

学业评价班级间比较平均分的限度与突破

王振华

(四川文理学院 教师教育学院, 四川 达州 635000)

摘要:在基础教育阶段,学业评价班级间比较呈现出的学业状况差异备受关注,基于深入分析平均分在学业评价班级间比较中具有重要分量的基础上,着重从学业评价政策、统计学、评价学三个角度厘清学业评价班级间比较平均分的限度。突破平均分的限度,需要确保学业评价政策不折不扣地落实;借助信息技术,收集学业评价班级间比较的多元信息;摆脱学业比较的“量化导向”,注重过程中的质性分析;唤醒教师的实践智慧,拒绝学业比较的固化思维。

关键词:学业评价;班级;比较;平均分

中图分类号:G645

文献标志码:A

文章编号:1674-5884(2018)06-0012-06

长期以来,在基础教育阶段通过学业评价班级间比较呈现出的学业状况差异,备受教务管理者、教师、家长的重视,教务管理者以此区分班级学业的优劣,教师以此作为改进教学的重要反馈,家长主要以此进行择校择班。当然,站在提升教育质量、促进师生共同发展的角度,关注班级间学业状况的差异无可厚非,但班级间的学业差异采用什么样的方式进行梳理和呈现是颇值得关注的议题。众所周知,班级间学业差异的梳理和呈现方式具有很强的导向性,对教师的教学产生了深刻的影响,特别是在应试的环境中,班级间的学业比较很容易滑向重分数轻分析的状态,这不得不让人警惕。

1 问题提出

1.1 学业评价班级间比较中的一些现象引人深思

到学校进行视导时,笔者听了一节初中二年级的英语试卷讲评课。L老师拿着卷子说:“这次期中考试题目多简单啊!连听力都没考,考的都是刚学过的语法项目,只要你会背单词和课文就应该能答上。可是你们班的排名却是倒数第三,

与年级成绩相比你们班的平均分低了8.11分,高分率低了6.81%,优秀率低了15.12%,及格率低了12.45%,差生率却高了3.22%,而且排名倒数第一的同学就在你们班……”^[1]为什么班级平均分在评价者的眼里是那样的重要?为什么L老师那么在乎经班级间平均分的比较生成的班级排名和热衷于各项指标的比较?此外,班级间的学业比较还会与教师考核直接挂钩。一位校长表示:“学校规定,如果所教班级平均分与平行班差距在5分以上的就属于教学事故,一年出现两次及以上的教学事故,年终考核即为不合格。”^[2]这所学校为何偏爱使用班级平均分来考核教师的教学业绩?

实际上,一场测验之后,班级间比较平均分确是通行的做法,然而隐藏在平均分背后导致班级间学业差异的因素,才更具深入分析的价值,以期有针对性地改进教师的教和学生的学。但是,比较平均分的惯用做法已弱化了教师对测评背后的教学过程和影响因素进行深入分析的动机,极易陷入追逐分数、分分计较的怪圈。教务管理者受制于班级间比较平均分的思维定势,会忽视班级间学业差异的多元化表达,没有强烈的意愿去寻

收稿日期:20180619

基金项目:四川省教育发展研究中心项目(CJF17057)

作者简介:王振华(1982-),男,河南洛阳人,讲师,硕士,主要从事教育统计与测评研究。

找除平均分之外使用更多的信息来描述班级间学业状况的不同。

1.2 研究的主要问题

基于上述事实与思考,本文尝试在深入分析平均分在学业评价班级间比较中具有重要分量的基础上,着重从学业评价政策、统计学、评价学三个角度厘清学业评价班级间比较平均分的限度,进而探寻如何突破平均分的限度,为开展学业评价班级间的比较提供更多的思路和借鉴。

2 平均分在学业评价班级间比较中的分量

2.1 平均分在学业评价班级间比较中从未缺场

学业评价班级间比较,平均分是从不缺少的比较指标。第一,平均分计算简便。一个班级的学业量化结果就是一组数据,计算班级平均分相当容易,即使是手工计算也用不了多少时间。平均分从量的角度反映了整个班级学业的一般情况。第二,平均分极易比较。即使一个年级有许多班级,但每个班级学生学业数据的平均分只有一个,班级之间比较学业平均分极易操作,完全是最简单的数学方法,比较大小一目了然。第三,平均分易于记录。班级间学业比较的结果需要用文本和图表进行表述,一个班级的学生学业状况浓缩为一个数据,全年级各个班级的学业数据可以清晰地呈现在一幅表格之中,无疑会方便整理和记录,形成文本描述。正是平均分的便捷性才使得其在学业评价班级间比较中备受青睐,具有极为重要的影响。

2.2 学业评价班级间比较中平均分的技术担当

在学业评价班级间比较中,平均分的技术担当主要体现在:第一,平均分精确清晰。作为根据全班学生学业数据计算得来的平均分,它精确清晰地描述了一个班级学生学业的总体水平。开展比较,没有比精确的数据更具准确性的比较指标了。第二,平均分简洁明了。面对班级之间众多的学业数据,寻找一个简洁的量来反映整个班级学业数据的分布状况,进而与其他班级进行比较,班级平均分是最容易被考虑的指标。平均分把全班学生学业数据整合为一个数据,简洁明了地反映了班级学业数据的整体分布状况。第三,平均分客观灵敏。平均分的客观性在于它是综合了班级里所有学生的学业数据计算得到,不因任何人

的想法而改变。同时,即使仅有一名学生的学业数据发生了变化,班级平均分亦会改变,平均分极为灵敏,与每位学生的学业数据息息相关。

2.3 平均分在学业评价班级间比较中的知识基础澄明

学业评价班级间比较是团体与团体的比较,从数据的角度而言,必须找到能够代表团体水平的量,才能形成班级间学业比较的数据基础,而班级平均分正好满足了这一需要。在测验、测量或调查中获得的大量数据,具有一种向数据中央某一点靠拢的趋势,这种趋势在统计学中称为集中趋势,它是数据分布的特征之一。用于描述数据集集中趋势的量数称为集中量数,平均数就是集中量数中最常用的一种,它是一组数据的代表值,用以说明一组数据分布的一般水平,可用于各组数据之间集中水平的比较^[3]。这从统计学的角度解释了为什么平均分能够作为班级学业水平的代表值参与比较。再者,有学者认为,现代课程与教学制度安排之初,就已经决定了成绩比较排名的必然性。成绩比较排名与当下教学制度在知识原理上有内在的亲缘关系,它的盛行是制度知识逻辑的当然的实践展开,或者说它本身就是当下课程与教学制度的有机组成部分,它们是同谋的关系^[4]。既如此,比较班级学业平均分就有了坚实的知识依据,成了理所应当之举。

3 反思与梳理:学业评价班级间比较平均分的限度

教育的本质在于培养人,教育就是个人获得知识或见解的过程,就是个人的观点或技艺得到提高的过程^[5]。把上述过程量化为描述个人学业的数据,进而整合为全班学生的学业平均分,这就需要剥离纷繁复杂的教育过程,从中选取能够量化的层面,必定无法兼顾更周全的学业信息,所以一直存在着从各个角度对班级学业平均分的考问。基于此,本文主要从以下三个角度深入分析学业评价班级间比较平均分的限度。

3.1 对照学业评价政策,检视其问题

教育部于2002年发布《教育部关于积极推进中小学评价与考试制度改革的通知》,明确强调:“改变长期以来以升学率作为唯一标准评价学校教育质量的做法,小学生的学习成绩评定应采用等级制,不得将学生成绩排队、公布。任何单位和

个人不得以任何形式按中考成绩给地区、学校和学生排队并公布名次。”^[6]2003年,教育部基础教育课程教材发展中心启动了“建立中小学生学业质量分析、反馈与指导系统”项目,旨在对教学过程进行管理和指导,对基础教育质量进行检测^[7]。2010年,中共中央、国务院发布《国家中长期教育改革和发展规划纲要(2010-2020年)》,着重指出:“不得以升学率对地区和学校进行排名,不得下达升学指标。”^[8]2014年,教育部发布《教育部关于全面深化课程改革 落实立德树人根本任务的意见》,明确指出:“统筹课标、教材、教学、评价、考试等环节,主要是纠正考什么教什么、单纯以考试分数评价教育质量的倾向。”^[9]2017年,国务院印发《国家教育事业发展规划“十三五”规划》,着重强调:“注重考查学生适应社会发展和终身发展的能力,防止单纯以升学率考核学校和教师、单纯以分数评价学生。探索实行利益攸关方共同参与的开放式评价,完善评价结果公开机制。”^[10]

基于上述政策,针对学业评价,归纳起来重在强调:一是不得按学业成绩排队,二是注重过程,三是改变单纯以考试分数进行评价。但对接教育现实仍会发现,当前许多学校的班级仍是按照成绩划分的,有所谓的重点班、实验班、普通班等,学校里的分化首先是从对班级的分化开始的^[11]。为什么政策的贯彻遇到了很大的阻力,一个关键的因素就是,当前学业考核过于注重从量的角度进行梳理,这必然会给计算班级平均分提供数据支持。当然,计算班级平均分也是从整体上量的角度梳理班级学业的重要方式,但以下问题不得不慎重考虑:其一,由平均分的比较产生的班级学业排名,给任课教师带来的是更多的压力还是更大的动力;其二,追逐班级间平均分的比较极易产生“结果导向”,忽视对学生学业过程的分析;其三,仅是单纯以分数来评价班级间学业的差异,不但违背了评价政策的要求,还会导致分数至上、分分计较。

3.2 借助统计学的知识,分析其局限

在特定条件(如平行班)下,平均分在一定程度上可以用来比较或说明不同群体之间的水平差异,但是,由于影响平均分的因素有很多,如试题难度、考试组织与评卷、学生基础与能力水平、考生规模、学风班风、教师水平与学校管理等都会对

学校或班级的考试平均分造成影响,因此,一般情况下,平均分不宜直接作为成绩或质量评价的指标^[12]。从统计学的角度来看,平均数能够反映一组数据分布的集中趋势,如果要清楚地描述这组数据的分布状况还需要说明其离中趋势(离散程度),即这组数据具有的偏离数据中央某一点(即平均数)的趋势,通常使用标准差来说明离中趋势。对于班级学业评价量化结果而言,就形成了通过一定的测评方式获得的一组数据,使用平均分就是在描述这组学业数据的分布状况,但如果仅是使用平均分就等于说只是关注到了这组学业数据的集中趋势,并没有说明其离散程度,显然没有描述出班级学业数据分布状况的全貌。

实质上,当从量的角度整体上对班级学业状况进行分析时,必须关注的是班级里所有学生的学业数据,进行所有数据分布状况的整体描述。鉴于上述分析,按照统计学对数据分布状况描述的方式,应当分别使用平均数和标准差来反映班级学业数据分布的集中趋势和离中趋势。简言之,当从整体的角度描述班级学业量化结果时,既要关注平均分,又要关注标准差,两者兼顾才能清楚地说明班级学业数据分布的主要特征。需指出的是,一组数据平均数的代表性受其标准差的制约,如果标准差越大,说明平均数的代表性越差;反之,标准差越小,平均数的代表性就越好。在多组数据比较时,当平均数相同或相近、标准差不同,或是标准差相同或相近、平均数不同时,均不能贸然下结论说各组数据分布状况相同。因为无论哪种情况它都只表明数据的分布情况,并不能判断出数据分布的优劣^[13]。从这个意义上来说,学业评价班级间比较如果只是简单地比较平均分就对班级间学业进行排序,显然是武断的做法。借助统计学的知识,审视学业评价班级间比较平均分,其局限主要有:第一,无法描述班级间学业数据分布离散状况的差异;第二,没有班级学业数据的标准差作支撑,平均分的代表性缺失了相应的基础;第三,平均数是统计学的基础概念,与其他概念的联系密切,如果仅是直接单纯地比较班级平均分就割裂了统计学的知识逻辑。

3.3 依据评价学的理论,考量其不足

价值是“评价”这一术语的根基,评价理论从“第一代评价”发展到“第四代评价”的过程中,评价理论的价值基础经历了从价值屏蔽到价值确定

与唯一,再到关注多元价值的过程^[14]。依据“第四代评价”理论,评价就是对被评事物赋予价值,其创立者美国著名教育评价专家库巴和林肯认为评价本质上是一种通过“协商”而形成的“共同的心理建构”,评价并不是“外在于人的”“纯客观”的过程,而是参与评价的所有的人,特别是评价者与其对象双方交互作用,形成共同的心理建构的过程。人头脑中的“建构”是以人们的价值观为基础而形成的,在价值观多元化的社会里,评价活动需综合考虑如何融合或沟通各方利益相关者的意见^[15]。当前,“第四代评价”理论受到了教育工作者的普遍认同,成为分析和改进评价工作的理论基石。

上述“第四代评价”理论,其核心要义在于:第一,推崇价值多元;第二,倡导平等协商;第三,强调共同建构;第四,注重评价过程。对接当前学业评价班级间比较平均分的做法,会发现在严酷的班级平均分比较的数据面前,其偏离了“评价不是为了证明,而是为了改进”的价值追求,过度关注班级间学业的差异,针对性的改进被悬置,评价者把自己的价值观上升为开展评价的唯一标尺,忽视了评价中的其他价值观念。同时,评价是一个涉及各方面利益相关者的复杂过程,真正实现《国家教育事业发展规划“十三五”规划》中所述“探索实行利益攸关方共同参与的开放式评价”,就需要在评价中充分聆听来自各方面的声音,不断协调各种分歧,平等协商共同建构,达成共识。显然,目前学业评价班级间比较平均分是一种自上而下、单向接受管理、过于看重分数的操作模式,亟须改进。此外,当前学业评价班级间比较平均分的评价过程已形成了严格的、固定的程序,弹性和灵活性不足,缺少对评价过程的系统反思,过于重视评价结果。

4 突破平均分在学业评价班级间比较中的强势影响

限于当前大班额班级教学还未从根本上得到缓解,加之任课教师面对繁重的教育教学工作,收集和梳理班级整体状况的学业信息难免简单化、单一化,采用班级学业平均分或是高分段学生学业平均分进行班级间比较就成了通行的做法。同时不得不承认,班级学业状况是一种客观的存在,对其梳理和分析会带有评价者主观的意愿,加上

目前班级学生人数普遍偏多,受制于各种因素,推进学业评价班级间比较的改革还会遇到不少困难,寻找撬动改革的有力支点,必须深入学业比较实践,探寻班级间学业比较中可松动的缝隙,找准适合实践变革的突破口。

4.1 确保学业评价政策不折不扣地落实

第八轮基础教育课程改革启动以来,评价始终是关注度最高的话题之一,在国家层面颁布了一系列关于学业评价改革的政策文件,地方教育行政部门也发布了具体的改革举措。但在应试的教育现实面前,政策到底落实得怎么样、如何结合实际更好地对接政策的要求,始终在考问着各级教育工作者。2013年6月,教育部印发《教育部关于推进中小学教育质量综合评价改革的意见》,仍着重指出:“单纯以学生学业考试成绩和学校升学率评价中小学教育质量的倾向还没有得到根本扭转,突出表现为:在评价内容上重考试分数忽视学生综合素质和个性发展,在评价方式上重最终结果忽视学校进步和努力程度,在评价结果使用上重甄别证明忽视诊断和改进。”^[16]2017年颁布的《国家教育事业发展规划“十三五”规划》再一次重申上述现象,表明学业评价政策走进评价现实并非易事,不折不扣地把学业评价政策转化为具体的评价操作,还面临许多挑战,主要克服的困难在于:其一,评价活动的提出者、推动者、实施者和使用者在多大程度上愿意放弃旧观念,接受评价政策的指引主动改进评价工作;其二,如何兼顾与平衡评价活动涉及的各方面利益相关者的思考和立场。

4.2 借助信息技术,收集学业评价班级间比较的多元信息

避免学业评价班级间比较陷入只是单纯地比较班级平均分的境地,收集学业评价班级间比较的多元信息是有力的应对之举。实际上,收集和描述整个班级的学业状况可以从很多方面切入:第一,从时间节点来看,学期前、学期中、学期末最为常见,还可选取某些周、月份;第二,从评价主体来看,除了学校和教师主导的评价外,还需收集来自学生、家长的评价意见;第三,从评价内容来看,既重视纸笔测验形成的评价信息,还兼顾学习过程中的状况;第四,从评价方式来看,量化式的班级学业信息备受关注,质性的评价资料也分量十足。由此产生了一个新的议题:面对这么多纷

繁的班级学业信息,如何整合在一起进行班级间比较呢?近年来,互联网+、大数据等信息技术的迅猛发展,为解决上述议题提供了技术支持。在教育领域,时时刻刻产生着各种数据,教育大数据就是教育统计插上了信息化的翅膀,评价者通过对学习者学习过程的关注与跟踪,在大数据技术的支持下提供最客观、直接和准确的学习分析和教育评价,使传统的经验式评价转向基于数据的过程性评价^[17]。随着信息技术的不断进步,学业评价中的多元信息收集、处理和分析将不再是难题。

4.3 摆脱学业比较的“量化导向”,注重过程中的质性分析

深受“科学实证主义”的影响,通过量化的方式描述班级学业状况一直被认为是严谨精确的做法。的确,班级学业状况转化为相应的数据,呈现出的是简洁的数字,符合精确梳理评价信息的需要。但一个严肃的问题必须厘清,班级学业评价使用数据表达的同时如何避免走向“量化导向”,即只要评估班级学业状况,就想方设法将学业信息进行量化。显然,如果热衷于追逐学业评价班级间比较平均分,量化思维会主导比较全过程,带来的是附着在冷冰冰的数据上的高利害影响。只要走进学生学业过程,便会发现鲜活多样的学业信息,除量化的学业数据外,对学生学业过程的质性描述同样含有客观、原始、朴素的学业元素。学生接受教育、学业获得提升的过程极其复杂,仅仅使用量化的数据不足以显示其全部信息,质性描述恰恰能弥补无法量化的学业信息的梳理与呈现。随着学生评价改革的不断深入,学业考核已从单纯反映学生学习结果的终结性评价向能够揭示学习过程的形成性评价扩展,其根本目的在于提高学生学习的质量,而不是为评价学生等级或给学生打分提供证据^[18]。重视班级学生学业过程中的质性评价,无疑将会丰富班级学业状况的表达,从更周全的层面为班级间学业比较提供评价信息支撑。

4.4 唤醒教师的实践智慧,拒绝学业比较的固化思维

在当前应试的环境中,面对学业评价班级间比较平均分的通行做法,难免会使部分教师形成班级间学业比较的固化思维,这与他们开展学业比较直接比较班级平均分的思维定势相关,长期

惯用这种比较方式容易使他们陷入“自我缠绕”的状态,一旦比较班级平均分构成的学业评价比较成为任课教师的比较观念,他们参与班级学业比较所得经验会不断强化这种观念,从而使自身陷入“观念—现实—经验—观念”的循环。其实,每位教师都会或多或少地对自身接触和所做的学业评价班级间比较进行反思,从中觉察出一些关键性信息,为学业评价班级间比较平均分的合理性寻找支点,这就是教师实践智慧的展现。教师的实践智慧就是指教师对教育合理性的追求,对当下教育情景的感知、辨别与顿悟以及对教育道德品性的彰显^[19]。以此为分析框架,直面应试环境中的广大教师,不禁要追问:在分数至上、分数竞逐的氛围中,在单向度的评价管理模式下,面对繁重的教学工作,到底有多少教师会用心地对学业评价班级间比较平均分进行充分的感知、辨别与顿悟,考问其合理性,并对其本应彰显的对人(学生、教师)的关注的教育道德品性进行反思?事实上,亚里士多德曾明确提到,实践智慧与科学一样,“也相关于普遍的东西”,实践智慧与具体的东西相关^[20]。这样就不必担心经过实践积累的教师没有机会具备相应的实践智慧,有可能他们的实践智慧还处在积累中的沉睡期,唤醒他们,为拒绝学业比较的固化思维开辟道路。

参考文献:

- [1] 王芳.从试卷讲评课看测评的得失[N].中国教育报,2013-05-29(09).
- [2] 张彩云.教师退出的标准在哪里[N].中国教育报,2016-02-23(09).
- [3] 梁红梅,王景英.教育统计学[M].长春:东北师范大学出版社,2005.
- [4] 顾云虎.成绩排名的学校制度知识依据及其批判[J].全球教育展望,2008(5):3-9.
- [5] 柳海民.教育学概论[M].北京:北京师范大学出版社,2015.
- [6] 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)[2018-03-09].http://www.moe.gov.cn/srcsite/A26/s7054/200212/t20021218_78509.html.
- [7] 张民生.建立中小学生学习质量的绿色评价系统[EB/OL].(2013-04-26)[2018-03-09].http://www.moe.gov.cn/jyb_xwfb/s5989/s6634/201304/t20130426_151275.html.

- [8] 国务院. 国家中长期教育改革和规划纲要(2010—2020年)[EB/OL]. (2010-08-02)[2018-03-21]. http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/moe_1485/201008/t20100802_93705.html.
- [9] 高毅哲. 统筹推进立德树人工作——教育部基教二司负责人答记者问[N]. 中国教育报, 2014-04-25(03).
- [10] 国务院. 国家教育事业发展规划“十三五”规划[EB/OL]. (2017-01-19)[2018-04-07]. http://www.gov.cn/zhengce/content/2017-01/19/content_5161341.htm.
- [11] 齐学红. 未来班级发展走向与班主任核心素养构建[J]. 教育科学研究, 2017(2): 19-21.
- [12] 吴金财. 考试成绩分析与评价: 如何使用平均分[J]. 教育测量与评价, 2013(8): 57-60.
- [13] 范晓玲. 教育统计学与SPSS[M]. 长沙: 湖南师范大学出版社, 2005.
- [14] 杜瑛. 西方教育评价理论发展的社会文化基础探析[J]. 教育测量与评价, 2012(10): 22-27.
- [15] 卢立涛. 回应、协商、共同建构——“第四代评价理论”述评[J]. 内蒙古师范大学学报(教育科学版), 2008(8): 1-6.
- [16] 教育部. 教育部关于推进中小学教育质量综合评价改革的意见[EB/OL]. (2013-06-08)[2018-04-14]. http://www.moe.gov.cn/srcsite/A26/s7054/201306/t20130608_153185.html.
- [17] 苏令. 大数据时代教育改革“靶向治疗”来了[N]. 中国教育报, 2016-03-30(04).
- [18] 蔡敏. 课堂评定: 学生学业评价的重要方式[J]. 教育科学, 2012(1): 22-26.
- [19] 邓友超, 李小红. 论教师实践智慧[J]. 教育研究, 2003(9): 32-36.
- [20] 杨文登, 谈心. 教师实践智慧的五种常见误解及其澄清——基于循证教育学的视角[J]. 教师教育研究, 2016(4): 1-7.

Limitation and Breakthrough of Average Score in Comparison of Academic Assessment between Classes

WANG Zhenhua

(School of Teacher Education, Sichuan University of Arts and Science, Dazhou 635000, China)

Abstract: In the period of basic education, the difference which is in comparison of academic assessment between classes draws a lot of attention. Based on the analysis of the importance of the average score in the comparison of academic assessment between classes, this study, from angles of academic evaluation policy, statistics and evaluation, clarifies the limitation of average score in the comparison. Ensuring the policy of academic assessment is implemented, collecting the multiple information of academic assessment between classes by using information technology, getting rid of the “quantitative guidance” of academic comparison and paying more attention to its qualitative analysis, and helping teachers come up with some ways to escape from the rigid and single thinking, which could break the limitations of average score.

Key words: academic assessment; class; comparison; average score

(责任校对 游星雅)