

学业质量增值评价数学模型的比较研究

——以西部某区教育实证数据为例

钟冬梅 卢锦运 贾玲

(重庆市教育评估院,重庆 400020)

摘要:增值评价是以学生成长为核心,以学校发展为目标的一种评价方式,可以更公平地评价学校效能。增值评价有多种数学模型,其中,与基于平均名次、基于标准分的两种模型相比,基于学生成长百分位和多层线性回归两种模型的优势在于能够避免优秀学生增值空间受限的问题。在多层线性回归模型中纳入学生性别、家庭社会经济地位指数等学校不可控的生源因素,可有效排除生源因素对学生的影响,更符合关注学校“净效能”的评价理念。

关键词:学业质量;增值评价;平均名次;标准分;学生成长百分位;多层线性回归

中图分类号:G40

文献标识码:A

文章编号:1673-0429(2023)03-0035-14

doi:10.19742/j.cnki.50-1164/C.230304

一、问题提出

随着教育的不断深入和发展,学校评价已成为教育领域中的热门话题之一。以往,在学校评价方面,忽略学校及教师的努力和付出,仅以学生学业发展的最终结果作为单一标准来评价学校效能,难以全面反映教育发展的真实状况和教育自身的内在规律,也无法充分展现教育公平性和社会公正性。2020年10月,中共中央、国务院印发《深化新时代教育评价改革总体方案》,对教育评价改革进行了总体的科学设计,首次提出了“四个评价”,即改进结果评价,强化过程评价,探索增值评价,健全综合评价。其中的增值评价,是指通过追踪研究设计,收集学生在不同时间点上的学业成绩数据,并考虑不受学校或教师控制的生源因素(如学生的初始成绩、人口学因素、家庭背景信息等)对学生成绩的影响,建立数学模型,在纵向追踪学生学业成绩变化的过程中,将这些因素对学生成绩的影响与学校或教师对学生成绩的效应分离开来,从而考察学校或教师对学生学业成绩影响的净效应,实现对学校或教师效能较为科学、客观的评价。简而言之,增值评价就是对学生学业成绩进步程度的评价,是一种以学生成长为

收稿日期:2023-03-22

作者简介:钟冬梅(1991—),女,重庆奉节人,重庆市教育评估院助理研究员。研究方向:教育评价与数据挖掘。

卢锦运(1985—),男,福建龙岩人,重庆市教育评估院教育评价研究与数据中心副研究员。研究方向:教育质量监测与教育信息化。

贾玲(1986—),女,重庆云阳人,重庆市教育评估院教育综合评估所助理研究员。研究方向:教育综合评价、教师评价。

基金项目:重庆市教育科学“十三五”规划2020年度一般课题“基于第三方专业测评的教师绩效评价创新研究”(2020-00-435)。

核心,以学校发展为目标的评价方式,其根本特征是将教育评价的参照系从外部统一的参照转向了自我增值的参照,更关注学生学习过程,更关注教师和学校对学生发展的贡献。

在学业评价中采用增值评价,可以更加客观、公正、全面反映学生的学业发展过程,而不仅仅是学生的最终成绩。且随着学校增值评价理论的不断发展和技术的日益完善,我们现在可以使用特定的数学模型来跟踪学生学业成绩的变化,并测算出学校或教师发挥作用的程度,以此来评价学校或教师对学生学习质量提升的贡献大小,为学校 and 教师提供更准确、更公正、更有针对性的评价结果,这一评价方法可以激发广大教育工作者和社会各界的积极性,帮助每所学校更好地发展。因此,学校增值评价研究具有重要的理论和实践意义。

目前,国内外常用的增值评价模型有基于平均名次、基于标准分、基于学生成长百分位和基于多层线性回归等多种模型,虽然各个模型在评价理念上都是通过量化学生的进步和变化程度来评价学校效能,但是各个模型在分析方法上存在一定的差异,最终的评价结果也有所不同。那么,在教育实践中,不同增值评价模型所测算的增值结果存在怎样的差异?究竟应该选取哪一种方法来开展增值评价更科学合理?如何科学地解释和运用学校增值评价结果?本研究采用西部某区 41 所初中教育质量评价相关数据进行实践分析,尝试探讨这些问题。

二、文献综述

学校增值评价最早起源于 20 世纪 80 年代美国底特律市。当时底特律的一些教育家开始尝试利用学生测试数据来评估教师的效果。这种方法后来被称为价值增值模型(Value-Added Model, VAM),并在 20 世纪 90 年代开始在美国其他城市和州以及其他国家推广使用^[1]。1992 年,增值评价正式纳入到美国田纳西州的评价体系,成为广为人知的田纳西州增值评价系统。截至 2004 年,美国共有 21 个州将增值评价纳入到其评价系统中^{[2][3]}。1990 年代初,加拿大开始实施学校增值评价,主要通过比较学生的历史成绩和预期成绩,以确定学校、教师和教育者对学生学业成就的影响程度。在法国,学校增值评价的实施始于 21 世纪初,是由法国教育部预测评估司组织的学校评估指标体系中的一个重要部分^[4]。我国香港增值评价的实施始于 2004 年,当时香港特区政府推出了“教育关怀计划”,其中包括引入增值评价作为评估学生学业进展的一种方法。自那时起,香港的学校逐渐采用了增值评价,并逐渐将其纳入到教学和学生评估中。目前,我国北京、上海、江苏、浙江、山东、湖南、广东、海南、重庆等地也都在探索实施增值评价,部分省市还在试行阶段,其评价结果的准确性和可靠性有待进一步检验。从国内外各地区应用增值评价结果来看,都主要是将评价结果用于评估学校、教师和教育计划的有效性,以及指导政策制定者、学校管理人员和教育工作者的决策,为教学改进、资源分配和教育政策制定提供重要的信息和指导。

虽然都运用了增值评价的理念来对学校、教师效能进行评价,但是不同国家和地区所采用的模型和方法并不一样。比如美国目前广泛使用的模型是学生成长百分等级(student growth percentile, SGP)来测算增值分数^[5]。新加坡主要是用学生预期的成绩和他们的实际成绩相比较,来测算学校对学生发展的作用,从而对学校效能做出评价^[6]。英国的学校效能研究者与地方教育当局则采用多层分析模型来评估本区域的学校效能和推进学校发展^[7]。我国研究者针对不同的模型也进行了研究,比如赖寒对国际常用的四种增值评价模型即获得分数模型、协变量校正模型、田纳西模型、持续效能模型的主要算法进行了系统梳理^[8];王旭东则梳理了五类常见的增值评价模型的分析算法、不足和优势,包括基于名次的增值评价模型、基于等第的增值评价模型、基于分数的增值评价模型、基于名次和等第的增值评价模型、基于分数和等第的增值评价模型^[9];周园、刘红云等主要研究了多层线性分位数回归模型的算法与应用^[10]。

国内学者在探索和理解增值评价理念的基础上,开展了一系列实证研究,推动了增值评价在国内的应用和发展。比如边玉芳、林志红以浙江省金华市部分高中学生为样本,通过增值模型分析验证了不同区域、类型的学校增值差异性^[11]。王家美等对江西省上饶市高中语文学科建立了增值评价模型,发现公办校与民办校、重点学校与非重点学校对学生学业发展的效能均存在显著差异,同时进一步分析了学校对不同性别学生、不同学习能力水平的学生成长的效能是否有所不同^[12]。范美琴等选取江苏省某市部分学生的中考和高考成绩数据进行了实证研究,并将增值评价与结果评价进行了对比分析,得出了结果评价与增值评价整体上具有一致性、局部差异较大,增值评价在解决生源起点不公的问题上具有独特优势的结论^[13]。邵越洋、刘坚基于北京市某区教育质量监测数据,基于多层线性回归建立了增值评价模型,纳入了学生家庭社会经济地位、学生基础认知能力等生源指标,并比较结果评价与增值评价在学校教育质量评价上的差异^[14]。

综上,增值评价已成为各国和国内各地区教育测量和评价体系中的重要组成部分,被广泛应用于学校和教师的评估中,而且各类增值评价模型的分析理念均得到了充分的思考和讨论,也进行了广泛的应用和实践,但是,缺乏运用实证分析来验证不同增值评价模型之间的差异并检验其科学性和公平性。

三、研究对象与方法

(一)数据来源

本文主要采用了西部某区开展的初中学校教育质量评价项目数据。学生群体为 2020 年升入初中的初一学生,2020 年 12 月进行了基线测试(以此次测验作为“基准”成绩),2022 年 6 月进行了后测;测试内容主要聚焦考查语文、数学、英语 3 个核心学科素养;测验工具主要为全区统考的期末考试卷。此外,本文通过问卷调查,采集学生的背景资料信息。每所学校有效学生样本量在 47~1 022 人之间。

表 1 各学校人数统计

学校	学生数	男生数	女生数	学校	学生数	男生数	女生数	学校	学生数	男生数	女生数
学校 1	395	222	173	学校 15	1022	563	459	学校 29	197	111	86
学校 2	297	175	122	学校 16	234	133	101	学校 30	455	265	190
学校 3	224	115	109	学校 17	675	352	323	学校 31	183	97	86
学校 4	470	262	208	学校 18	206	110	96	学校 32	15	10	5
学校 5	275	160	115	学校 19	66	35	31	学校 33	298	170	128
学校 6	394	215	179	学校 20	259	143	116	学校 34	111	70	41
学校 7	903	519	384	学校 21	155	92	63	学校 35	111	75	36
学校 8	261	153	108	学校 22	305	177	128	学校 36	82	46	36
学校 9	289	180	109	学校 23	116	65	51	学校 37	137	80	57
学校 10	443	248	195	学校 24	141	85	56	学校 38	73	42	31
学校 11	297	154	143	学校 25	501	280	221	学校 39	354	181	173
学校 12	300	174	126	学校 26	71	33	38	学校 40	247	133	114
学校 13	291	179	112	学校 27	609	366	243	学校 41	384	228	156
学校 14	381	200	181	学校 28	156	82	74	总人数	5220	2956	2264

(二) 变量说明

根据以往研究,学生生源如学生性别及经济因素等指标会显著影响学生的学业质量^[14],因此,本文除关注基测及后测的学生学科成绩变化外,自变量还加入了学生的性别、家庭社会经济地位得分。

学生社会经济地位指数的测算主要是参照了 PISA2018,由学生父母职业、学历、家庭拥有物相关变量合成,学校内学生社会经济地位指数平均后即为学校平均社会经济地位指数。将两次测试的语文、数学、英语 3 科成绩各自标准化,分别得到 3 科成绩的标准分,以此作为本研究的基线成绩和后测成绩。

表 2 各自变量和因变量说明

类别	变量	说明
自变量	学生前测成绩	2020 年 12 月的基线测试成绩
	学生性别	1 为男生,0 为女生
	学生家庭社会经济地位指数	通过将学生父母职业、受教育程度、家庭拥有物相关变量而合成分数 SES
	学校平均社会经济地位指数	各学校所有学生 SES 得分的平均值即为学校平均 SES
因变量	学生后测成绩	2022 年 6 月的后测成绩

表 3 学生基线测试和后测成绩描述统计

学生学业成绩	满分值	最小值	最大值	平均得分	标准差
语文基线成绩	120	10.35	106.55	80.39	12.4
数学基线成绩	100	9.9	98.3	63.51	17.76
英语基线成绩	100	13.7	99.23	63.8	18.99
语文后测成绩	120	1	114	84.61	17.5
数学后测成绩	100	0.5	99.5	60.14	23.63
英语后测成绩	100	3	100	62.27	24.03

(三) 分析方法

本文主要采用了基于平均分名次、基于标准分、基于简化后的学生成长百分位和基于多层线性回归的增值评价模型四种方法进行测算。

1. 基于平均名次的增值评价模型

每次考试,先按每个学生分数由低到高排序(位次数值越大代表成绩越好),然后将各学校所有学生的名次求平均,计算出各学校两次测试的学生平均名次 $N_i(i=1,2)$,按 $\Delta N=N_2-N_1$ 得到平均名次差值(即为相对增值),差值为正数,代表进步,差值为负数,代表退步,差值越大,代表学校进步或退步的程度越大。

2. 基于标准分的增值评价模型

基于标准分的增值评价模型直接将两次成绩的标准分数相减的差值,即 $\Delta Z=Z_2-Z_1$,计算出各学校的增值。与前一种方法在原理上,都是看学生在这群体中两次成绩的相对位次变化,操作简单,便于普通教育工作者和一线教师理解。

3. 基于学生成长百分位的增值评价模型

该模型的主要步骤是:计算每个学生在所有样本学生中的前测成绩百分位,然后计算每个学生在同一前测成绩百分位的学生群体中的后测成绩百分位,来作为其学业成长百分位。例如:基线测试时某学生成绩百分位为 50,而后测成绩为 60 分,要计算其成长百分位,首先找出基线成绩百分位都为 50 的学

生,再计算该学生后测成绩(60分)在这些学生的后测成绩中的百分位,这个百分位就是该学生的成长百分位。这一方法的主要原理是每个学生都是在与其基础相同的学生群体比较进步幅度,能够有效减轻不同基础的学生群体比较时进步空间不同而产生的不公平性,但要求前测成绩每个得分点的学生样本数量要足够多才行。本研究结合数据实际,对该方法进行了调整,主要是采用K-MEANS聚类方法,根据学生各科基线成绩将学生进行聚类,同一类的学生表示前测成绩相当,以此解决前测成绩相同的学生人数较少的问题,再计算各学生在同类学生中后测成绩的百分位,求各学校所有学生的平均百分等级代表学校的增值分数,高于50代表正增值,低于50代表负增值。

4. 基于多层线性回归的增值评价模型

在这个模型中,每个学生的成绩可以被认为是由多个因素影响的,包括学生自身的能力、家庭环境、学校资源、教学方法等等。因此,多层线性回归模型可以建立在多个层面上,包括学生、教师、学校和区域等不同的层面。在每个层面上,模型都会考虑到该层面上的因素对学生成绩的影响,并给出一个对应的回归方程。该模型具体算法为将学生后测成绩作为因变量,将学生基线成绩作为预测变量,通过加入影响学生成绩的学校不可控因素校正模型,进一步测算学生后测成绩预期值,实际后测成绩与预期成绩的差值(残差)即为学生的“增值”,对各所学校所有学生的增值数据求平均,即为学校的增值得分,分数大于0为正增值,小于0为负增值。

教育领域的数据一般具有嵌套结构,在构建完整的多层线性回归模型之前,本研究先构建学生层和学校层都不纳入任何解释变量的零模型^[15],如下公式:

$$Y_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = r_{00} + u_{0j}$$

其中, Y_{ij} 表示第*j*所学校第*i*个学生的学科成绩, β_{0j} 表示第*j*所学校的截距, e_{ij} 相互独立,服从期望值为0,方差为 σ^2 的正态分布, r_{00} 表示所有学校的平均截距, $Var(u_{0j}) = \tau_{00}$, σ^2 表示学业成绩学生层面的变异, τ_{00} 表示学业成绩的学校层面的变异。

$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$ 表示学生学业成绩总的变异中,可由学校层面解释的方差比例,也是学校内学生之间的相关系数ICC(也称校际差异)。

本研究在零模型基础上纳入了学生层面的性别及家庭社会经济地位指数(SES),以及学校层面的学生平均社会经济地位指数,构建了随机截距模型,如下公式:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + \beta_2 SEX_{ij} + \beta_3 SES_{ij} + e_{ij}$$

$$\beta_{0j} = r_{00} + SES_SCHOOL_j + u_{0j}$$

模型中, X_{ij} 为第*j*所学校第*i*个学生的基线成绩, β_1 为其斜率。 SEX_{ij} 和 SES_{ij} 分别为学生层面的性别变量(1为男生,0为女生)和社会经济地位指数, SES_SCHOOL_j 为第*j*所学校的平均社会经济地位指数。

(四) 数据处理

本文主要运用R软件、SPSS22.0进行模型拟合与增值计算,运用EXCEL进行数据管理和数据可视化。

四、研究结果

(一) 四种模型的增值评价结果与学生基线成绩的关系

从表4可知,基于平均名次和基于标准分测算的学生语文、数学、英语3科增值得分均与对应的前

测成绩显著负相关,相关系数在-0.628 到-0.586 之间,说明基于这两种方法,由于学生基础成绩越好,进步空间会越小,学生的增值得分就会被限制,这样的增值评价对于优秀学生或者优质生源学校不公平。

同时,基于学生成长百分位和基于多层线性回归模型测算的学生语文、数学、英语 3 科增值得分均与对应的前测成绩不存在相关性,说明基于这两种模型,学生的基线水平不会影响学生增值得分,不同层次学生的进步空间均相当。

表 4 基于四种方法的增值得分与前测成绩的 Spearman 相关系数

	基于平均名次	基于标准分	基于成长百分位	基于多层线性回归
语文学科前测成绩	-0.586 **	-0.625 **	0.003	0.007
数学学科前测成绩	-0.613 **	-0.623 **	0.005	0.002
英语学科前测成绩	-0.624 **	-0.628 **	0.002	0.001

注: **指 $p<0.01$ 。

(二) 学生成长百分位模型与多层线性回归模型的比较

两张模型都可以在一定程度上解决优秀学生增值空间有限的“天花板效应”问题。其中,基于学生成长百分位的增值评价模型主要是通过将学生前测成绩进行分类,让学生在同起点的学生群体中比较,来解决这一问题;基于多层线性回归的增值评价模型则主要是通过引入学校不可控因素,测算学生的预期成绩,将学生实际成绩与预期成绩的差值作为增值,也有效解决了优秀学生进步空间小的问题。表 5 对两种模型下学生增值得分与学生社会经济地位指数进行了相关分析,发现基于成长百分位模型,学生的三学科增值得分均与学生社会经济地位指数显著正相关,相关系数在 0.300 到 0.358 之间,说明这一方法因为只考虑了两次成绩,而没有排除由于其他因素对学生成绩的影响。基于多层线性回归模型的学生增值得分与学生社会经济地位指数不相关,可以看出,这一结果是排除纳入的关键变量影响后,学校对学生进步程度的“净效能”,其增值结果不受学生基线成绩、学校不可控的生源状况的影响。

表 5 两种模型下学生三学科增值得分与学生社会经济地位指数的 Spearman 相关系数

	基于成长百分位	基于多层线性回归
语文学科	0.300 **	0.004
数学学科	0.339 **	-0.002
英语学科	0.358 **	-0.007

注: **指 $p<0.01$ 。

(三) 多层线性回归中纳入学校不可控的生源因素的影响

本文先构建不包含任何层次变量的零模型,来计算 3 个学科成绩的组内相关系数(intraclass correlation coefficient,简称 ICC)以检验校际变异占总变异的大小,当 ICC 大于 0.059 时,组内达到中度以上相关,组内相关对估计回归系数、标准误所产生的影响不可忽视^[16]。结合图 1 可知,学生语文、数学、英语三科学业成绩的校际差异分别为 23.46%,22.36%,24.54%。均超出 20%,有必要采用多层线性回归进行分析。

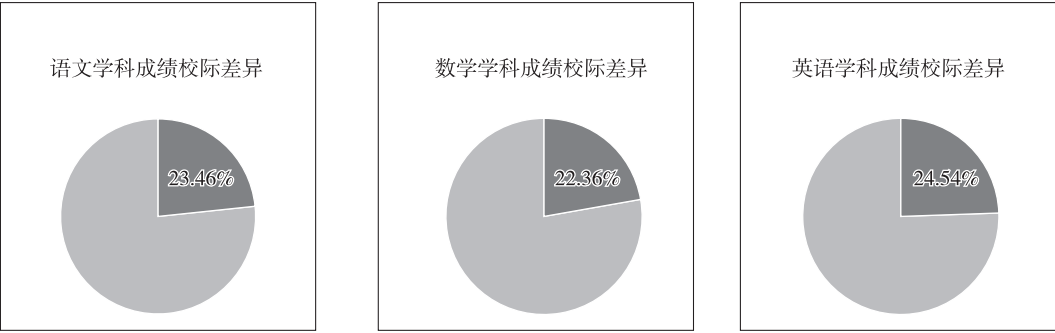


图 1 学生三科学业质量的校际差异(组内相关系数)

由表 4 可知,对三个学科进行模型拟合时,基线成绩、学生 SES、学校平均 SES 系数均大于 0,且均显著($P=0.000$),说明这三个变量都对学生的后测成绩具有显著的正向作用;其次,在语文和英语学科上,男生显著低于女生(系数分别为 -0.436 、 -0.275 , $P=0.000$),数学学科上模型截距和性别变量不显著($P>0.05$)。

表 6 各学科多层线性回归拟合系数

	参数	系数	标准误差	df	T	P
语文学科	截距	-0.254	0.044	27.715	-5.81	0.000
	性别	-0.436	0.017	10 973.237	26.211	0.000
	学生 SES	0.169	0.01	10 966.241	16.736	0.000
	基线成绩	0.145	0.008	10 975.35	17.35	0.000
	学校平均 SES	0.339	0.067	29.393	5.082	0.000
数学学科	截距	-0.031	0.03	35.188	-1.059	0.297
	性别	0.025	0.017	10 951.3	-1.501	0.133
	学生 SES	0.205	0.01	10 941.052	20.047	0.000
	基线成绩	0.155	0.008	10 957.472	18.296	0.000
	学校平均 SES	0.44	0.046	39.038	9.624	0.000
英语学科	截距	-0.141	0.027	37.442	-5.269	0.000
	性别	-0.275	0.016	10 979.661	16.692	0.000
	学生 SES	0.215	0.01	10 968.857	21.455	0.000
	基线成绩	0.13	0.008	10 982.937	15.799	0.000
	学校平均 SES	0.502	0.042	41.874	12.057	0.000

为进一步清楚地解释生源信息变量对学生成绩、学生增值分数的影响,本研究还挑选了 10 个典型学生进行个案对比分析。表 7 中的 5 个学生来自同一个学校,且具有相同的社会经济地位,从学生 1-3 (皆为男生)的对比分析看出,语文前测成绩越高,模型测算的学生的预期后测成绩得分也越高,因为前测成绩对学生后测成绩有显著的正影响。但从学生 1-3(男生)与学生 4-5(女生)对比分析发现,学生 1-3 前测成绩高于学生 4-5,但预期成绩低于学生 4-5,从模型拟合系数看出,该校女生的语文成绩显著高于男生,所以在计算预期成绩时,女生的预期值也会高于男生,说明在模型中加入性别变量,对于存在性别差异的学科做增值分析时,可以通过调节预期值来排除这一因素对增值结果的影响。

表 7 五个学生数据看性别差异对语文成绩的影响

	社会经济地位	性别	语文前测成绩	语文后测成绩	语文预测成绩	语文增值得分
学生 1	-1.828	男	0.013	-1.806	-0.826	-0.981
学生 2	-1.828	男	0.118	-2.549	-0.81	-1.739
学生 3	-1.828	男	0.569	-1.32	-0.745	-0.576
学生 4	-1.828	女	-0.584	-1.121	-0.476	-0.644
学生 5	-1.828	女	-0.33	0.251	-0.439	0.69

表 8 是来自同一个学校的三个男生,数学前测成绩得分相同,学生 6,7,8 的家庭社会经济地位指数得分依次增加。由表可知,在性别、基线成绩一样时,社会经济地位越高的学生,其数学预期成绩会越高。

表 8 三个学生数据看学生社会经济地位差异对学生数学成绩的影响

	性别	社会经济地位	数学前测成绩	数学后测成绩	数学预测成绩	数学增值得分
学生 6	男	-2.332	-0.367	-0.937	-0.739	-0.198
学生 7	男	-0.691	-0.367	-1.381	-0.402	-0.979
学生 8	男	0.145	-0.367	0.671	-0.231	0.902

表 9 的两个男生来自不同学校,两个男生的社会经济地位指数相同,英语前测成绩得分相同,但学校 25 的平均社会经济地位指数高于学校 28,因此学生 10 的英语预期成绩高于学生 9。

表 9 两个学生数据看学校平均社会经济地位差异对学生英语成绩的影响

	学校	性别	社会经济地位	学校平均社会经济地位	英语前测成绩	英语后测成绩	英语预测成绩	英语增值得分
学生 9	学校 28	男	-1.795	-0.633	0.180	-1.863	-0.821	-1.043
学生 10	学校 25	男	-1.795	-0.392	0.180	1.237	-0.700	1.937

(四) 增值评价结果与传统评价结果的比较

基于前面模型的比较研究,本研究选取了基于多层线性回归的增值评价与传统的结果评价,来进一步对比分析。

当考虑学校起点因素与生源因素时,可以发现结果评价排序靠前的学校,一般都拥有较高的起点,且拥有较好的生源条件;从表 10 可知,语文、数学、英语三个学科的结果评价与基线测试排序都显著正相关,相关系数分别是 0.848,0.881,0.844;结果评价与学校平均 SES 排序也显著正相关,三科相关系数分别是 0.923,0.958,0.954。说明学校的生源基础即学生起点水平、学生家庭社会经济地位是影响学校结果评价的主要因素。但是,增值评价排序与学校的起点因素、学校平均 SES 相关性很小,增值评价的排序与结果评价的排序存在中等程度相关(英语学科不相关),即从学校评价结果看,生源因素不会对增值评价的排序产生影响。

表 10 学校增值结果与各因素的 Spearman 相关系数

		学校平均 SES 排序	学校基线测试排序	学校结果评价排序	学校增值评价排序
语文学科	学校平均 SES 排序	1			
	学校基线测试排序	0.851 **	1		
	学校结果评价排序	0.923 **	0.848 **	1	
	学校增值评价排序	0.176	0.176	0.467 **	1
数学学科	学校平均 SES 排序	1			
	学校基线测试排序	0.854 **	1		
	学校结果评价排序	0.958 **	0.881 **	1	
	学校增值评价排序	0.265	0.296 *	0.460 **	1
英语学科	学校平均 SES 排序	1			
	学校基线测试排序	0.857 **	1		
	学校结果评价排序	0.954 **	0.844 **	1	
	学校增值评价排序	-0.002	0.081	0.238	1

注: **指 $p < 0.01$.

以英语学科为例,表 11 显示 41 所样本学校各变量及其评价结果(排序)的情况,便于直观比较两种评价方式对学校排序的影响,并同时纳入学校起点因素与平均社会经济地位指数。从表可知,运用传统学校结果评价来比较学校效能时,起点水平和生源因素较好的学校,基本都是位列前茅,但运用增值评价调查学校为学生学业发展所做的贡献大小时,则不受起点水平和生源因素的影响。例如表中的学校 5,学校平均社会境地位排名第 4,基线成绩排名第 2,在时隔一年半后的后测中排名第 4,名次幅度变化较小,但是其增值得分为负值,排在了第 22 位;相反,学校 17 的社会经济地位、基线测试成绩和结果评价的排序分别位居第 16,11,10 位,但其增值评价排序跃居第 1 位。

表 11 样本学校在各指数级英语学科测试成绩的排序结果

学校 代码	学校平均 SES		学校基线测试		学校结果评价		学校增值评价	
	分数	排名	分数	排名	分数	排名	分数	排名
学校 8	0.673	7	0.246	1	0.455	7	-0.049	23
学校 5	0.728	4	0.216	2	0.516	4	-0.023	22
学校 10	0.541	9	0.215	3	0.498	5	0.113	9
学校 7	0.680	6	0.208	4	0.614	3	0.117	8
学校 2	0.968	2	0.183	5	0.692	1	-0.016	20
学校 11	0.524	10	0.173	6	0.385	9	-0.003	19
学校 6	0.719	5	0.166	7	0.427	8	-0.093	31
学校 4	0.730	3	0.161	8	0.495	6	-0.056	24
学校 1	1.108	1	0.155	9	0.688	2	-0.122	35
学校 9	0.603	8	0.145	10	0.247	12	-0.185	37
学校 17	-0.066	16	0.117	11	0.315	10	0.363	1

续表11

学校 代码	学校平均 SES		学校基线测试		学校结果评价		学校增值评价	
	分数	排名	分数	排名	分数	排名	分数	排名
学校 39	-0.970	38	0.103	12	-0.792	36	-0.075	27
学校 12	0.273	11	0.026	13	0.266	11	0.105	11
学校 16	-0.045	15	0.010	14	0.069	13	0.110	10
学校 28	-0.633	27	-0.001	15	-0.385	29	0.091	14
学校 20	-0.152	19	-0.005	16	-0.108	18	0.051	17
学校 14	0.154	13	-0.009	17	-0.156	20	-0.222	38
学校 24	-0.320	23	-0.018	18	-0.241	22	0.007	18
学校 19	-0.099	18	-0.025	19	-0.020	16	0.095	13
学校 23	-0.313	22	-0.034	20	-0.261	23	-0.018	21
学校 27	-0.596	26	-0.083	21	-0.387	30	0.069	15
学校 13	0.157	12	-0.084	22	-0.025	17	-0.087	30
学校 21	-0.220	20	-0.089	23	-0.284	25	-0.070	26
学校 22	-0.244	21	-0.098	24	0.017	15	0.238	4
学校 18	-0.097	17	-0.109	25	-0.239	21	-0.119	33
学校 30	-0.657	29	-0.126	26	-0.376	27	0.159	7
学校 26	-0.401	25	-0.135	27	-0.382	28	-0.083	29
学校 15	0.141	14	-0.151	28	0.027	14	-0.065	25
学校 35	-0.924	34	-0.157	29	-0.610	33	0.100	12
学校 37	-0.951	36	-0.165	30	-0.863	37	-0.129	36
学校 25	-0.392	24	-0.169	31	-0.151	19	0.170	6
学校 33	-0.770	32	-0.184	32	-0.734	34	-0.121	34
学校 31	-0.754	30	-0.204	33	-0.323	26	0.271	2
学校 29	-0.641	28	-0.206	34	-0.280	24	0.241	3
学校 32	-0.761	31	-0.231	35	-0.531	31	0.062	16
学校 40	-1.009	39	-0.257	36	-0.589	32	0.213	5
学校 34	-0.781	33	-0.302	37	-0.751	35	-0.095	32
学校 42	-1.127	41	-0.333	38	-1.144	39	-0.337	39
学校 36	-0.925	35	-0.381	39	-1.227	40	-0.484	41
学校 41	-1.033	40	-0.461	40	-0.913	38	-0.083	28
学校 38	-0.963	37	-0.702	41	-1.230	41	-0.421	40

图 2-4 分别呈现了 41 所样本学校在语文、数学、英语学科上的增值评价结果,这里给出了学校增值估计值的分布范围,即以增值估计值为基础,考虑了标准误和置信水平(95%)而给出一个区间,如果置信区间的下限大于零,则为正增值,表示学校学生的进步明显,实际水平显著高于预期水平;如果置信区

间的上限小于零,则为负增值,表示学校学生退步明显,实际水平要显著低于预期水平;如果置信区间的上限大于零且下限小于零,则为零增值,表明学校学生实际水平与预期水平相当。从结果来看,该区 41 所学校中,语文、数学、英语实现了正增值的学校数分别为 14,11,12 所;语文、数学、英语为负增值的学校数分别为 9,8,8 所;语文、数学、英语为零增值的学校数分别为 18,22,21 所,将此结果反馈给所在区域教育行政部门和学校,可以帮助其有针对性地制定行政决策和改进策略。

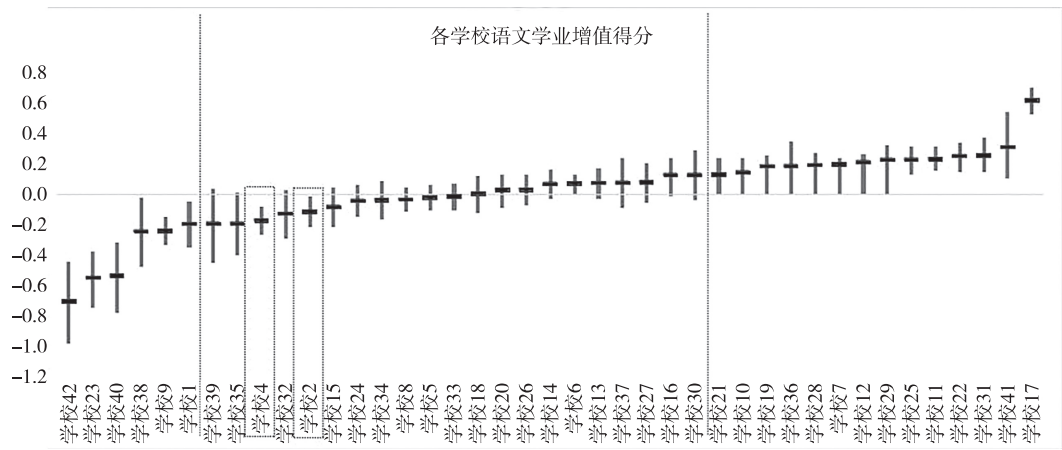


图 2 各学校语文学科增值评价结果

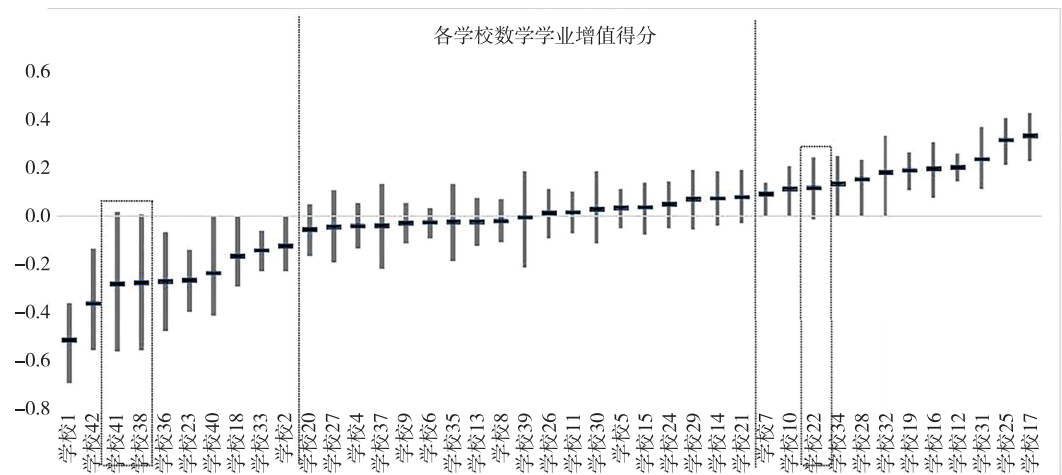


图 3 各学校数学学科增值评价结果

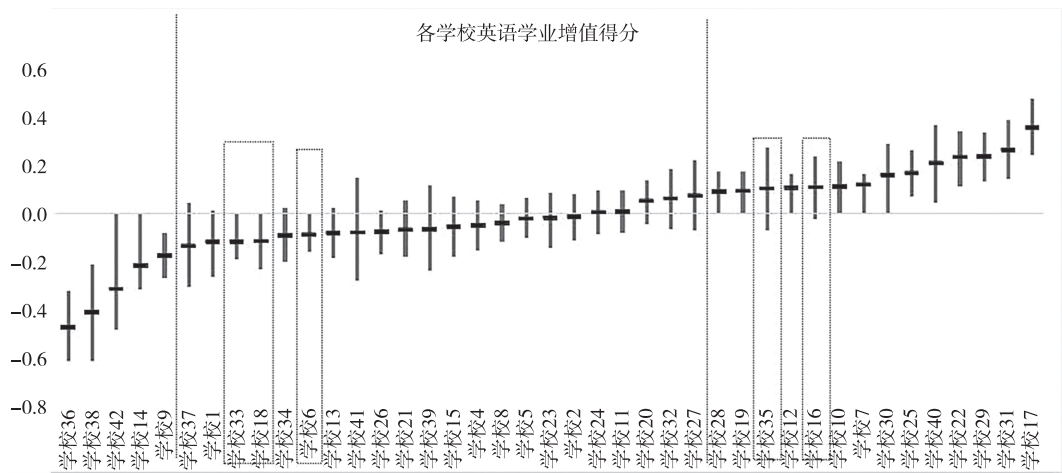


图 4 各学校英语学科增值评价结果

五、研究结论与讨论

(一) 研究结论

1. 基于学生成长百分位模型与多层线性回归的增值评价模型都能在一定程度上避免“天花板效应”和“地板效应”。学生成长百分位模型主要是通过让学生与起点水平相当的同类学生群体比较来看增值,多层线性回归模型则是通过比较学生的实际水平与预期水平来看增值。与基于学生平均名次、学生成绩标准分的增值评价相比,这两种模型都有效地解决了优秀学生增值空间受限的“天花板效应”,以及起点水平较低的学生更容易有较大增值的“地板效应”。

2. 基于多层线性回归的增值评价模型更符合关注学校“净效能”的评价理念。与基于学生平均名次、学生成绩标准分、学生成长百分位等增值评价模型相比,基于多层线性回归的增值评价模型不仅考虑了学生起点水平,还考虑了学生背景因素的影响,更强调学校“净效应”,可以准确地评价学校在可以做出改变的方面所体现出来的努力及成效,从而为每个参与评价的学校提供了新的改进思路 and 方向,更加符合现代教育评价的理念与趋势。

3. 学业质量增值评价与传统的结果评价相比,较好地提升了学校评价公平性,能够促进学校及教师良性竞争,挖掘学生学习潜力。因为学生成绩在很大程度上取决于学校控制范围以外的因素,只根据最终结果对学校进行评价是不公平的^[17]。增值评价可以让不同生源质量的学校进行科学的比较,破解了“强者恒强、弱者恒弱”评价瓶颈,在学校评价的公平性上迈出了一大步,促进了学校及教师的良性竞争。且从学生发展层面来讲,增值评价能够帮助每个学生定位自己的起点以及增值幅度,更科学地帮助学生了解自身在各个学科上的优势与不足,挖掘出每一个学生的发展潜力。

(二) 讨论

1. 各级各类教育工作者可以根据实际选择不同模型探索增值评价

基于平均名次和标准分的增值评价模型虽然未能很好地避免“天花板效应”和“地板效应”,但与传统结果评价相比,也在一定程度上淡化了原始分数的绝对性,消除了不同学科、不同测试之间的差异性,增加了前后测试的可比性。对于中小学管理干部和一线教师而言,想要探索增值性评价,可先从这两种相对简单、容易理解、便于操作的模型入手,并运用评价结果充分调动师生积极性,有针对性地改进和调控教学行为,提高教学质量,但在运用评价结果时需要注意其评价的局限性,特别是针对优生和差生群体,这种类模型的评价结果存在较大误差。

基于学生成长百分位的增值评价基本理念也容易被一线教育工作者和学生、家长理解、接受,数据分析步骤相对也比较清晰,是值得探索应用的一种增值评价方法。基于多水平线性回归的增值评价模型,既能计算中小学生在一定时期内学业发展水平的“增值”,又能在纷繁复杂的学习场景中剔除非教师和学校因素的影响,且在处理涵盖学校、班级、学生不同层级的分层嵌套数据时,只有多水平分析方法是最合适的。这一模型提高了对学生在学习过程中或学习后的进步程度的评价信度与效度。不过,多水平回归模型的操作难度较大,一线教育评价工作者很难理解且使用多水平回归分析方法来开展增值评价。教育行政部门、学校可以引进第三方专业评价机构,由专业的数据分析人员进行模型测算,充分发挥这一数学模型在学生增值评价过程中的导向性和实用性,运用增值评价结果,合理制定管理决策,激发每一所学校进步的活力,引导教师关注每一个学生的进步。

2. 科学量化学生学业增值相关影响因素的效能是践行增值评价理念的关键

学生学业发展受学生、家庭、班级、学校、区域等不同层级、不同类型因素的综合影响,同时,学生的

学业增值、成长进步也是这些因素共同影响的结果。探索增值评价,就是为了合理区分不同因素对学生的影响,关注学校和教师对学生发展的净效能,进而实现科学、客观评价学校教育教学工作绩效,合理减轻普通学校和薄弱学校在传统结果评价中的不公平感,提高改善学校教育的积极性;同时,引导教师关注全体学生成长,充分尊重学生个体差异,挖掘学生潜能,促进全部学生共同进步和发展。

在实施增值评价过程中,通过量化方式设计回归方程,运用层层嵌套的增值模型,全面考虑学生、班级、学校、区域等不同层级变量对学生发展的影响,把不受学校、教师控制的生源因素如家庭社会经济地位、学生初始水平等因素的影响剥离开来。当然,除了本文所考虑的家庭社会经济地位、学生性别等,笔者认为进一步考虑家长教育观念、教育能力、教养风格等家庭因素的影响,可以更加精准地刻画学校和教师净效能,是践行增值评价理念的关键。

3. 系统构建学生成长发展数据库是发挥增值评价效用的有力支撑

要发挥增值评价效用,需要在教育评价实践中充分融入增值评价理念,以发展的眼光看待学生的成长与进步。探索建立适应新时代要求的增值评价,需要科学全面采集学生增值评价相关数据,积累学生学业成绩数据,记录影响学业增值的学生、家庭、同伴、教师和学校相关数据,建立数据库之间的链接,构建能够互联的学生成长发展数据库。从而为建立增值评价模型提供数据支持和技术支撑。

构建教育发展数据库需要相关专业机构组建教育学、心理学、测量学、统计学等相关学科专业团队,制定科学的增值评价指标体系,研发测评工具。大规模实施测评,是一项大体量、技术及平台要求较高的系统工程,具体实施时,可以采用逐步试点的方式,先在小范围开展增值评价的实践,总结经验、优化技术,逐步推广到更多的学校和教育领域,服务广大学校教育教学。在试点过程中,需要注意发挥专业机构的作用,建立科学、客观、公正的评价标准和方法,确保评价结果真实可信、科学准确。

[参 考 文 献]

- [1] 张雅楠,杜屏. 增值评价在美国教师评价中的运用和发展[J]. 全球教育展望,2017(1):67-78+89.
- [2] 边玉芳,王烨晖. 增值评价:学校办学质量评估的一种有效途径[J]. 教育学报,2013(1):43-48.
- [3] 辛涛,张文静,李雪燕. 增值性评价的回顾与前瞻[J]. 中国教育月刊. 2009(4):40-43.
- [4] 吴志华,王红艳,王晓丹. 大规模教育评估的兴起、问题与发展——加拿大教育评估的启示[J]. 外国中小学教育,2011(8):1-5.
- [5] 周园,刘红云,袁建林. 增值评价中学生增长百分位模型及其估计方法概述[J]. 教育导刊,2019(11):61-67.
- [6] 马晓强,彭文蓉,萨丽·托马斯. 学校效能的增值评价——对河北省保定市普通高中学校的实证研究[J]. 教育研究,2006,(10):77-84.
- [7] 任玉丹. 英国学校增值性评价模式对推进我国教育公平的启示[J]. 教育探索,2011(5):152-154.
- [8] 赖寒. 学校增值评价模型概述[J]. 教育现代化,2016(21):247-248.
- [9] 王旭东. 教学质量增值评价常见模型与实践应用思考[J]. 考试研究,2020(5):11-16.
- [10] 周园,刘红云. 教育增值评价中嵌套数据增长百分位估计方法探析:多水平线性分位数回归模型的应用[J]. 中国考试,2020(9):32-39.
- [11] 边玉芳,林志红. 增值评价:一种绿色升学率理念下的学校评价模式[J]. 北京师范大学学报(社会科学版),2007(6):11-18.
- [12] 王家美,戴海琦,周延. 教育增值评估的实证研究——以江西省上饶市30所高级中学语文学科的评估为例[J]. 中国考试(研究版),2009(9):3-9.
- [13] 范美琴,高柳萍. 基于中考和高考成绩数据的高中学校教育教学效能增值性评价[J]. 中国考试,2019(10):6-13.

- [14] 邵越洋,刘坚. 增值评价:关注学校为每一位学生的成长助力——以北京市某区教育实证数据为例[J]. 中国考试, 2020(9):40-45.
- [15] 王晓华. 基于 PISA 数据的学校增值评价研究[J]. 教育发展研究, 2021, (10):8-17.
- [16] 刘宝存,黄秦辉. PISA 高绩效地区学生的全球素养:个体和学校因素的影响[J]. 华东师范大学学报(教育科学版), 2023(2):38-52.
- [17] 杜晓梅,郝春东. 对中小學生进行增值评价的探究[J]. 教育教学论坛, 2021(29):13-16.

Comparative Study of Mathematical Models for Evaluating Academic Quality Enhancement: A Case Study of Empirical Data from Education in a Western District

Zhong Dongmei Lu Jinyun Jia Ling

(Chongqing Education Evaluation Institute, Chongqing 400020, China)

Abstract: Value-added evaluation is an evaluation method that focuses on students' growth and aims to evaluate school effectiveness more fairly. There are various mathematical models for value-added evaluation. Compared with two models based on average rank and standard deviation, the advantages of models based on student growth percentile and multilevel linear regression are that they can avoid the problem of limited value-added space for excellent students. In the multilevel linear regression model, uncontrollable student factors such as gender and family socio-economic status index can be included to effectively eliminate the influence of student factors on students, which is more in line with the evaluation concept of focusing on the "net effectiveness" of schools.

Keywords: academic quality; value-added evaluation; average rank; standard deviation; student growth percentile; multilevel linear model

[责任编辑:刘力]