

增值评价：基于中高考数据的学校质量发展绿色评价模式

周业宇 叶剑祥 雷根生

摘要：为探索增值评价在区域教育质量评价中的实际应用效果，选取 25 所学校共 10 721 名高三年级学生的中考与高考成绩作为样本，综合运用结果评价和标准分法、效应量法以及 SGP 法等多元化的增值评价方式，以评价不同学校的教育质量表现，进一步验证了增值评价方法的可靠性和有效性。但增值评价也具有一定的局限性，需理性地对待增值评价结果，同时还需加强技术赋能教育质量增值评价，不断提升教育质量增值评价的准确性和有效性。

关键词：增值评价；区域数据；学校质量

教育评价事关教育发展方向，教育行政部门和学校都非常重视教育评价的作用，但目前使用的评价方式和标准还较为单一，多为通过计算平均分、及格率、优秀率、升学率等数据，基于终结性评价指标对学校效能和教师教学成果进行评价。这种评价不仅会导致“为考试而教”，而且会阻碍教育均衡发展的推进^[1]。为扭转当前教育评价的不科学倾向，坚决克服“五唯”的顽瘴痼疾，中共中央、国务院于 2020 年 10 月印发《深化新时代教育评价改革总体方案》，提出“坚持科学有效，改进结果评价，强化过程评价，探索增值评价，健全综合评价，充分利用信息技术，提高教育评价的科学性、专业性、客观性”^[2]的主要原则。本研究基于区域范围内的中考与高考（以下简称“中高考”）数据，从结果评价和增值评价两个层面深入剖析，旨在探索一种更加全面、绿色的学校教学质量发展评价方式，以期通过评价促进区域教育质量快速发展。

一、增值评价的内涵

与传统评价相比，增值评价不再单纯地将结果作为评价的唯一标准，而是将结果与之前的基

础作比较，通过二者的差距来判定评价对象的进步水平^[3]。增值评价通过学生在特定时间段内的当前成绩与基线成绩的差距，来评估学生在特定时间段内的成绩增长情况。用动态的评价视角，考量学生、教师和学校所取得的进步，能显著提升评价的科学性、客观性和公平性。它不仅是评估教师、学校及其所在区域工作质量的有效工具，而且对提高教育质量、促进教育公平、优化教育教学管理、改进教学实际等方面具有重要参考价值。

这种方法侧重于观察学生在一段时间内的相对进步，帮助教育评价者从学生的绝对水平中分离出反映优质教学成果的实践，如高效利用课堂时间、清晰和结构化的课程设计、教师的专业发展、纪律良好的课堂环境等。

近年来，国内外学者已普遍认识到在评估教育主体和教学项目时，采用增值方法作为考试成绩平均分的替代指标的重要性。全球各地的学校正逐渐采纳并实施增值评价和学生标准化测试，以便更有效地监控和评估学校的教育质量。其中，较常使用且易实践的增值评价方法有 4 种，包括标准分法（Standard Score）、分层回归法（Hierarchical

作者简介：周业宇，浙江省丽水中学，正高级教师、特级教师；

叶剑祥，浙江省丽水市教育招生考试中心主任，高级教师（通信作者）；

雷根生，浙江省丽水市教育招生考试中心副主任，高级教师。

Multiple Regression)、效应量法(Effect Size)以及学生成长百分位法(Student Growth Percentile),其计算原理、适用范围以及优劣势如表 1 所示。

二、增值评价的研究与结果

(一) 数据来源

本研究采用的数据来源于 2023 年某市普通高中毕业生。为确保样本的代表性与有效性,剔除了中高考成绩不完整的样本,并排除了高考录取人数不足 20 名的学校。经过筛选,最终得到的有效样本 10 721 名。这些学生分别来自 25 所学校,涵盖了重点高中、特色高中和民办学校,且均参与了 2020 年的中考以及 2023 年的高考。鉴于高考成绩涉及个人隐私,且权威部门未予公开,本研究在高考成绩的处理上,采用了学生所录取专业的最低投档线作为替代指标。

(二) 统计方法

增值评价的 4 种方法中,分层回归法计算比较复杂,需要较高的统计分析技能,同时非专业教育评估者很难理解和接受,因此本研究使用了标

准分法、效应量法和 SGP 法计算学校增值效应并作对比。研究中,效应量法采用 Cohen's *d* 值(效应量小中大参考指标依次为 0.2、0.5、0.8),中高考原始分使用 Excel 软件计算单个学生的标准分(*Z*),按 $G = 70 + 10 \times Z$ 转换成标准分数,以中高考成绩标准分数为前测成绩(x_1)、高考成绩标准分数为后测成绩(x_2)构建模型,再用 Excel 软件计算单个学生的 ΔZ 和 SGP,进而用 SPSS26.0 软件和 Excel 软件得到不同学校以及不同特征群体的 ΔZ 、*ES* 和 SGP 及排序。

(三) 研究结果

本研究得到中高考的描述性统计结果、增值评价与结果评价的学校排序结果,以及 3 种评价方式与学校生源因素的相关分析结果。

1. 描述性统计结果

从学生层面分析,经过高中阶段的学习,学生的学业能力出现了进一步的分化现象。在总体平均水平保持稳定的情况下,分数的分布范围呈现出扩大的趋势,表现为最高分数基本不变、最低分数显著下降、标准差增大,见表 2。这一现象反

表 1 4 种增值评价方法的比较

	计算原理	使用范围	优势	劣势
标准分(<i>Z</i>)法	$\Delta Z = \frac{x_1 - \bar{x}_1}{S_1} - \frac{x_2 - \bar{x}_2}{S_2}$ x_1 、 x_2 为 2 次考试的分数, $\bar{x}_1$ 、 $\bar{x}_2$ 为 2 次考试的平均分, S_1 、 S_2 为 2 次考试的标准差	不同条件下测试分数间进行比较和分析,预测学生表现	不受测试难度的影响,不同性质的分数稳定、可比性强,不同性质的分数可以相加	统计上的复杂和以负数或小数的呈现, <i>Z</i> 分不易被普通人群理解与接受,可能会扭曲数据的原始分布形状
效应量(<i>ES</i>)法	Cohen's $d = \frac{\bar{x}_1 - \bar{x}_2}{S}$ Pearson's $r = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}}$ S 为考试的标准差, n_1 、 n_2 为 2 次考试的样本数, <i>ES</i> 大小判定标准: 小、中、大的 d 依次为 0.2、0.5、0.8; r 依次为 0.1、0.3、0.5	研究前,根据期望的 <i>ES</i> 可以确定最小样本量。数据收集完成后,计算并报告 <i>ES</i>	<i>ES</i> 能判断显著性差异是否有实际意义,且不受样本大小的影响,可避免“ p 值操控”	<i>ES</i> 的判定受研究的特殊性、研究设计类型、估计指标的使用前提等影响
分层回归法	$y_{ijt+1} = \gamma_{(t+1)} + \beta y_{ijt} + \mu_{j(t+1)} + \varepsilon_{ijt+1}$ i 代表学生, j 代表学校, t 代表年份, γ 为学校的均值, β 为学生在第 t 年成绩的斜率, μ 为学校的随机效应, ε 为学生的随机效应	常用于研究多个变量的相互关系,并考察每个变量的独特贡献	每次测试的样本量不必相同,得到每个学生的增值分,再计算学校增值分的区间	对统计分析技能有一定的要求,增值评价具有天花板效应
学生成长百分位(<i>SGP</i>)法	$SGP = \frac{x_2 - x_1}{x_1} \times 100\%$	适用于评价学生个体,也适用于班级、学校、地区等的评价	实际中更易理解,且较好地解决了优秀学生绝对上升空间的天花板效应	需要较大的样本量以保证每名学校进行有效对比

映出，优秀学生的学习需求得到了基本满足，然而对于学业后进生关注仍有待加强。

表 2 中高考的描述性统计

		样本量	最小值	最大值	均值	标准差
中考成绩	学生	10 721	30.31	89.36	70.17	9.60
	学校	25	50.19	82.70	70.16	8.08
高考成绩	学生	10 721	21.53	88.85	70.04	9.98
	学校	25	52.82	80.16	70.03	6.52

从学校层面分析，不同学校的生源基础存在显著的差异，而这种差距经过高中阶段的培养后逐渐缩小。具体表现为学校的最低分数有所提高，最高分数有所下降，同时标准差也在降低，见表 2。实际上，相较于优质高中，生源基础相对薄弱的学校在促进学生学业发展方面取得了更为显著

的成效，其中学校的最高增值得分达到了 10.96 分，见表 3。

2.3 种增值评价结果的对比

25 所样本学校各变量及其评价结果（排序）的情况，如表 3。在结果评价排序中，高考成绩和中考成绩列代表 2 次考试成绩及排序的结果。在增值评价排序中，标准分法、效应量法和 SGP 法代表使用这 3 种方法计算出的增值及排序的结果。

与传统的结果评价排序相比，增值评价呈现出显著的特点，为学校整体表现的评估提供了更加全面和深入的视角，有助于更加客观地评估学校在教学质量、学生成长及综合素质提升等方面的实际表现，从而为学校的改进与发展提供有力支

表 3 25 所样本学校的中高考与增值评价排序结果

学校	学校类别	中考成绩		高考成绩		标准分法		效应量法		SGP 法	
		分数	排名	分数	排名	分数	排名	分数	排名	分数	排名
校 1	1	71.21	9	73.47	6	0.2249	5	0.23	5	4.14	5
校 2	1	82.70	1	80.16	1	-0.2538	23	-0.25	23	-2.91	23
校 3	1	76.36	5	75.30	4	-0.1056	16	-0.11	16	-1.19	16
校 4	1	79.94	2	77.65	3	-0.2295	21	-0.23	21	-2.80	22
校 5	1	74.36	7	72.78	8	-0.1588	18	-0.16	18	-2.04	19
校 6	1	75.48	6	73.35	7	-0.2130	19	-0.21	19	-2.54	20
校 7	1	66.72	13	69.83	12	0.3112	3	0.31	3	5.64	4
校 8	1	78.63	3	77.70	2	-0.0935	15	-0.09	15	-0.95	15
校 9	2	69.97	11	70.90	10	0.0924	8	0.09	8	2.25	9
校 10	2	54.13	24	65.09	16	1.0960	1	1.10	1	20.84	1
校 11	2	64.43	17	65.33	15	0.0906	10	0.09	10	1.85	12
校 12	2	66.25	15	66.35	14	0.0107	14	0.01	14	0.77	14
校 13	2	70.36	10	71.14	9	0.0779	11	0.08	11	2.16	10
校 14	2	77.35	4	74.84	5	-0.2512	22	-0.25	22	-2.62	21
校 15	2	61.73	22	63.44	20	0.1710	6	0.17	6	3.61	6
校 16	2	63.48	18	64.04	18	0.0564	13	0.06	13	1.30	13
校 17	2	68.95	12	69.86	11	0.0913	9	0.09	9	1.93	11
校 18	2	50.19	25	52.82	25	0.2634	4	0.26	4	5.95	3
校 19	2	65.23	16	61.97	22	-0.3251	24	-0.33	24	-4.99	24
校 20	2	57.80	23	60.98	23	0.3178	2	0.32	2	6.23	2
校 21	2	73.11	8	69.19	13	-0.3924	25	-0.39	25	-5.29	25
校 22	2	62.00	21	60.52	24	-0.1474	17	-0.15	17	-1.81	18
校 23	3	62.44	20	63.12	21	0.0683	12	0.07	12	3.26	7
校 24	3	62.76	19	63.91	19	0.1151	7	0.12	7	2.99	8
校 25	3	66.45	14	64.27	17	-0.2179	20	-0.22	20	-1.43	17

学校类别说明：1— 重点高中，2— 特色高中，3— 民办学校

持。若将结果评价作为衡量学校优劣的标准，则往往会出现“强者恒强”的现象，例如校 2 的中考和高考均名列榜首。当采用增值评价的方式更全面深入地评估时，可以发现这所学校在助力学生成长方面并未达到理想效果，其排名位于倒数之列；而校 10 在中考排序仅居第 24 位，其增值评价排序跃居第 1。

为更直观地比较增值评价与结果评价应用于学校评价的差异，绘制了样本学校的学业成绩评价和增值评价排名变动分布情况图（如图 1），增值评价拓宽学校整体表现的维度，与结果评价存在较大差异。

结果评价的结果显示，25 所样本学校在两次考试中的成绩排名变动相对较小。具体而言，有 12 所学校的排名位次之差为 1 名，5 所学校的排名没有变化，排名变动的最大值为 8 名。相较于结果评价，增值评价在揭示学校整体表现时展现了较大的差异和更为丰富的信息。首先，成绩排序靠后的学校在增值评价中表现出色，而成绩排序靠前的学校则在增值评价中表现不佳，说明增值评价更为全面地反映了学校的进步和发展潜力。其次，增值评价的排名变动幅度显著增大，变动范围由 $[-6, 8]$ 扩大至 $[-22, 23]$ ，这显示了增值评价在揭示学校间差异方面的敏感性和准确性。再次，3 种增值评价方法对学校进行评价的结果一致性较好，特别是标准分法和效应量法得出的增值排序高度一致，且这两种方法与 SGP 法也呈现出相对的一致性，超过 84% 的学校在 3 种评价方法中的位次之差在 1 名以内，这进一步验

证了增值评价方法的可靠性和有效性。

三、结论与讨论

基于以上研究，在区域教学质量评价过程中，教育行政部门可采用结果评价与增值评价相结合的方法，既关注学业成绩等直观指标，又重视学校的成长和进步，精确地反映学校教育和教师教学的实际成效，减轻学校间的比较压力，纠正“唯生源论”的片面看法。同时，学校也可以利用增值评价这一工具，让学生、教师都能清晰地看到进步的轨迹，从而有效激发他们学习、工作的积极性和创造力，推动学校整体教育质量的持续提升。

通过实施增值评价，能够对不同层次学校的进步情况进行持续的追踪与评估，全面挖掘学校的独特优势和潜在能力，并据此提供定制化的支持与指导方案，有针对性地进行教学改进，以推动学校特色化和高质量发展，有效地缩小县域间和学校间的教育差距，促进重点高中、特色高中和民办学校之间的和谐发展、错位发展和特色发展。

通过实施增值评价，持续跟踪和分析学生的增值情况，能够及时发现教育过程中存在的问题，更科学地制定教育政策，合理分配教育资源，确保每位学生都能享受到适合的教育。为不同的教师和学校设定各自的发展目标，激发良性竞争，推动教育资源的共享与优化，促进经验的交流与借鉴，让更多学生享受到优质的教育资源。

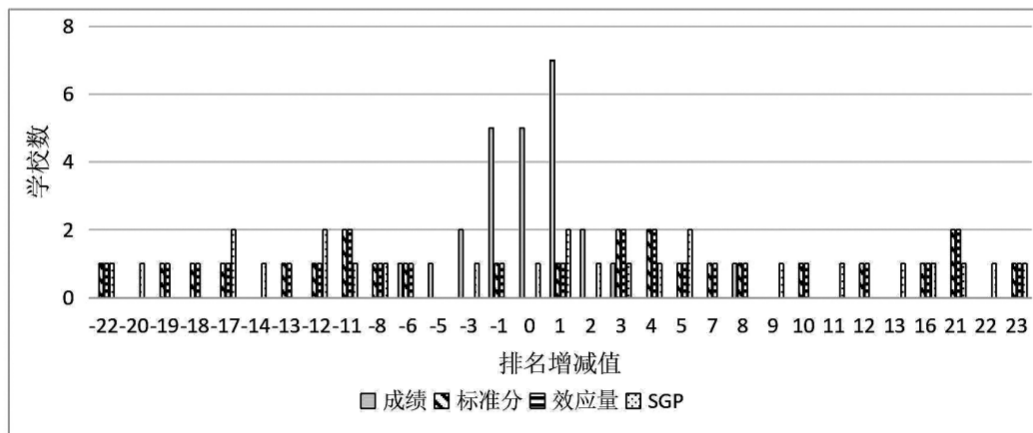


图 1 25 所样本学校的学业成绩评价与增值评价排名变动分布情况

（下转第 64 页）

(上接第 60 页)

虽然增值评价能关注学生、教师和学校进步,使得评价更加科学、客观且公正,然而,若增值评价仅局限于关注学生考试成绩的提升,而忽视了其学习过程中的其他成长与进步,那么这种评价方法也将不可避免地陷入“唯分数论”的狭隘误区。在进行区域教育质量增值评价时,仍需注意两点:第一,理性地对待增值评价结果。教育质量的影响因素错综复杂,增值评价不是“万能药”,借助多种评价方式对教育发展事实进行综合分析和判断,要回归教育现场,找到“治病良方”^[4]。调查发现,尽管一些教育质量相对较弱的学校在增值评价中取得了比结果评价更好的成绩,但其教育质量在公众心中的认可度依旧偏低。第二,关注

技术赋能教育质量增值评价体系的完善与发展。增值评价中要运用统计分析技术,评价水平的提高依赖于统计分析技术的发展。要有效化解增值评价工具中的技术瓶颈,就必须做到智能技术与增值评价的融合创新^[5]。智能成绩管理系统与网络阅卷平台等技术工具,能够辅助构建涵盖学生、教师及学校发展的数据追踪系统,分析预测学生整体发展趋势与未来发展方向,进一步促进区域教育质量快速发展。此外,还可以通过引入第三方评估机构,构建与区域教育质量发展实际需求更为契合的分析模型与评价工具,提升区域教育机构的数据采集、处理及分析能力,确保增值评价所需数据的精确性与可信度。

参考文献:

- [1]王帅鸣,姚守梅,王梦,等.基于成长百分等级模型开展增值评价的实证研究[J].中国考试,2020(9):46-51.
- [2]中共中央 国务院印发《深化新时代教育评价改革总体方案》[EB/OL].(2020-10-13)[2024-04-28].http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202010/t20201013_494381.html.
- [3]关丹丹,韩宁,章建石.立足“四个评价”、服务“五类主体”进一步深化高考评价改革[J].中国考试,2021(3):1-8.
- [4]王家祺,周园,王梦璐.区域教育质量增值评价结果运用探析——以济南市部分区域学生增值评价为例[J].中国考试,2023(3):52-59.
- [5]刘志军,范韵婧.我国增值评价研究:现状、问题与展望[J].中国考试,2023(4):35-41.