, **40**(10): 72-78.
- [31]徐士强, 赵风波. 美国田纳西州教育增值评价模式及其论争[J]. 全球教育展望, 2009,(9): 52-55.
- [32]周燕, 边玉芳. 美国 TVAAS 的解读及其对我国教育评价的启示[J]. 全球教育展望, 2012, (3): 11.
- [33]邓森碧, 边玉芳. 教师效能增值模型的研究与应用[J]. 教育学报, 2012,(4): 113-121.
- [34]张文静, 辛涛, 康春花. 教师变量对小学四年级数学成绩的影响: 一个增值性研究[J]. 教育学报, 2010,(2): 69-76.
- [35]凡细珍, 任杰. 增值理念下的新疆双语教学质量评价——对小学五年级汉语学科的实证研究. 中国考试, 2013,(10): 17-23.
- [36]Corcoran S.P. Can Teachers Be Evaluated by Their Students' Test Scores? Should They Be? The Use of Value-Added Measures of Teacher Effectiveness in Policy and Practice. Education Policy for Action Series. Annenberg Institute for School Reform at Brown University (NJ1), 2010.
- [37]National Bureau of Economic Research. Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates[R]. North Carolina, 2013.
- [38]Gordon R.J., T.J. Kane, and D. Staiger. Identifying effective teachers using performance on the job[M]. Washington, DC: Brookings Institution, 2006.
- [39]Chudowsky N., H.I. Braun, and J.A. Koenig. Getting value out of value-added: Report of a workshop[M]. Washington, DC: National Academy Press, 2010.
- [40]Corcoran S.P., J.L. Jennings, and A.A. Beveridge. Teacher Effectiveness on High-and Low-Stakes Tests. Society for Research on Educational Effectiveness, 2011.
- [41]Schmidt W.H., R.T. Houang, and C.C. McKnight. Value-added research: Right idea but wrong solution. Value added models in education: Theory and applications, 2005: 145-164.
- [42]Young M.J. Vertical scales. In S.M. Downing and T.M. Haladyna (Eds.), Handbook of test development. Lawrence Erlbaum Associates Publishers, 2006.