

方法与模型 :引领教育评估的变革

龚春燕,何怀金,贾玲,卢锦运
(重庆市教育评估院,重庆 400020)

摘要:大数据时代,依赖经验,依赖粗放方式、单一手段、定性思维进行学习行为判断和教学决策,这种沿袭自远古的教育评价范式必须改变。基于数据的教育评估,是教育评估最前沿的课题。重庆市教育评估院自2009年成立以来,一直致力于教育数据分析研究,研发教育质量监测工具,建构了分析系统,形成一套较完整的教育数据采集、处理模型或方法,挖掘了多元多层数据,促进了教育评估的改革。

关键词:大数据;教育评估;评估方法;评估模型

中图分类号:G40-058.1 文章标识码:A 文章编号:2095-3380(2016)01-0038-04

DOI:10.13794/j.cnki.shjee.2016.0009

Methods and Models: The Leaders of a Reform of Educational Evaluation

Gong Chunyan, He Huaijin, Jia Ling, Lu Jinyun

(Chongqing Educational Evaluation Institute, Chongqing 400020)

Abstract: In the age of big data, the old paradigm of educational assessment which relies on the experience, extensive mode, simple approach and qualitative thinking to give a judgment of learning behavior and teaching decision-making should be changed. Educational evaluation based on data is the most advanced contents now. Since 2009, Chongqing Educational Evaluation Institute has been working on the research of educational data analysis, developing educational assessment tools, constructing the analytic system, forming methods and models of educational data collection and processing and mining multi-layer data to promote the reform of educational evaluation.

Keywords: Big data; Educational evaluation; Evaluation methods; Evaluation models

笔者曾指出,大数据具有“5V”的特点,即体量大(Volume)、类型杂(Variety)、速度快(Velocity)、价值高(Value)和真实性强(Veracity)等特点。^[1]世界各国对大数据的研究和应用都高度重视,推动了产业升级与新兴产业诞生,引起了科技界对科学研究方法论的重新审视。目前我们接触到的基于大数据的现代教育测评工具都有专业性强、技术过于复杂、实施成本高等问题,不便于其在省或地市县或学校的开展及普及。我们认真

研究脑科学、教育学、心理学、教育测量学、心理测量学、教育统计学、信息技术科学等各领域的理论与技术,并应用这些技术较好地监测和评价融入到日常学校的教学之中,建立了全面的监测评价大数据库,把复杂的测评分析及报告技术转化为易操作的软件,动态跟踪评价教育教学活动的发展状况,评价学生成长、教师专业发展及学校管理,为改进和实施教育行政决策提供科学依据。

收稿日期:2015-06-25,修回日期:2015-11-30

一、学习与践行 改革传统评价方法

(一) 研究教育测评技术

2009年,作者在教育部基础教育质量监测中心和北师大“认知神经科学与学习”国家重点实验室的帮助下,首次开展了重庆市基础教育质量监测。探索实践证明,基于现代测评技术的监测评价能准确地掌握教育质量现状、全面评价学生成长、探索教育评价规律,为教育决策、学校管理和教师教育教学提供科学依据,从而指导教育研究和改革。但对地市存在以下问题:一是专业性强,专业人员不足。监测工具的研发、大规模数据的采集、量尺分数的合成、监测结果的统计分析,以及兼具科学性和通俗性的报告撰写需要大量专业人员,工作技术含量高、专业性强、需多领域专家共同参与,难以普及和大规模推广。二是工作量巨大,实施成本高。工具研发、数据采集、清理、分析和报告撰写都需要大量专业人才的长时间参与,成本高,难度大,普通区县和学校无法承担。三是结果反馈不及时,无法进行过程性评价。经验表明,监测评价的数据处理流程复杂,专业性强,需要较长时间才能发布结果报告,时效性差,无法及时用于改进教育教学和服务行政决策;同时,其缺乏过程性评价,缺乏对学生个体的评价和追踪,无法有效地改变当前学校仅通过传统考试进行评价的“唯分数”质量观。

针对上述问题,通过与美国、英国、中国香港和台湾,以及北大、北师大、西南大学等高校的各领域专家的合作,尤其是到香港教育学院王文忠(国际著名教育评估监测的专家)工作室学习监测理论与技术,重庆市教育评估院确立了“应用信息技术整合教育测评技术编写相应软件平台,开发专业且简单易操作的软件,采集教学活动数据,构建大数据库,实现监测评价各环节的自动化、智能化,实现监测评价的过程化”的研究目标。

(二) 组建工具研发团队

科学、合理、可操作的评价指标体系和工具量表,是保证监测评价质量的基础。本研究选择了学术水平较强的专家、经验丰富的教育工作者组成了工具研发团队,经教育部监测中心多次培训,组建了语文、数学、英语、科学及创造力、阅读能力、学生心理发展和教师

发展、学校发展等若干专家小组,几年来召开各类研讨会、分析会、工作会150余次,调研、访谈上百次,保证了工具研究的实效性和科学性。

专家组有高校的专家,更多的是来自学校的骨干教师,通过国家级、市级专家队伍的线上、线下的培训,逐步形成了“评价指标体系构建”、“评价工具研发”、“评价工具试用、使用”、“数字化评价平台分析工具质量”、“高校及科研院所专家指导”、“完善评价工具”的工作模式,逐渐建成了可供区县和学校使用的监测工具库。

(三) 实施分层试测

立足重庆大城市、大农村的实际,把重庆教育划分为三个层面,即教育发达区、教育发展区、教育欠发展区,作者选择三个有代表性的样本区县,实施了全样本监测。从2011年到2014年的连续4年中,在全市范围内选择了100所不同层次学校进行了试用,每次涉及学生约8000名,被选学校的所有教师、管理干部和校长全部参测。在2010年和2013年,全市有40个区县1400余所学校、23251名教师、12436名学校管理干部、1400余名校长、145866名学生,共约35万人次参加了监测。同时开展了对全市普通高中、职业中学的全面数据评价。通过试测使用,完善了软件,建立了标准化的工作流程,构建了相应的监测评价机制。同时对区域教育均衡发展、学校办学规模、区域城乡一体化进行试测,逐渐完善模型。

二、方法与模型 创新监测评价手段

结合重庆市的实际情况,建立了一系列教育教学质量测评模型,并将这些数学模型用于实际评价和数据分析。

(一) 义务教育均衡系数测算模型^[2]

$$CEE = \sum_{i=0}^r \lambda_i G_{Bi} + \rho \sum_{j=1}^{n-r} \tilde{\lambda}_j V_{Bj}$$

其中 λ_i 和 $\tilde{\lambda}_j$ 表示指标权重, G_{Bi} 表示采用基尼系数算法计算出的指标基尼系数值, V_{Bj} 表示采用差异系数算法、经线性变换后的指标差异系数值, ρ 表示基尼系数和差异系数间的等值系数,即 $\rho = \frac{\bar{G}_{合格}}{\bar{V}_{合格}}$, $\bar{G}_{合格}$ 表示合格

区县前 r 个指标的基尼系数平均值, $\bar{V}_{合格}$ 表示在剩余 $n-r$ 个指标差异系数的平均值。

(二) 城乡一体化教育资源配置模型^[2]

根据资源类型,分析“粒型”和“块型”资源进行配置,具体算法如下。

1.“粒型”资源配置模型

区域内生均拥有教育资源为:

$$\bar{x} = \left(\sum_{i=1}^n x_i \right) / \left(\sum_{i=1}^n S_i \right)$$

校间生均资源的差异系数为:

$$V = \frac{1}{\bar{x}} \sqrt{\frac{\sum_{i=1}^n (\bar{x}_i - \bar{x})^2 \times S_i}{n}}$$

每所学校对应的生均拥有的教育资源为:

$$\bar{x}_i = \frac{x_i}{S_i} \quad (1 \leq i \leq n-1)$$

学校数为 n , x_i 为第 i 所学校现有资源, S_i 为该校学生数,将 \bar{x}_i 与 \bar{x} 作差后按由小到大排序为 $\bar{x}_1 \cdots \bar{x}_n$,则可以算出第 i 所学校与第 $i+1$ 校的资源差为

$$D_i = (\bar{x}_{i+1} - \bar{x}_i) \times S_i \quad (1 \leq i \leq n-1)$$

则第 i 所学校分配到的资源为:

$$S_{Di} = \sum_{i=1}^{n-1} D_i$$

2.“块型”资源配置模型

每所学校对应的生均拥有的教育资源:

$$\bar{x}_i = \frac{x_i}{S_i} \quad (1 \leq i \leq n)$$

其中 n 为学校数, x_i 为第 i 所学校拥有的某项块型教育资源, S_i 为该学校的学生数。 \bar{x} 为县(区)域内生均拥有教育资源的标准, $\bar{x}_1 \cdots \bar{x}_n$ 为 \bar{x}_i 与 \bar{x} 作差后的序列,设 $S_{fi} (1 \leq i \leq n)$ 为预计达到的学生规模, C_{min} 为最低投入的成本单位,则有每所学校投入的单位成本为 $C_{pi} (1 \leq i \leq n)$,则第 i 所学校达到规划期预计的投入为 $C_i = (S_{fi} \times \bar{x} - x_i) \times C_{pi}$,该学校对应的当前均衡效益指数为:

$$E_i = \frac{S_i \times (\bar{x}_i - \bar{x})^2}{C_i}$$

设该资源可能进行的最小投入为 K ,将向每所学校分配的教育资源(资金)设为 R ,则分配方式如下:

- (1)对 $C_i \leq K (1 \leq i \leq n)$ 的学校不投入;
- (2)对 $C_i > K$ 且 $(S_i \times \bar{x} - x_i) \times C_{pi} < K$ 的学校暂不投入;
- (3)对 $C_i > R$ 的学校暂不投入。

将剩余的 M 所学校按当前均衡效益指数 E_i 由大到小排序,排序后对应学校需投入资源为 C_i , C_{min} 为所有 M 所学校所需投入最少的学校资源, S_c 为已分配的资源总和。第 1 所学校分配资源为 C_1 , $S_c = S_c + C_1$; 当 $C_2 \leq R - S_c$ 时,或者政府能适当增加一部分达到 $C_2 \leq R - S_c$ 时,配给第 2 所学校的资源为 C_2 , $S_c = S_c + C_2$, 否则不分配; $C_3 \leq R - S_c$ 时,或者政府能适当增加一部分达到 $C_3 \leq R - S_c$ 时,分配给第 3 所学校的资源为 C_3 , $S_c = S_c + C_3$, 否则不分配……当 $R - S_c < C_{min}$ 时,如果政府能适当增加一部分达到 $C_{min} \leq R - S_c$ 时,分配对应的学校的资源为 C_{min} , 否则不再分配。

(三) 基于IRT项目反应理论的量尺分数合成模型^[3]

现代心理测量理论中项目反应理论 IRT(item response theory)认为,根据受测者回答问题的情况,可以建立数学模型,推测受测者对不同测验的相对能力情况(量尺分数)。

研究中,采用了适合不同需要的 IRT 能力分数合成数学模型,并设计了相应的软件自动进行估计合成。举例如下。

(1)用于 2 级评分的客观题测量的双参数模型:

$$P(\theta) = \frac{\exp[Da(\theta-b)]}{1+\exp[Da(\theta-b)]} \text{ 或 } P(\theta) = \int_{-\infty}^{a(\theta-b)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

其中 $D=1.702$, μ 为区分度, b 为难度。后者为正态分布,较复杂,常用前一个模型。

(2)用于多级评分主观题的三参数模型,其中 $D=1.702$,

$$P(\theta) = c + (1-c) \frac{\exp[Da(\theta-b)]}{1+\exp[Da(\theta-b)]} = c + \frac{1-c}{1+\exp[-Da(\theta-b)]}$$

a 为区分度, b 为难度, c 为猜测系数。

(四) 多维模型改革教学方法

5 年的研究中,建构了一系列用于监测评价的统

计分析模型,并用软件实现运算。除一般描述分析、差异分析、相关分析、回归分析、聚类分析等外,还有一系列分析模型。

1.结构方程模型

学生成长的背景因素如教师因素、家庭环境因素、学校因素等对学生成长的影响程度分析,采用了结构方程模型。

2.多层线性模型

学生、班级、学校、区县等嵌套的多层因素对学生发展的影响采用多层线性模型进行分析。

3.学习进步评价模型

针对学生个体引入了学习进步评价模型(增值评价),将数字化的学生成长信息,通过构建数学模型对学生的进步程度进行价值判断,实现学生成长过程的量化评价,促使关注全体学生的发展和进步。

除此之外,针对大数据挖掘,还研究实现了C4.5算法、SVM算法、KNN分类算法、朴素贝叶斯模型算法、CART分类法和回归树等统计分析方法的具体

应用。

(五)学校的办学规模测算模型

$$CEE_{\text{规模测算}} = \text{int} \left(\sum_{i=1}^7 \lambda_i \frac{G_i}{\rho_i} \right)$$

CEE_{规模测算}表示最终测算出的核定学校规模,λ_i表示第i项指标的权重,G_i表示第i项指标的原始数据,ρ_i表示第i项指标的日均要求,int(A)表示对人数A取整。表1是根据上述公式对8所学校测算的招生规模。

指标的确立是通过大数据对4年来学校多项办学指标数据与学生质量进行相关分析,挖掘出最影响学生学业质量的7项办学指标:校地面积、校舍面积、实验室(含微机室)、体育运动场面积、专任教师、图书、教学仪器设备值,各项指标的权重根据指标与学生质量的相关系数确定,具体为:

$$\lambda_{\text{实验室}}=0.2 \quad \lambda_{\text{体育运动场}}=0.2 \quad \lambda_{\text{专任教师}}=0.2 \quad \lambda_{\text{校地面积}}=0.15 \quad \lambda_{\text{校舍面积}}=0.15 \quad \lambda_{\text{图书}}=0.05 \quad \lambda_{\text{教学仪器设备值}}=0.05。$$

表1 部分学校根据办学规模测算模型算出的建议招生规模

学校名	学段	目前规模	实验室(m ²)	体育运动场面积(m ²)	专任教师数(人)	校地面积(m ²)	校舍面积(m ²)	图书(册)	教学仪器设备值(万元)	最终核定招生规模(人)	建议招生规模(人)
学校1	小学	1118	147	1000	65	26080	12910	14000	240	943	1000
学校2	小学	1392	867	4868	75	33966	27247	30340	1200	2201	1500
学校3	小学	4593	390	5500	214	18246	20289	45600	50	1563	2000
学校4	小学	724	472	7560	60	12206	13956	10000	25	916	800
学校5	初中	4191	3805	9960	205	106102	65284	6000	344	2994	3000
学校6	初中	971	170	3600	171	13320	8537	7649	60	822	800
学校7	初中	3784	1040	10500	168	71626	44636	65325	576	2359	2500
学校8	初中	4350	555	11332	397	26664	21785	33380	602	2202	2500

上面一系列的模型,在实践中反复运用,促进了教育的变化。校长不再是多招学生,而是注重内涵发展了;区县财政拨款投入由过去拍脑袋变成了分析一系列数据再做决定,有效促进了教育均衡与城乡一体化发展。

参考文献

[1] 龚春燕,何怀金,贾玲,等.重庆模式:大数据评估促进教

育决策科学化[J].中小学管理,2015(8):9-12.

[2] 龚春燕.对义务教育均衡发展系数测算模型与评估思考[J].人民教育,2012(12):10-11.

[3] 龚春燕.大数据:教育监测评估发展[M].重庆:重庆出版社,2014:16-22.

[4] 辛涛.项目反应理论研究的新进展[J].中国考试,2005(7):18-21.