

【教师管理】

探索增值性评价:教师绩效考核改革的突破口

赵德成

【摘要】增值性评价在控制影响学生成绩有关因素的前提下,分析教师对学生成绩变化的贡献程度,这种评价方法可在很大程度上提高教师绩效考核的客观性与公平性,有力引导教师关注和改善工作结果,因此将在未来教师人力资源管理实践中发挥重要作用。在分析美国哥伦比亚学区相关评价经验的基础上,建议我国在推进教师增值性评价中要做到:修订有关政策,为增值性评价实施创造数据条件;做好评价数据库建设,全面收集相关数据;关注测验效度,确保数据质量;积极稳妥地推动有关改革,防止急躁冒进。

【关键词】增值性评价;教师评价;绩效考核;教师绩效工资;增值模型;美国哥伦比亚学区

【作者简介】赵德成,北京师范大学教育学部(北京 100875)。

【原文出处】摘自《中小学管理》(京),2020.10.16~20

【基金项目】本文系北京市教育科学“十三五”规划重点课题“中小学学校治理现代化:问题、原因分析与改进”(课题批准号:CCAA2020037)的研究成果之一。

最近,中央深改委审议通过的《深化新时代教育评价改革总体方案》明确强调,各类教育评价要积极“探索增值评价”。在这样的政策背景下,增值性评价逐渐成为我国教育管理领域的一个热点问题,如何在教师绩效考核中有效应用增值性评价也备受关注。

一、将增值性评价引入教师绩效考核有何现实意义?

增值性评价在控制影响学生成绩有关因素的前提下,将一名教师在一个学年中所教学生的测验分数,与这些学生在上一学年或上几学年的分数以及同一年级其他学生的分数进行比较,从而分析教师对学生成绩变化的贡献程度。这种评价方法起源于20世纪70年代,进入21世纪以来,随着计算机技术的发展及数据库建设的普及,加之教育问责制度的推动,在美英国家教育管理实践中的应用越来越广泛。相对于传统的教师绩效考核方法,增值性评价具有一些明显优势,将成为未来教师绩效考核改革的突破口。其优势可归纳为以下几点。

1. 克服传统评价方法的主观性弊端

近些年来,国内各地市都在努力探索教师绩效考核的机制与方法,多数中小学遵循“定性 with 定量相

结合”“教师自评与学科组评议、年级组评议、考核组评议相结合”以及“形成性评价和阶段性评价相结合”原则,^[1]采用多主体多元化的评价方法。这种考核思路借鉴了工商管理领域常用的360度考核法,克服了评价自上而下、教师被动接受、领导说了算的制度弊端,在一定程度上体现了管理民主,避免了“晕轮效应”“个人偏见”与“考核盲点”等实际问题,较好地反映了教师的工作绩效。但必须承认的是,这种多主体评价仍然不能有效克服评价中的主观性,有人会在评价中凭个人感觉、偏好甚至恩怨评分。更普遍的情况是,在教师群体中,很多人本着“你好、我好、大家好”的“好人主义”给自己和同事都打高分。

相较于多主体评价,增值性评价对教师绩效的考核更加客观、准确和有效,可成为传统评价的补充办法。增值性评价在控制了非教师因素的前提下,用分数增长变化的幅度说话,能够有效克服传统评价的主观性。

2. 引导教师关注和不断改进工作结果

教师绩效考核主要是评价教师依法执教以及完成学校规定岗位职责和工作任务的情况。而在实践中绩效考核评什么,与管理者对绩效的理解有很大关系。对绩效的理解不同,绩效考核的重点和方法

也就不同。在学术界和公众眼中,人们对绩效的理解主要有三种。^[2]第一种观点认为绩效是结果,是在特定时间内、由特定的工作职能或活动所产生的产出记录;第二种观点认为绩效是行为,是员工与工作相关的实际表现;第三种观点则将前两种观点整合起来,认为绩效是过程与结果的结合,这种整合、平衡的观点在学术界与实践中得到更为广泛的认同,应该成为指导教师绩效考核实践的重要依据。

目前,在中小学教师绩效考核实践中,既存在过分强调结果特别是把中高考成绩与升学率作为重要考核指标的问题,又存在片面关注过程、忽视结果导向、对学生成绩和发展重视不足的问题。相对而言,前一种问题已经受到较多重视,而后一种问题则亟待更多的关注。近些年来,我国一直倡导发展性教师评价,淡化分数、甄别与奖惩,反对以分数作为唯一标准^[3]和以升学率作为考核指标^[4],所以义务教育阶段的教师评价多强调过程、工作表现和工作量,这是相关改革成功的标志。但值得警惕的是,在有些学校,特别是没有升学压力的小学,出现了从一个极端走向另外一个极端、过分关注过程的现象。教师们努力工作,把过程做足,但这些工作究竟在多大程度上促进了学生成绩的提升以及其他方面的多样化发展,工作效果和效率怎么样,分析略显不足,在不同程度上引发高一二级学校和家长的不满。

在教师评价实践中引入增值性评价,以教师对学生成绩变化的实际贡献作为一项重要的考核指标,将在一定程度上平衡过程与结果的关系,适当增加结果在教师绩效考核中的权重,从而引导教师强化结果导向,为改善结果而不断分析与优化工作过程,最终更好地促进学生的学习与发展。

3. 提高教师评价的公平性

公平性是影响教师评价实效性的重要因素。很多地方和中小学在教师评价实践中淡化分数,除了担心以分评教可能导致教师为分数而教,还因为学生分数的高低乃至变化受到多种因素的复杂影响,不是教师个人所能决定的。如果教师评价简单地以学生分数作为主要依据,会被教师们认为不公平,从而影响评价激励和发展性功能的发挥。如果再将教师评价的结果与教师留用、晋升、奖惩等高利害决策

联系起来,很可能引发教师的不满。在这一意义上而言,难以保障公平性已然成为中小学建立多劳多得、优绩优酬、能进能出、能上能下人事制度的一个重大技术障碍。

而增值性评价通过追踪研究设计,收集学生在一段时期内不同时间点上的标准化测验成绩,在控制学生先前测验成绩及其他教师无法左右的各种因素之后,考查教师因素对学生成绩变化的净影响。有了增值性评价,教师绩效水平的高低主要取决于自己的努力、投入和专业能力,受生源、学校办学条件、同事工作表现等其他因素的影响小到可以忽略的程度,那么教师们对绩效考核公平性的主观感受就会明显改善,就更愿意接受绩效考核的结果。如果绩效工资、职称晋升等相关制度也与之配套,区域内对学生分数增值贡献大的教师可以有更多的机会获得奖励和晋升,那么教师也就更愿意向薄弱学校流动,所谓的“后进生”也可以获得教师更多的关注与支持,这将在很大程度上推动区域教育均衡与公平。

二、美国哥伦比亚学区开展教师增值性评价的核心经验

如何在教师绩效考核中有效应用增值性评价,国外是否有可资借鉴的成熟经验?美英国家在教师绩效考核中应用增值性评价的地区有很多,其中美国哥伦比亚学区(DCPS)是启动相关探索较早、持续时间最长、成效较为显著的学区之一。该学区现有116所学校,4025名教师和4万多名K-12年级的学生,是美国首都华盛顿特区最大的一个公立学区。^[5]2009年,为推动教师绩效工资改革,该区开始对教师进行较为严格的绩效考核,个体增值测量是考核内容中一个重要的组成部分。其核心经验可以概括为以下三点。

1. 全面考核教师绩效,增值性评价只是其中一个方面

哥伦比亚学区大力推动教师绩效工资改革,将教师薪酬与学生学业成绩增长情况联系起来,加大奖惩力度,拉开收入差距,力图建立优绩优酬的工资分配制度。因此,他们在教师绩效考核中重视学生学业成绩及其增值。但他们不唯分数,而是从个体增值测量(IVA)、教师评定的学生成绩数据(TAS)、教

学观察框架(TLF)、学校共同体承诺(CSC)、核心专业精神(CP)等五个方面考核教师绩效。⁶⁾表1列举了五个方面的评分办法与分数合成方式。

由表1可见,IVA是教师绩效考核体系中一个重要的组成部分,主要针对任教于考试学科的教师,包括4~10年级英语教师、4~8年级数学教师,以及其他有年度标准化测验的学科教师。在考试学科教师绩效考核中,IVA占35%的权重,TAS占15%的权重,两个反映学生学习结果的方面总计占一半分数。在非考试学科教师的绩效考核中,TAS仍然占15%的权重,其余反映教师绩效过程表现的指标占另外85%的权重。在计算得分时,以TAS为例,A老师教得好,得3.5分;B老师弱一点,得2分。因为这个教学分权重是15分,所以A老师和B老师分别有52.5分(3.5×15)和30分(2×15)计入总分。

可以说,哥伦比亚学区全面考核教师绩效,较好地平衡了结果与过程的关系,既强调结果,又关注过程,而且对考试学科与非考试学科教师区别对待,不搞“一刀切”,尊重学科差异。

2. 控制来自学生和班级层面的多种变量,准确分析教师的净影响

哥伦比亚学区采用多元线性回归分析方法构建增值模型,在模型中控制了来自学生和班级层面的多个变量。其中学生变量有同一学科之前的测试成绩、其他学科此前的成绩、免费午餐的资格、减费午餐的资格、特殊教育状态、英语熟练水平、学习障碍、转学经历以及上一学年的出勤率;班级变量有上学年的班级平均成绩、上学年学生成绩的变异程度、减免午餐的学生比例。控制变量的选择一方面基于前期对数据的探索性分析,另一方面也回应了教师对绩效考核的关切。比如:许多教师认为学生成绩的多少及变化在很大程度上受其学习基础的影响,这个基础不仅包括学生上一学年在同一学科上的测验成绩,而且包括学生上一学年在其他学科上的测验成绩。多元回归分析结果支持教师们的意见,两个变量的回归系数在统计学意义上都达到了十分显著的水平。增值模型控制了无关变量的干扰,就可以更准确地考查教师工作对学生成绩变化的净影响。在这样的考核制度下,教师不用担心任教于薄弱学校或者生源不好而

影响自己的绩效考核,从而更愿意服从工作安排,为促进教师资源分配上的均衡创造了条件。

3. 注重细节,确保增值性评价的可靠性与公平性

因为深知绩效考核在绩效工资制度体系中的重要性,哥伦比亚学区以务实的态度谨慎推动绩效考核与增值性评价,在考核数据收集、处理与分析细节上做了很多精心设计,以保证评价的客观可靠与公平公正。比如:在IVA评价中,为保证数据的可靠性,学区要求参评教师的选课人数不得少于15人,以

表1 美国哥伦比亚学区教师绩效考核评价内容与评分办法

评价内容	评分办法	分数加权与合成			
		考试学科教师		非考试学科教师	
		权重(%)	分数范围	权重(%)	分数范围
IVA	基于学生多个年度参与标准化测验的成绩,构建控制学生个体和班级层面若干变量的增值模型,分析教师对学生成绩增长的净影响;模型会结合每位教师的表现为其生成一个IVA百分等级,然后转化成1-4点的量表分数。	35	35-140	-	-
TAS	校长和教师依照学区确定的学科内容标准,共同商定学生成长的具体目标;教师采用多种方式评价学生学习,并给出终结性评价分数;校长或副校长基于教师评定的分数及证据,对教师所教学生的成长表现给1-4分。	15	15-60	15	15-60
TLF	采用学区自己开发的教学实践评分量规,从教学计划、教学过程和教学有效性三个维度,对教师的教学进行观察和评估;一般一个学年每名教师要接受正式听推门课4节,每节课评1-4分;教师教学分数为4节推门课的平均分。	40	40-160	75	75-300
CSC	测量教师在学科教学之外的表现,包括与同事的合作关系、对学校目标的贡献、与家长的关系等;评价由学校管理者实施,在寒假前和暑假前各进行一次,评分为1-4分。	10	10-40	10	10-40
CP	从出勤、迟到早退、遵守政策与规定、尊重他人等4个维度上,考查教师的敬业精神;每年寒假前和暑假前各评价一次。	CP通过减分机制在考核中发挥作用,如果教师在一次评价中被认定“略低于标准”,考核总分扣10分;如果被认定为“明显低于标准”,扣20分;两次评价最高扣分为40分。			

保证学生样本的代表性及数据的可靠性。如果考试学科教师的选课人数少于15人,他就不能/无须参加IVA评价,而是参加非考试科目组教师只有四个方面的考核。又如:学区还根据选课学生的出勤率和听课时间,给每个学生赋予相应权重,出勤率低和听课时间少的学生在回归分析中占的权重相应减少,以确保IVA估计结果不受学生投入水平的影响。

三、教师绩效考核中应用增值性评价的对策建议

在我国,关于增值性评价的研究尚处于起步阶段,严格意义上在教师人力资源管理实践中应用增值性评价的实践案例几乎没有报道。随着《深化新时代教育评价改革总体方案》的颁布,增值性评价将逐渐从研究和政策文本走向实践,成为教师绩效考核的新趋势。那么,在教师绩效考核实践中做好增值性评价需要注意哪些问题,笔者结合我国实际提出如下建议。

1. 修订有关政策,为增值性评价实施创造数据条件

增值性评价的核心数据是学生参加标准化测验所取得的成绩。这就必然面临一个现实问题:近些年来不少地市为减轻中小学生过重学业负担,明确要求严格规范考试和评价工作。比如:许多省市规定,除组织初中毕业升学考试外,区县不得在初中其他年级和小学组织统一考试。而增值性评价一般不能由学校单独实施,需要在学区乃至区县、地市等更大层面上统筹推动,如果没有跨校的统一考试,那么增值性评价就无法实施。

有人建议将增值性评价与区县教育教学质量监测数据联系起来。但值得注意的是,多数质量监测项目采用概率比例抽样或其他方式抽取一个学生样本参与,而不是针对区域内所有学生;测验内容只覆盖语文、数学、科学等部分学科,且通常情况下为避免学校之间对学生学业成就水平进行比较,参与质量监测的学生只完成若干题组中的部分测验,而不是完整的学科测验;另外,质量监测项目只针对4年级、8年级等少数年级,缺乏逐学年的连续追踪数据,所以质量监测项目也难以以为教师增值性评价提供数据支持。有关部门需要对有关政策和工作安排进行系统反思与改进,一方面使收集区域性跨年级学生

学科学业成就连续数据成为可能,另一方面又不能因此加剧应试教育倾向,增加学生负担。

2. 做好评价数据库建设,全面收集相关数据

评价数据库建设是增值性评价实施的前提。增值性评价采用追踪测验的研究范式,不仅要保存学生在多次标准化测验中的成绩数据,而且要收集来自学生、家庭、班级、教师、学校等多方面的相关数据。增值模型构建是一个不断探索与完善的过程。已有研究发现,在同一地区针对同一批教师,采用不同的模型,在增值分析中控制不同的变量,可能导致对教师绩效的高估或者低估。^[7]有时候,在同一地区,针对不同学科教师的增值性评价也需要采用不同的模型,因为对学生成绩有重要影响的因素具有学科差异。要确保模型能纳入各种影响学生成绩但教师又无法控制的因素,使模型能为管理者和教师广泛接受,我们就需要注意数据的全面性,尤其是在改革初期的探索阶段。当然,这种全面也是有限度的,究竟收集什么数据不仅要充分借鉴已有研究,而且要结合当地实际,广泛听取干部和教师意见,收集各种可能有用的信息。而且在收集过程中要注意减轻师生负担,如果有些数据在其他数据库中已经存在,就不要重复收集,只要开放数据接口和做好链接就可以。

3. 关注教育测验效度,确保数据质量

在评价考试改革与大规模质量监测项目的推动下,我国在教育测验领域取得了长足的进步,多维项目反应理论、概化理论、认知诊断模型、垂直等值方法等前沿理论与技术都已开始在实践中得到不同程度的应用。但需要注意的是,在各种教育测验编制与改进过程中,测验效度这个基础性问题尚未受到足够重视。

效度反映了一个测验题目或一整套测验准确测量欲测特质(如口语交际能力、逻辑推理、批判性思考)、针对既定目标做出推论和解释的有用性程度,它是衡量测验质量的首要指标。当前,全球最具影响力的测验质量分析框架——《教育与心理测试标准》,在1999年版^[8]及2014年最新版^[9]中都强调效度是教育测验最根本的要求,将其放在首要位置进行分析与讨论。如果测验不能有效描述学生在预测特质上的表现,不能准确推论学生达成既定学

习目标的程度,比如一项历史测验本意要考查学生对历史知识的掌握程度,但实际考查到的是学生的阅读理解能力,那么即使它的难度、区分度、信度合乎测量学要求,也是不合格的测验。题目参数估计、等值化处理等复杂技术必须在测验效度达标的前提下才有意义。

同样的道理,只有学科测验的效度合乎测量学要求,基于测验成绩及相关因素所进行的增值性评价才是可以接受的,也才能在实践中真正发挥对教师教学行为以及课程改革的积极导向作用。所以,各区县和学区在推动增值性评价过程中,必须高度重视学科测验的效度,改善测验质量,同时也要确保其他数据的真实性、有效性与可靠性。

4. 积极稳妥地推动有关改革,防止急躁冒进

美国哥伦比亚学区的教师绩效工资改革能在加大奖惩力度、拉开收入差距的背景下成功推动十多年,不仅是因为他们构建的增值模型控制了多种非教师因素,从细节上保障了评价的公平性,而且更重要的经验是他们在改革中采取积极而谨慎的态度。他们不盲目追求分析技术的复杂性,没有采用多水平模型,而是采取便于理解与实施的多元线性回归分析模型;他们在控制变量选取上不仅参考已有研究,而且积极听取和响应教师的关切;他们没有因为联邦教育部和资助方要求加强教师工资与学生分数增值的联系,就在实践中扩大化应用增值性评价,在他们的教师考核方案中,增值性评价只是教师整体绩效考核的一部分,而且只针对考试学科教师。

为此,各区县和学区探索在教师绩效考核实践中应用增值性评价,应借鉴哥伦比亚学区的经验,积极稳妥地推动有关改革,防止急躁冒进,尤其需要预防以下可能出现的问题。(1)搞“一刀切”,在所有学科都推动教师增值性评价。(2)顾此失彼,从原来的没有增值性评价变成过分依赖增值性评价,忽视教师基于自编测验评定的学生成绩、学生全面发展状况以及教师绩效过程表现,引发和加剧应试教育。(3)将合格率和升学率等指标纳入教师增值性评价,既违背有关政策,又可能导致教师关注部分学生而忽视另外一部分学生。(4)有条件要上,没条件

也要上。测验质量、数据条件以及教师心理准备程度都将在不同程度上影响增值性评价的实效性,如果条件不具备,非要推动增值性评价就可能引发各种问题。比如:有地方编制的学科测验或质量监测题目不仅存在效度不良的问题,难度也亟待关注。他们不是严格依据课标编制测验,而是人为降低测验难度,让每个学生和学校都能获得体面的分数,都“获得成功”,但难度过低时题目的区分度也会下降,基于这种测验的学生分数增值性分析,其信效度就难以保证。

总之,在教师绩效考核实践中应用增值性评价,关系到每一位教师的切身利益与专业发展,大而言之更是关系到学校、地区乃至国家的教师队伍建设。政策制定者、管理者和研究者要加强合作沟通,积极稳妥地推动有关改革。

参考文献:

- [1][4]教育部.教育部关于做好义务教育学校教师绩效考核工作的指导意见[EB/OL].(2008-12-31)[2020-8-19].http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s7051/201412/xxgk_180682.html.
- [2]赵德成.教师绩效工资改革难以深入推动的原因及对策[J].中小学管理,2020(9):28-32.
- [3]教育部.教育部关于积极推进中小学评价与考试制度改革的通知[EB/OL].(2002-12-18)[2020-8-20].http://www.moe.gov.cn/srcsite/A26/s7054/200212/t20021218_78509.html.
- [5]DCPS. DCPS at a Glance: Enrollment[EB/OL].(2019-12-08)[2020-8-20].<https://dcps.dc.gov/node/966292>.
- [6]Gitomer D H, Crouse K, Joyce J. A Review of the DC IMPACT Teacher Evaluation System[R/OL].(2018-02-26)[2020-08-20].https://www.researchgate.net/publication/280718857_A_review_of_the_DC_IMPACT_teacher_evaluation_system.
- [7]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):88-95,112.
- [8]美国教育研究学会,美国心理学会,美国教育测量学会.教育与心理测试标准[S].燕妮琴,谢小庆,译.沈阳:沈阳出版社,2003:12-15.
- [9]AERA, APA, NCME. Standards for Educational and Psychological Testing[S].Washington, DC: American Educational Research Association, 2014: 11.