

本科生学业增值评价方法的 应用、比较与反思*

——以某市师范专业为例

冯丽,陈薪硕,万华

(贵州师范大学教育学院,贵州贵阳 550025)

摘要:通过对G省S市大学师范专业本科生专业课程学习进行增值评价的探索 and 比较,发现在四种增值评价方法的应用结果中既显示出整体上的差异性与同质性特征,又凸显出个体间的个性化特征。据此得出应遵循评价方法特性,提高增值评价的实用价值;拓宽学生评价视野,发挥增值评价的育人功能。文章建议:建立智能技术支持体系,助力增值评价实践;探索生本增值评价,聚力赋能个性评价。

关键词:本科教育;增值评价;学生评价;教育评价

中图分类号:G647,G40-058.1 文献标识码:A 文章编号:2095-3380(2024)01-0056-06

Application, Comparison and Reflection of the Methods of Value-added Evaluation for Undergraduate Education Students

—A Case of Teacher Major of a City

FENG Li, CHEN Xinshuo, WAN Hua

(School of Education, Guizhou Normal University, Guiyang, Guizhou 550025)

Abstract: Through the exploration and comparison of the value-added evaluation of professional course learning of undergraduate teaching major in the university of S city in G province, it is found that the application results of the four value-added evaluation methods not only show the difference and homogeneity of the whole, but also highlight the individual characteristics. Therefore, it is concluded that the practical value of value-added evaluation can be improved by following the characteristics of evaluation methods, and the educational function of value-added evaluation can be given play to by broadening the horizon of student evaluation. This paper puts forward the suggestions of establishing intelligent technical support system to improve value-added evaluation practice, and exploring student-based value-added evaluation to empower personality evaluation.

*基金项目:2022年贵州省高等学校教学内容和课程体系改革项目“师范生专业课程增值评价的教学改革研究——以《微格教学理论与实践》课程为例”(2022069/0422100)。

收稿日期:2023-10-31,修回日期:2023-11-28

Keywords: Undergraduate education; Value-added evaluation; Student evaluation; Education evaluation

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》(以下简称《总体方案》),明确提出“改进结果评价,强化过程评价,探索增值评价,健全综合评价”,旨在打破以往“唯分数”“重结果”的评价模式,强调更科学有效的教育评价,以迎合建设高质量教育体系与培养高素质人才的时代需求。通过《总体方案》可以看出,增值评价是新时代深化教育评价改革的新要求,是新时代教育高质量发展的新诉求。如何正确应用增值评价、保障其最大实施效度,需要开展深入的研究。

近年来,增值评价已然成为教育领域的研究焦点,我国学者从不同视角对其进行探究,主要涵盖理论研究与实践应用两方面。张亭予从学生视角出发,提出增值评价是一种新的教育评价方法,学生增值评价可以准确掌握学生成长状态、科学预估其发展水平并记录其增值轨迹。^[1]张辉荣、刘丹认为可以从两个方面来认识增值评价的内涵,一方面,增值评价是一种学校效能或教师效能评价方法,另一方面,增值评价是对学生自身成长评价的依据。^[2]辛涛、张文静等认为增值评价具有潜在的诊断性功能,实现了关注点的变化,保障更加公平的比较,与绩效责任紧密相连,且能满足所有学生的需要等特点。^[3]周园、刘红云将多水平线性分位数回归估计应用于增值评价,对某市31所高中学校的数学成绩进行评价。^[4]王旭东指出教学质量增值评价的常见模型可以划分为五类九种,并将这些模型的应用举例介绍。^[5]

通过对已有研究的梳理发现,国内学者对学生增值评价的理论研究较为丰富,但实践层面更多是以一所学校或是区域内多所学校等中宏观研究为主要对象,对学生个体的微观研究相对较少,评价分析与反馈忽视了学生个体差异,无法凸显学生个体增值。同时,研究大多采用单一的增值评价方法,不同方法的比较研究相对较少。鉴此,本研究尝试用多种方法对学生个体增值进行评价,旨在通过比较与分析,揭示各种方法的内在特性,突出学生个体的增值差异,促进增值评价实践探索的不断丰富与完善。

一、四种增值评价方法介绍

本研究主要对两类四种评价方法进行研究,一种是传统评价方法,即基于原始分排名法,是最为熟知、沿用最久的评价方式;另外三种评价方法是在众多增值评价方法中选择出的相对简单易懂、易操作且能关注个体的评价方法。

1. 原始分数排名

原始分数最为典型的应用在高考制度。原始分数使用广泛的原因在于对考试成绩计算与呈现方面简单易懂,符合公众对分数呈现的相对公平性的诉求。但也存在严重缺陷,一是可比性差,不同科目与场次测试之间的分数难以对比,受试题难度、考生水平等影响极大;二是可加性不足,即单科或单项分数对于总分的贡献权重会受难度影响。原始分排名法是指基于学生测试的原始分数以班级或年级为一个考察整体进行排名,依据排名的高低做比较,不考虑学生的自身基础、科目差别、内容难易差距,单以每一次的原始成绩进行排名,并以排名先后评价学生学习的优劣,显然是不客观、不准确与不公正的。

2. 标准分法

事物比较的前提是有共同的参照和标准,为此将标准分法引入教学管理。标准分是依据统计学与测量学原理,将原始分按照一定规则转化为具有相同意义、单位以及共同参照点且能揭示学生分数在群体中所处位置的分数。具体方法是将数据集中的每一个数据减去平均值后除以标准差。将学生当前的成绩与前一次或基线成绩相减,便能得出学生的增值分数,这是一种计算学生增值成绩的直观简便的方法。因为两次成绩的分布位次已经发生变化,即使分数相同,但位次已然不同,所以要将两次成绩进行标准化处理。计算公式如下:

$$\Delta Z = \frac{X_2 - \bar{X}_2}{S_2} - \frac{X_1 - \bar{X}_1}{S_1}$$

上述公式中, Z 表示标准分, ΔZ 是指前后标准

分的差值, X_1 、 X_2 是指学生在两次测试中个人所得的原始分数, \bar{X}_1 、 \bar{X}_2 是指两次测试中所有学生的平均分, S_1 、 S_2 是指两次测试分数的标准差。当 $Z=0$ 时,表示学生分数等于总体平均分;当 $Z>0$,说明学生分数大于总体平均分;当 $Z<0$,则相反;这是标准分的重要特性。标准分法能反映出各个学生成绩在全体学生成绩中的位置,也能比较某一位同学在两次或多次测试中所得分数的优劣情况。因此,标准分可以用于横向与纵向比较。横向比较包括对同一学生不同学科成绩的比较,对不同学生同一学科成绩或总成绩的比较,对不同班级或年级同一学科成绩的比较;纵向比较是对同一学生同一学科在不同阶段的终身比较。^[6]

3. 效应量法

效应量又称为效应大小或效应尺度,是反映研究中处理因素与反应变量间关联大小的无量纲统计量,与测量单位无关、具有单调性且不受样本容量影响的一种衡量效应大小的指标。其中 Cohen's d 是度量两个平均值之间差异的标准化指标,可用于确定两组数据之间的差异情况。效应量应用于学生成绩评价,可不受内容与时间的限制进行多重比较,允许独立于原始成绩的比较。效应量的主要优势是独立量纲,借此可对学生学业成就的各种不同影响做相关比较。

杨德发提出“效应量可以定量描述教学干预对学业成就的影响,对于教师而言,可以运用效应量来计算自己教学的效应量,从而评估教学效果”。效应量多用于群体间,计算公式如下:

$$\text{效应量} = \frac{\text{平均值(后测)} - \text{平均值(前测)}}{\text{分布(标准差)的平均值}}$$

根据上述公式,个体效应量的计算公式转化如下:

$$\text{个人效应量} = \frac{\text{个人成绩(后测)} - \text{个人成绩(前测)}}{\text{分布(标准差)的平均值}}$$

效应量比较大都是应用于群体之间,个人效应量的计算是建立在每名学生对整体方差的贡献相同的假设上,将合并分布(标准差)作为每名学生的估计量。效应量的值越大代表差异越大,即增值越大。它可以对不同科目、不同班级以及不同时间的测试进行比较,为教师评价学生成绩提供了另一种思路。

4. 百分位等级法

百分位等级是指将原始分数分成1-100个等级,再通过公式计算出某一个原始分数所对应的百分位等级,实质上是一种累积百分比,其含义是在参加本次测试的整个群体中低于该原始分数的学生占比。

百分位等级法与标准分法既具有相似性也存在差异性,相似性体现在二者均是对学生个体成绩在全体学生成绩中所处位置的计算;差异性则是具体计算方法不同。计算公式如下:

$$PR = 100 - \frac{100R - 50}{N}$$

上述公式中,PR称为百分等级,R代表某一个学生的原始成绩在全体学生中的排名,N是全体学生的总人数。这种评价方法直接反映出学生所处的相对位置,用以评价其学业成绩,能减少利用原始分数直接比较所造成的误差,排除了不同测试的试卷难易程度的影响。在实际运用中,可用于对学生前后两次测试的表现进行对比,判断学生的变化,学生由此知道自己在班级或年级中所处的相对位置,教师借此了解每个学生的具体变化情况。

二、四种方法评价结果分析

本研究的目的是通过四种评价方法的实践应用,对比结果的异同,从而厘清四种评价方法的特性,拓展学生评价视野,为增值评价的探索应用提供参考。

以普通高校师范专业核心课程“师范生微格教学理论与实践”的学习增值为研究对象,分别在课程开始之前和结束之际对同一位教师教授的91位学生微格教学技能的水平进行测试。两次测试内容、评分标准以及评分人员均相同,前测与后测成绩由3位专家综合评定得出,对于评分差距大于5分的同学的成绩进行仲裁,最后形成学生的前后测成绩,以此作为判断每位学生学业增值情况的数据基础。

评价标准参考中小学教师资格考试中面试环节试讲方面的评价流程与标准,结合专业人才培养目标和课程教学大纲,将师范生需要重点掌握的微格教学技能区分为语言、导入、板书、讲解、提问、评价、教学组织、结束八项核心技能,并采用特尔菲法

对各项技能的权重、分值与评价指标进行讨论、修订并最终确定评分标准。依据学生学号排序分别将其编号001-091,分别采用原始分排名法、标准分法、效应量法以及百分位等级法对91名学生的学业成绩进行评价,并进一步分析不同评价方法下学生学业成绩的变化情况。

1. 四种评价方法的异同性分析

表1出示了编号前10位学生在课程开始前后两次测试成绩用不同方法进行评价排序的结果,其中传统评价排名中的前测成绩与后测成绩是指依据学生个体的成绩在该批学生成绩中的排名,增值评价排名中的标准分法、效应量法与百分位等级法是指利用这三种方法计算出学生个体成绩增值的排名。

表1 编号前10位学生两次测试成绩
传统评价与增值评价的排序

学生编号	传统评价排名		增值评价排名		
	前测成绩	后测成绩	标准分法	效应量法	百分等级法
001	68	86	83	70	66
002	42	68	70	63	74
003	4	4	51	66	48
004	15	32	61	66	65
005	53	22	24	27	15
006	5	43	81	84	82
007	46	85	87	84	83
008	78	64	22	19	27
009	87	83	29	16	39
010	55	88	89	87	80
……					

(1) 排名差值差异性分析

四种评价方法得到的学生排名呈现出不同程度的差异。综合差值的平均值与最大值可知:一是传统评价中前后两次原始成绩排名变化幅度较大,最大差值为66,差值的平均值为23.84;二是三种增值评价方法的排名结果与传统评价中后测成绩的排名结果差距明显,效应量法、标准分法、百分位等级法与后测成绩之间的差值最大值依次是76、73、51,差值平均值分别是25.36、20.7、20.45;三是增值评价中的三种评价方法也存在差异,但相对较小,其中差距最小的是标准分法与效应量法,其后依次为标准分法与百分等级法、效应量法与百分等级法,其间差值最大值分别为19、25、27,差值平均值

分别是5.65、6.56、9.2。

(2) 排名结果的同质性分析

部分学生在两种或三种评价方法下的排名结果相同。一是两种评价方法下排名结果相同的共有24名学生。其中标准分法与效应量法排名结果相同的最多,共8名学生;而后是效应量法与百分等级法、标准分法与百分等级法,均有5名学生;标准分法与后测排名有3名学生,前后测排名有2名学生,效应量法与后测排名有1名学生。二是三种评价方法下排名结果相同的共2名学生。首先是编号为014的同学,前后测排名分别为第18与84名,三种增值评价方法下的排名结果相同,均为第90名。其次是编号为015的同学,后测排名、标准分法及效应量法的排名结果相同,均为第91名,前测排名26,百分等级排名89。上述2名学生的学业成绩大多位于班级后50%,尤其是014和015这2名学生,均位于班级末尾。因此,出现同质性的原因可能是学生的增值情况在班级中所处水平趋于稳定。

2. 四种评价方法的个体性分析

增值评价的本质是观测进步,集体进步是由个体成员的进步所构成的,因此,客观、合理地反映个体增值变化和趋势显得尤为重要。个体增值测量与结果的阐释具有一定的“情境性”,即不同的测量方法会得到不同的结果阐释。

(1) 不同评价方法下的“第一名”

标准分和效应量评价方法下的第1名相同,其余评价方法下出现的第1名不同。首先是编号036的同学,前后测排名分别为第1与16名,三种增值评价方法下排名分别是第75、84、60名。其次是编号为048的同学,标准分与效应量法中排名第1,前后测排名为第90与46名,百分位等级排名第10。再次是编号049的同学,后测排名第1,而前测排名第3,三种增值评价方法下排名分别为第23、37、45名。最后是编号为091的同学,百分位等级排名第1,前后测排名为第85与23,而标准分和效应量排名分别是第3和第5。

(2) 不同评价方法下的结果差异

四种评价方法下学生个体的排名结果差异明显,以下四位同学最具代表性。一是编号为009的学生,前后测排名为第87与83名,三种增值评价方

法下的排名依次为第29、16与39名。二是编号为034的学生,前测排名第2,后测排名第5,三种增值评价方法下的排名分别为第56、71与51名。三是编号为040的学生,前后测排名由第12名变成第9名,三种增值评价方法下的排名分别是第40、57、44名。四是编号为075的学生,前后测排名为第69与68名,三种增值评价方法下的排名结果分别是第35、30、47名。上述学生的原始分数均有所增加,而排名后退,传统评价与增值评价的排名结果差异明显。

三、结论与建议

通过四种评价方法在本科生学业评价中的应用与比较可以得出以下几点结论。

一是遵循评价方法特性,提高增值评价的实用价值。

首先,各评价方式各具特性。一是各评价方式之间差值绝对值的平均值与最大值均存在不同程度的差异,其原因在于不同评价方法特点不同、侧重各异,传统评价中的原始分排名是依据分数排名,增值评价是对比前后两次测试的增值情况排名,标准分和效应量法侧重于标准差,百分位等级法是以平均数为基准。部分同学在不同评价方法下的排名结果一致,是由于增值评价基本原理相同,凸显学生增值情况。二是上述四种评价方法下的“第1名”各异,不同评价方式下同一名学生的成绩情况既可能名列前茅也可能居于末位,在班级中所处层次变化明显,尤其是传统评价与增值评价之间的差距,这是因为不同评价方法各具特性。其次,根据四种评价方法的具体应用,优缺点突显。一方面是传统评价,依据原始分数排名通俗易懂,操作简单。缺点在于过分注重学生分数,忽视个体差异,且缺乏可加性与可比性,应用局限。另一方面是增值评价,上述三种增值评价方法较为简单、易于理解、便于操作,可用于学生个体的评价。当然也有不足,标准分法的缺点是“极端”分数,即零分转换的科学性备受质疑;效应量法的缺点是公式较为复杂,难以理解,鲜少用于个体评价;百分位等级法的缺点是无法解决不同层次学生的分类评价问题。综上可知,增值评价虽更为科学、客观与公正,但其共同缺点是要多且高,多数教师对其产

生“可望而不可及”的心理,严重削弱了增值评价的实际应用。根据比较而言,以班级或年级为单位的本科生学业增值评价使用标准分法相对较好,但不可否认仍有局限,因而还需去探寻真正能用于课堂教学的增值评价方法。

二是拓宽学生评价视野,发挥增值评价的育人功能。

目前有关增值评价应用的研究大多依托庞大的数据库,以区域、学校为单位,样本量成千上万。以大规模数据调查为基础的增值评价有利于检验地区教育质量水平与学校办学质量与效能;但却忽视了个体性,只能看见整体的进步与成长,无法关注到教师及学生个体。就教师而言,一方面,难以获取大量数据,缺乏数据支撑,无法应用主流的增值评价方法;另一方面,与区域或学校教学质量相比,教师更关注自己班级学生的学业情况,更需适用于小样本数据的增值评价方法。就学生而言,每一个学生都有其自身特性,大规模增值评价湮没了学生个体特征,难以实现“尊重学生个性差异”的教育理念。近年来,逐渐有学者或教师认识到以上不足,开始关注增值评价在课堂教学中的应用,例如,李彦婷以个案分析的形式,将增值评价应用于学生个体,关注学生的情感与个性差异,帮助教师有的放矢地给予学生帮助,促使学生学业实际增值;^[7]王安琪对增值评价运用于英语课堂进行了探究,认为英语课堂里的增值评价可以让学生看见自己的进步,让社会看见老师的努力。^[8]此类基于教师教学探讨增值评价方法的应用弥补了以往增值评价个性缺失的不足。

关于增值性评价,本文提出以下两点建议。

一是建立智能技术支持体系,助力增值评价实践。教师是评价学生的主要对象,探索增值评价不应只停留于理论研究及区域应用,更应注重学生个体情况,将应用主体回归教师。因此,应从多方面为教师提供支持,以推动增值评价的实践应用。第一,教师要正确认识增值评价,认识增值评价的价值与工具使用,在应用中掌握其实践特性,厘清优缺点。虽然与传统评价相比,增值评价更加难以理解和操作,但却是推动教育评价改革的重要举措,教师要克服畏难情绪,去了解、认识、学习、应

用增值评价。第二,提供相应的培训支持,为教师开设增值评价相关的培训课程,通过培训使教师认识、掌握增值评价的操作,切实推动增值评价的 applications。第三,构建增值评价应用帮扶平台,针对教师在实际应用中遇见的困难与问题提供帮扶,帮助教师建立使用增值评价的信心。第四,建立增值评价案例库,教育行政部门应牵头收集增值评价实践应用效果较佳的优质案例,给教师提供参考与借鉴,促进增值评价实践发展。

二是探索生本增值评价,聚力赋能个性评价。加强“以学生为本”的增值评价方法或技术的探索,结合时代特性,将数字技术引入增值评价,促使增值评价易理解、易操作、注重“以学生为本”,既为教师的实践应用提供便利与支持,也为学生个体学业增值评价聚力与赋能。其一,注重发展以学生个体为中心的增值评价,探索适用于小样本数据的科学有效且实用的增值评价方法。其二,鼓励增值评价软件开发,鼓励多方参与,促进多方联动,基于增值评价的理论内涵,开发出能专门计算增值评价的软件,有效解决增值评价操作难的问题。其三,搭建数据共享平台,对内,给教师提供数据支持,实时探索和应用增值评价,保障学校内部各部门、学科、教师之间的信息畅通,有利于公正、科学、有效地实施学生评价;对外,促进各级学校之间信息沟通;增强国内外增值评价领域内专家与一线教师之间的交流合作,促进理论与实践的完美契合。这样做有利于引进多元主体参与增值评价,接受来自各方面的监督与评价,促进增值评价科学公正、公开透明的持续发展。

总之,增值评价是对传统评价的融合与补充,是对培养全面发展人才的适切回应,是实现教育高质量发展的重要举措,进一步发掘与探索增值评价将是未来很长一段时期内教育领域的主线任务。

参考文献

- [1] 张亭亭. 如何评价职业学校教师教学的有效性: 基于学生知识图谱增值模型策略[J]. 中国职业技术教育, 2022(17):49-54.
ZHANG Zhuyi. How to Evaluate the Effectiveness of Teaching of Teachers in Vocational Schools: a Strategy Based on the Value-added Model of Students' Knowledge Graph [J]. China Vocational and Technical Education, 2022(17):49-54.
- [2] 张辉蓉, 刘丹. 基于改进的学生学习增值评价的模式构建及应用研究[J]. 中国考试, 2023(6):1-10.
ZHANG Huirong, LIU Dan. Model Construction and Application of Improvement-based Value-Added Assessment of Student Learning [J]. Journal of Chinese Examinations, 2023(6):1-10.
- [3] 辛涛, 张文静, 李雪燕. 增值性评价的回顾与前瞻[J]. 中国教育学刊, 2009(4):40-43.
XIN Tao, ZHANG Wenjing, LI Xueyan. On the Use of Value-added Assessment in Retrospect and Prospect [J]. Journal of the Chinese Society of Education, 2009(4):40-43.
- [4] 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.
ZHOU Yuan, LIU Hongyun. Exploration of Growth Percentile Estimation Method for Nested Structural Data in Educational Value Added Evaluation: Application of Multilevel Linear Quantile Model [J]. Journal of the Chinese Society of Education, 2020(9):32-39.
- [5] 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.
WANG Xudong. Common Models and Practical Thinking of Value-added Evaluation on Teaching Quality [J]. Examinations Research, 2020(5):11-16.
- [6] 徐子仪. 标准分在数学教育评价和管理中的应用[J]. 黄冈师范学院学报, 2001, 21(5):23-24.
XU Ziyi. Application of Standard Scores in Mathematics Education Assessment and Management [J]. Journal of Huanggang Normal University, 2001, 21(5):23-24.
- [7] 李彦婷. 增值评价对学生和教师发展的意义[J]. 内蒙古教育, 2021(4):25-28.
LI Yanting. The Significance of Value-added Evaluation to the Development of Students and Teachers [J]. Inner Mongolia Education, 2021(4):25-28.
- [8] 王安琪. 增值评价在英语课堂中的运用探究[J]. 海外英语, 2021(12):171-172.
WANG Anqi. Exploration of the Application of Value-added Evaluation in English Classrooms [J]. Overseas English, 2021(12):171-172.