

# 新时代基于高考数据的区域 教育教学评价实践与展望

——以江苏省为例

杨从意，范美琴，朱文琪

**[摘要]** 在习近平新时代中国特色社会主义思想的指导下，我国进行了力度空前的高考改革。高考改革对基础教育，特别是高中教育形成倒逼机制，各项教育教学活动越来越离不开考试数据的分析和支持。高考数据是教育考试机构拥有的最重要资源库，发挥其优势和价值是推进教育治理现代化的一个重要突破口。立足高考数据，面向基础教育开展评价服务是推动区域教育改革发展的的重要抓手，是向服务型考试机构转型发展的必然选择，也是考试评价自身深化发展的重要方向。江苏省区域教育教学评价通过建立科学的评价指标体系、开发智能化统计评价系统、开展多层次数据价值判断，研制出了较为优质的高考评价报告，分为地区版和学校版两类，主要包括生源、成绩、录取概况，考生答题水平分析，近三年成绩比较，质量综述与教学建议等内容，定量与定性分析相结合，汇聚形成区域教育教学评价地图，具有评价内容多元、方式多样、结果细致的特点。经过两年的实践，江苏省已建立起较为稳定的评价指标体系和报告体系，初步实现了对基础教育教学的精准反馈和指导，取得了一定的社会效果和反响。下一步，将通过引入增值性评价，加强成因分析，重视个体评价等，完善区域教育教学评价工作，更好发挥评价的激励、诊断和育人功能。

**[关键词]** 考试评价；高考数据；基础教育教学

进入新世纪以来，国外考试机构纷纷改弦易辙，有意拉远自己与考试(test)的距离，贴近自己与评价(assessment)的距离。国内部分考试机构也积极顺应潮流，成立考试评价研究机构。一时间，“考试评价”成为考试机构的热名词。考试评价有广义和狭义之分。广义的考试评价是一个大概念，指运用考试手段开展教育评价的过程，是国际通行的教育评价主要手段；狭义的

**[收稿日期]** 2018-03-06

**[作者简介]** 杨从意，江苏教育考试院，电子邮箱地址：2320158166@qq.com；范美琴，江苏教育考试院，电子邮箱地址：496151254@qq.com；朱文琪，江苏教育考试院，电子邮箱地址：905553791@qq.com。

考试评价是教育与心理测量学的一个分支,指通过相关技术手段开展考试数据分析,将其结果应用于政策优化和教育改进的过程。运用高考数据开展区域教育教学评价,是狭义考试评价的具体实践,它对推动区域教育改革、促进考试机构转型升级、深化评价工作发展内涵具有重要意义。

## 一、背景和意义

### (一)国家教育改革发展之需

当前世界各国的教育改革正风起云涌,我国也正处于教育全面深化改革攻坚期。一方面,教育正自上而下开展着自我革命。以新高考改革为龙头的考试招生制度改革正在向纵深推进,而新高考改革对基础教育,特别是高中教育改革形成倒逼机制。在这种新的改革形势下,各项教学活动和 Learning 组织越来越离不开相关考试数据的分析和支持。立足于高考数据开展区域教育教学评价,精准诊断学情学力,精确评价教学成果,将对树立科学的教育质量观,重塑基础教育教学生态,对基础教育改革发展起到强有力的推动作用。另一方面,教育与经济社会发展的结合更加紧密。教育的模式、形态、内容和学习方式发生深刻变革,教育治理呈现出新特点。《国家教育事业发展“十三五”规划》明确提出“加强和改进教育统计,推动各级教育行政部门和学校开展深度数据挖掘和分析,运用互联网、大数据提升教育治理水平,更好地服务公众和政府决策”。可见,“大数据”在本质上已经转化成了一种新的思维方式和问题解决办法,发挥高考大数据的优势和价值成为推进教育治理现代化的一个重要突破口。因此,立足高考数据强化考试评价是推动教育发展的重要途径。

### (二)考试机构转型发展之需

以服务对象需求为导向,提供优质测评服务是国外著名教育考试机构生存与发展的根基(戴家干,2008)。受其影响,国内教育考试机构也提出了“从单一考试机构管理向考试机构管理与教育测评服务并重转变,从行政管理型向服务管理型转变”的发展定位,要求努力发挥专业化考试服务机构的重要作用。其中,充分利用独有的数据信息资源积极开发评价服务,是教育考试机构向服务型专业化机构转型发展的重要抓手。高考数据是教育考试机构拥有的最重要资源库,如果对它的分析仅局限在试题评价层面,而忽视了对基础教育的评价和指导作用,这实际上是一种资源浪费,也是教育考试机构的功能性缺失(戴家干,2008)。从著名考试机构的发展路径来看,它们都将基础

教育作为服务的一个重要对象，ETS通过测评服务进入基础教育市场，并拓展包括学习支持体系在内的K-12工程；香港考试评核局全面接受香港教统局委托开展全港基础教育系统评估和学生诊断评估，为全港基础教育改革服务（乔丽娟等，2006）；天津教育考试院也承接了当地的基础教育质量监测任务。可见，服务型教育考试机构建设需要树立促进教育发展的大服务理念，将基础教育作为一个重要服务领域。因此，立足考试数据开展评价服务是教育考试机构转型发展的必然选择。

### （三）考试评价深化发展之需

在当前最全面系统的新高考改革背景下，教育测量和考试评价技术的使用范围越来越广，对考试评价工作提出了更高的要求 and 期待。服务基础教育改革发展是考试评价深化发展的重要方向。教育部考试中心，北京、天津教育考试院等先后都面向基础教育研制了高质量的评价产品。江苏教育考试院自2004年开展高考自主命题以来，便积极运用考试评价改良命题技术，但面向基础教育的评价仅做了一些研发，尚未付诸实践。近年来，江苏教育考试院提出“完善教育考试生态系统，从过去只关注教育考试组织管理这一环节，着力向两头延伸，前期重视考试研究，中间加强组织管理，后期重视考试评价，形成完整的教育考试链”的发展思路。在此背景下，2015年11月江苏教育考试院正式成立科研处，职能之一就是牵头开展考试评价工作，研发面向各类服务对象的评价产品。至此，考试评价工作有了更好的组织保障和运行机制。2016年起，江苏教育考试院启动面向地区和学校的高考评价试点工作，先后在江苏省4市10校开展了基于高考数据的区域教育教学评价实践，取得了较好的社会效果和积极的社会反响。

## 二、设计与研发

为确保区域教育教学评价的高质量、高水准，江苏教育考试院与国内一流的科研团队建立了合作机制，保障工作科学推进。主要经历了以下研制阶段：

### （一）建立评价指标体系

评价指标体系是区域教育教学评价的基础和灵魂。为了实现高考数据分析对基础教育教学的精准反馈和指导，江苏教育考试院本着“科学性、实用性、导向性和前瞻性”的原则，构建了一套较为严密的评价指标体系，力争实现对地区和学校教育教学客观、公正、科学、系统的评价。一是以科学理论指导为基础。评价指标体系建设要有科学的理论作为支撑，使得指标体系概

念上严谨,逻辑上合理,准确抓住评价对象的实质。江苏教育考试院建立评价指标体系的基本依据是现代教育测量学的原理和方法,遵循由易到难的原则,起步阶段主要基于经典测量理论(CTT),而后逐步将新一代的认知诊断理论(CDT)应用于指标体系建设。二是以基层实际需求为目标。面向区域的高考评价,最终目的是为了帮助地区和学校改进教学实践,促进学生发展,助力教师成长,服务教育决策。因此,江苏教育考试院加强需求调研,先后多次赴试点区域开展座谈访谈,充分听取一线教师、管理人员、教研人员、教育行政人员对评价指标体系的意见建议,充分考虑基层实际需求,确保针对性、实用性和有效性。三是以优秀评价项目为指导。江苏教育考试院充分学习和借鉴了国际知名评价项目 PISA、TIMSS 的分析评估框架,学习和参考了北京、天津等同行考试机构面向基础教育开发的评价报告体系,选取具有一定前瞻性的指标作为导向,力争兼容并蓄、全面多元、突出特色。在广泛听取各界意见建议的基础上,结合江苏省教育和考试实际,基本构建起具有鲜明独创性的,符合江苏省高考实际情况、高中教育教学实情和考试评价技术规范的高考评价指标体系。该体系一共包括 5 个一级指标,22 个二级指标和 65 个三级指标。

## (二)开发统计评价系统

开发统计评价系统是实现区域教育教学评价快速化、智能化的关键环节。江苏教育考试院基于教育测量学视角下的大数据分析核心思想,分为整理标准数据、建立数据资源池、设计报告模板、开发评价软件、校对核验数据五个主要步骤来实现。其中,标准数据的整理,力求充分整合命题和阅卷环节的过程数据,将试题考察的知识点、能力要求等关键属性纳入教育考试大数据中心,将考生的报名、成绩、录取数据统筹使用,建立完备的考生数据信息表;数据资源池的建立,基于 Oracle 数据库,依据评价指标体系和教育统计专业算法,自主研发算法模型,根据数千种指标维度组合计算形成数据资源池;报告模板和评价软件采取联动开发模式,将微软旗下的 VB 和 Office 联合应用,采用 Word 建立标准化表格模板,采用 Excel 宏录制绘制各类图形,个别复杂图形采用 R 统计软件绘制,采用 VB 开发面向图表的数据库应用程序,建立可视化、便捷化、智能化的操作界面。

## (三)开展数据价值判断

统计数据背后涌现的规律和效益,远不是数据结果本身所能涵盖的。基于高考数据结果开展价值判断,是了解学生学习过程发生机制、了解教师教学过程存在问题、了解教育现状发展态势的重要环节。主要分两个步骤来实

现,一是组织统计测量专家撰写质量综述,归纳数据统计分析得出的主要结论,提炼考生学业水平表现的主要特征,实现对考生学业表现质量的综合性评判。二是组织学科评价专家撰写诊断性教学建议,通过对数据结果多维度、多角度、多元性的比较、分析和推演,实现对区域的学力诊断、教学诊断和调控诊断,并提出富有针对性的对策和建议。其中,教学诊断要求评价专家尽可能指出教学的薄弱环节及学生的认知盲区,并结合学习条件、教学资源、教学设备等给出教学改进的具体方法和策略。调控诊断要求评价专家根据区域整体教育现状及发展态势,就如何促进高位均衡、特色发展、优质增效等方面给出有效管理策略。可以说,开展数据价值判断是江苏区域教育教学评价的一个亮点工程,不仅给出了高水平的教育教学建议,还给出了如何进一步挖掘数据潜在价值的有效示范,充分践行了“评价是为了改进”的理念要求。从两年的试点效果来看,深受教育管理部门和基层学校的欢迎。

### 三、主要内容

基于高考数据的区域教育教学评价报告分为地区版和学校版两类报告,地区版主要面向各省辖市,学校版则面向各级各类学校,两类报告的主要内容相似,包括生源、成绩、录取概况,考生答题水平分析,近三年成绩比较,质量综述与教学建议几个方面的内容。

其中,生源分析,是通过分析试点区域生源结构状况与特点,帮助基层从教学的外部因素理解考生高考学业水平表现现状与差异;成绩分析,是通过将试点区域成绩与全省同类校、高水平校、全市、全省平均水平等多个比较对象群进行差异性比较,帮助基层找到自己的相对水平位置,发现自身的相对优势与不足;录取分析,是通过综合分析试点区域的上线率、录取率和临界生概况,试图挖掘有提升潜力的目标群体。以下示例成绩分析。

表1 某校各科目原始分概况比较(理科类)

科目	平均分	标准差	差异 系数(%)	优势系数				百分 等级
				同类校	高水平校	本市	全省	
语文	112.34	8.22	7.31	0.24	-0.13	0.70	0.52	92.73
数学	113.01	10.38	9.19	-0.13	-0.60	0.63	0.28	81.31
数学附加	26.70	4.10	15.37	-0.19	-0.48	0.62	0.20	75.72
英语	80.35	8.80	10.95	0.08	-0.52	0.92	0.53	87.48
物理	74.62	12.83	17.20	0.10	-0.40	1.01	0.55	88.68

续表

科目	平均分	标准差	差异 系数(%)	优势系数				百分 等级
				同类校	高水平校	本市	全省	
化学	81.09	9.76	12.04	-0.31	-0.79	0.71	0.25	74.83
生物	75.15	10.61	14.12	0.07	-0.39	0.78	0.51	86.78

注：1. 优势系数，表示学校平均分与比较群体平均分(如全省同类校、高水平校、本市、全省)的标准化差异，计算公式为学校与比较群体平均分差值除以学校标准差。分值大于0，则说明学校与比较群体相比具有优势；分值小于0，则说明具有劣势；数值则表示具体优劣的程度，通常，绝对值在0.2以下认为优势或劣势小，0.5左右认为优势或劣势中等，0.8以上认为优势或劣势大。

2. 同类校，是指基于全省理科总分，将全省普通高中分为一类校前(1/3)、二类校(中1/3)、三类校(后1/3)三种类别，与学校所处类别相同的学校群体。

3. 高水平校，是指理科总分平均分处于全省前10%水平的学校。

4. 百分等级，是指学校各科目平均分在全省所有普通高中各科目平均分分布中所处的百分位置。

依据表1，从差异系数可知，数学附加、物理成绩的离散程度较大。由优势系数可知，与同类校相比，语文最具优势，化学劣势最大；与高水平校相比，各科目均具劣势，化学劣势最大；与本市平均水平相比，各科目均具中等以上优势，英语、物理优势大；与全省平均水平相比，各科目均具中等优势。由百分等级可知，语文为优势学科，超过了全省92.73%的学校；数学附加、化学相对较弱，仅超过全省75%左右的学校。综合来看，语文、英语、物理最具优势，数学附加、化学相对较弱。

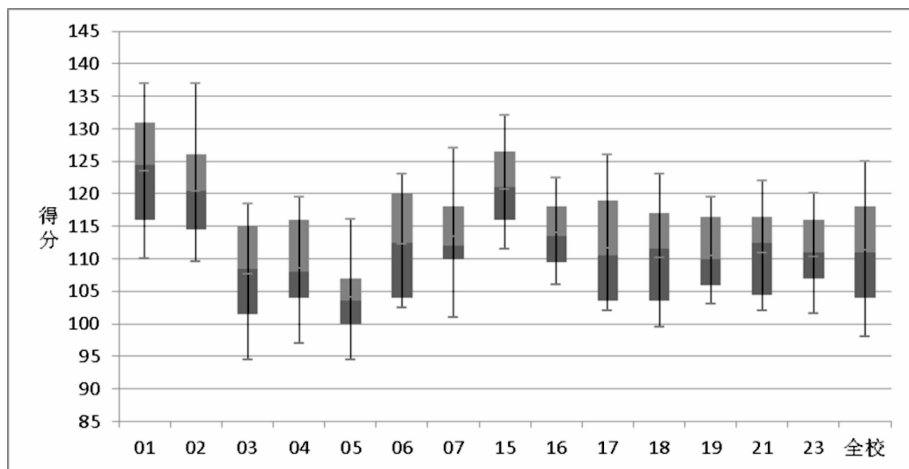


图1 某校各班级考生单科成绩比较(语文)

图1为某校各班级语文成绩分布箱图,箱图由上至下分别标识成绩的90、75、50、25、10百分位数,50百分位数附近的一小短横标识平均分,该图可以反映各班级成绩分布的集中趋势、离散趋势以及偏态信息。从全校平均分所在位置划一条横线,可见01、02、15班语文平均成绩大幅领先全校平均水平,从箱体高度来看,01班虽平均水平最高,但均衡性不及02、15班;05班平均成绩最低,与全校平均水平差距最大,且整体离散程度小;17班几乎与全校平均水平分布一致;04、07班呈现明显正偏态分布,75、50百分位点之间的人数比例明显高于50、25百分位点之间的人数比例;21班呈现明显负偏态分布,50、25百分位点之间的人数比例明显高于75、50百分位点之间的人数比例。综合可对各班级成绩分布形态有整体的认识与了解。

考生答题水平分析,是分科目客观呈现考生在不同题型、不同知识板块、不同试题上的作答情况,帮助地区和学校更好把握考生的具体优势和不足,为改进教学提供参考。以下举例分析。

表2 某校各知识板块得分情况(语文)

知识板块	平均分	差异系数(%)	得分率				优势系数				百分等级
			本校	高分组	中间组	低分组	同类校	高水平校	本市	全省	
语言文字运用	13.62	14.65	0.91	0.96	0.91	0.86	0.24	-0.19	0.38	0.22	78.86
古代诗文阅读	25.00	15.77	0.68	0.75	0.69	0.60	0.44	-0.52	0.53	0.32	79.00
现代文阅读 (文学类)	10.96	20.19	0.55	0.60	0.55	0.49	0.31	-0.43	0.43	0.23	74.82
现代文阅读 (论述类)	11.49	16.78	0.64	0.68	0.64	0.59	0.10	-0.47	0.31	0.10	59.67
写作	44.32	15.09	0.63	0.71	0.63	0.55	-0.25	-0.84	-0.03	-0.27	32.11

注:高、中、低分组是指基于学校考生单科成绩的两个百分位数  $P_{33.3}$  和  $P_{66.6}$  将考生分成前1/3(高分组)、中1/3(中间组)、后1/3(低分组)。

由表2可知,该校现代文阅读(文学类)作答得分率最低、离散程度最高;“语言文字运用”板块得分率最高、离散程度最小;每一知识板块高、中、低分组得分率依次降低。由优势系数可知,前三知识板块相对于同类校、本市、全省均具中等优势,与高水平校有差距;“现代文阅读(论述类)”与同类校、全省平均水平相当,较本市有中等优势,弱于高水平校;“写作”较四个比较群均具劣势。由百分等级可知,五个板块中“现代文阅读(论述类)”表现较弱,“写作”表现最弱。综合来看,考生在各知识板块表现水平不均,主要是“写作”拖了后腿。

表3 某校单题得分情况(数学)

题号	分值	平均分	得分率					优势系数		
			本校	高分组	中间组	低分组	同类校	高水平校	本市	全省
13	5	1.18	0.24	0.45	0.20	0.06	-0.04	-0.22	0.34	0.20
14	5	0.55	0.11	0.24	0.06	0.02	-0.02	-0.15	0.23	0.13
15	14	13.73	0.98	1.00	0.99	0.95	0.00	-0.05	0.34	0.22
16	14	12.43	0.89	0.93	0.91	0.83	0.09	-0.08	0.54	0.40
17	14	11.06	0.79	0.90	0.84	0.63	-0.11	-0.40	0.63	0.38
18	16	7.71	0.48	0.70	0.43	0.31	-0.07	-0.38	0.58	0.32
19	16	5.56	0.35	0.45	0.37	0.22	-0.03	-0.29	0.59	0.36
20	16	2.14	0.13	0.28	0.09	0.03	-0.01	-0.25	0.41	0.28

由表3可知,该校在第13、14、20题得分率较低。从高、中、低组各题得分率来看,第14、20题主要拉开了高分组与中间组的差距,第17、19题主要拉开了中间组与低分组的差距。从与同类校的优势系数来看,各题作答水平与同类校平均水平相当。从与高水平校的优势系数来看,第13、17、18、19、20题与高水平校平均水平差距较大,达到中等程度劣势,其中第13、20题是难题,与高水平校有差距容易理解,第17、18题难度并不大,但与高水平校差距较大,考虑有可能与考生作答时间的把控有关。

表4 某校不同层级考生选择题受选率统计(英语)

题号	A				B				C				D			
	全部	高分组	中间组	低分组	全部	高分组	中间组	低分组	全部	高分组	中间组	低分组	全部	高分组	中间组	低分组
30	34.06	45.06	32.11	24.94	16.26	7.95	15.93	24.94	22.25	18.07	25.00	23.73	27.18	28.43	26.72	26.39
32	6.80	5.78	6.62	7.99	15.78	7.95	16.18	23.25	35.36	31.57	35.78	38.74	41.99	54.46	41.42	30.02
35	14.64	7.71	16.18	20.10	68.77	86.27	69.85	50.12	10.19	4.10	8.58	17.92	6.39	1.93	5.39	11.86
43	1.62	0.48	0.98	3.39	5.74	1.45	5.64	10.17	91.34	97.83	92.40	83.78	1.21	0.24	0.98	2.42
70	3.88	1.21	3.92	6.54	76.46	92.05	80.64	56.66	15.86	5.54	12.75	29.30	3.56	1.21	2.21	7.26

注:阴影标注为各题正确选项。

由表4可知,高、中、低组考生在各题正确选项上的受选率均依次降低,说明考生在这5道试题上的作答水平与总分表现水平一致。30题D项、32题C项均超过了自然猜测率25%;30题D项高分组选择率还高于中低组,说明高分组对D项所考查知识点的掌握程度还不及中低组;35题A项、43题B项、70题C项受选率明显高于其他诱选项,高、中、低组考生选择率均依次增加,这些信息教学上可重点关注。



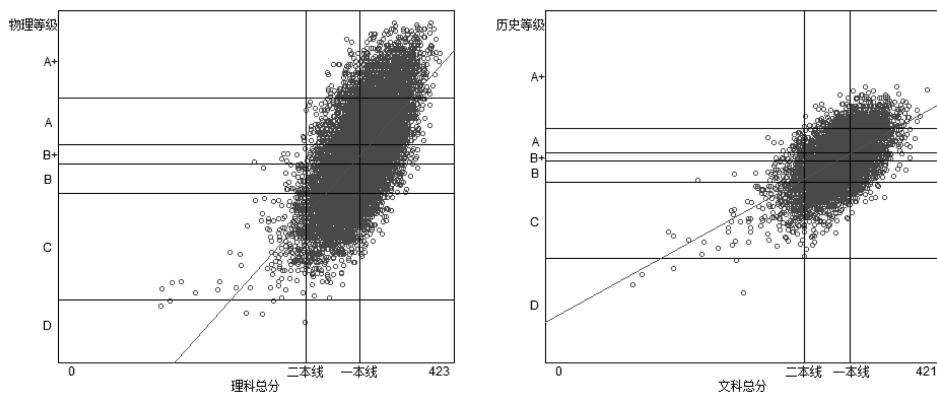


图2 某市选修科目等级水平与录取总分匹配分析(物理、历史)

图2左图为某市理科必选科目物理等级分与理科总分的匹配分析图,右图为文科必选科目历史等级分与文科总分的匹配分析图,横坐标为语数外总分,纵坐标为选修科目等级分。由左图可以看出,整体上来看,该市考生物理等级分与理科总分之间呈现出正相关趋势,等级分与总分匹配良好;从细节上来看,仍有少部分总分达一本线的考生物理等级分为B以下,也有少部分总分落在二本线下的考生物理等级分为B以上,局部存在总分等级分不匹配情形。右图反映的匹配信息与左图类似。左、右图比较来看,理科等级分与总分相关趋势线的斜率更大,趋势线上升得更陡峭,匹配度更高。

除此以外,报告中还包含近三年成绩比较、质量综述与教学建议等内容,由于篇幅关系,不再一一举例。值得说明的是,江苏省在评价报告的推介阶段还专门设计专家解读环节,邀请国内教育测量一流专家亲赴试点地区和学校举办评价报告现场解读会,普及统计指标和方法,并针对试点区域统计分析结果开展“一对一”分析讲解,帮助基层快速掌握读懂和使用报告的科学方法,鼓励结合教育经验和学科背景开展深度挖掘,实现评价报告价值最大化。同时,也将现场解读作为完善优化评价工作的重要途径,促进双改进、双提升。

#### 四、总结与展望

江苏省基于高考数据的区域教育教学评价初步取得成效,建立了一套较为完备、系统的评价体系,推出了一批较为丰富、实用的评价成果。归纳起来有三个特点:一是评价内容趋于多元,既包括常规的报名数据、成绩数据,又包括命题过程性数据和录取数据,内容丰富、全面多元。二是评价方式趋

于多样,既有专业的统计分析,又有精准的价值判断,定量评价与定性评价相结合,汇聚形成区域教育质量地图。三是评价结果趋于细致,既可实现知分知位,知晓本地本校的考试表现及相对位置;又可实现知长知短,知晓教育教学的长处和短板;还可实现知进知退,知晓教育教学的进步和退步,使得考试数据真正产生化学效力,将“基于数据说话”的现代教育治理方式向前推进。当然,目前的评价也有许多需要拓展和深化的领域。

一是引入增值评价。报告中的“年度比较”是为了反映区域教育教学质量的变化趋势。因为不同年份的试题试卷、考生、授课教师等都不一样,无法直接比较最终的考试结果。故采取常模比较的办法,即通过比较区域不同年份位于全省高、中、低分组考生比例和不同年份各选修科目等级考生比例的变化趋势来概括和评价教育教学质量发展趋势,这是一种较为简单、粗糙的办法。下一步,将引入增值性评价来科学评价区域教育教学质量的年度变化趋势。这是一种起源于美英的、国外通用的学校效能评价模式,通过构建严谨规范的测量模型,来测算每个地区或学校在提升学生学业成绩中发挥的平均“净效应”,将其界定为衡量学校教育教学质量的有效指标。在此基础上,建立面向区域的增值评价体系,研制增值评价报告,全面呈现区域增值的整体表现、分类表现和变化趋势,关注学生的成长与进步,有效解决因起点不同带来的评价不公问题,引导地区、学校之间更趋公平合理的比较,凸显中等水平学校的优势,激发薄弱学校转变和发展的动力,更好发挥评价的激励作用。

二是加强成因分析。目前报告主要提供地区或学校在各比较组群中的相对位置,在各类别考生上的表现水平差异,在各板块试题上的作答水平高低等较为宏观的信息,对比精确反馈和精准指导的要求还有差距。下一步,将致力于甄别影响考生学业表现水平的主要因素及形成机制,致力于回答“考生学业表现水平为什么高或低,高或低表现水平考生所在的学校教育教学有什么特征”,旨在揭示考生学业水平表现构成的“黑匣子”,为学校管理和教师专业发展提供精准支撑。这种影响因素的剖析应该是多维度、多层次的,既要分析具有普遍性的学校资源、教师资质和人口学特征,还要尝试将教育教学过程性要素纳入研究体系,从而为学校制定“有根据”的改进行动路线提供数据支撑。此外,鼓励试点地区教研部门和学校积极开展区域范围内的相关研究,结合各地教育教学实际形成各级各类的微报告,共同探索影响考生学业表现水平的决定性因素,更好发挥评价的诊断作用。

三是重视个体评价。目前开展的是面向地区、学校和班级的群体性评价,还未涉及面向学生的个体性评价。而随着各级各类教育的不断推进,学

生个体的重要性被提到了新的高度。《国家中长期教育规划纲要》明确提出“为每个学生提供适合的教育”，江苏省也适时提出“适合的教育才是最好的教育”科学教育理念(葛道凯, 2017)。在这种形势要求下, 评价工作也要立足于服务个体发展的要求, 关注学生个体的差异, 建立以学习者为中心的评价。就高考评价而言, 可以开发面向学生个体的学业质量评价报告, 就个体同一学科的不同方面进行比较评价, 不同学科之间成绩与能力的差异进行横向比较评价, 两个或多个学科内的学业表现进行纵向比较评价, 努力揭示每一个个体独有的特点、优势和不足, 为其更加清晰地认识自我、改进学习策略、挖掘学习潜能, 选择发展方向等提供帮助, 引导学生健康成长, 更好发挥评价的育人功能。

### [参考文献]

- 戴家干, 2008:《改造我们的考试》, 北京: 高等教育出版社。
- 葛道凯, 2017:《“适合的教育”才是最好的教育》,《光明日报》12月14日。
- 国家教育事业发展“十三五”规划, [http://www.moe.edu.cn/jyb\\_xxgk/moe\\_1777/moe\\_1778/201701/t20170119\\_295319.html](http://www.moe.edu.cn/jyb_xxgk/moe_1777/moe_1778/201701/t20170119_295319.html)(2018-01-30)。
- 国家中长期教育改革和发展规划纲要(2010—2020年), [http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/info\\_list/201407/xxgk\\_171904.html](http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/info_list/201407/xxgk_171904.html)(2018-01-30)。
- 扈岩、任涛, 2017:《高考改革背景下考试评价对高中走班制指导作用分析》,《内蒙古师范大学学报(教育科学版)》第8期。
- 梁文艳, 2017:《中小学教师教学质量评价及其影响因素研究基于增值性模型的分析》, 北京: 北京师范大学出版社。
- 乔丽娟、赵彤璐, 2006:《对我国教育考试机构专业化建设的几点思考》,《中国高等教育》第7期。

## Practice and Prospect of the Regional Education and Teaching Assessment Based on Gaokao Statistics: Taking the exploration of Jiangsu as an example

YANG Cong-yi, FAN Mei-qin, ZHU Wen-qi  
(Jiangsu Provincial Education Examination Authority)

**Abstract:** The current new Gaokao reform has formed a forced mechanism for basic education, especially for senior middle school education, which calls for more analyses and supports from examination statistics. Gaokao statistics is the most important database that

an education authority has. Making the best use of its advantage and value, is an important breakthrough in promoting the modernization of education management. Providing assessment services to rudimentary education institutions through the exploration of Gaokao statistics has already become an important fulcrum of promoting regional education reformation development, an inevitable choice during the transformation to service-oriented education authorities, and an important direction for the further development of examination assessment. The regional education and teaching assessment in Jiangsu province have developed a sound Gaokao assessment report through the construction of a scientific assessment index framework, the development of an intelligent statistical evaluation system, and the value judgement of multi-level statistics, which is comprised of two versions; the regional version and the school version. The content includes test takers' demographic background, score, matriculation status, analysis of test-takers' response proficiency, performance comparison of the last three years, quality review, and teaching suggestions. The report combines qualitative analysis with quantitative analysis, produces regional education and teaching assessment map, and features a variety of assessment content, various analysis methods, and detailed results. After two years' development and use, Jiangsu province has established a stable assessment index system and report system, gradually realized the accurate washback and guidance to rudimentary education and teaching, and obtained a good social reputation and response. In the next step, we will introduce Value-added Evaluation, strengthen the cause analysis, and highlight individual assessment to improve the regional education and teaching assessment and better play the incentive, diagnostic and education function of assessment.

**Key words:** Examination assessment; Gaokao statistics; basic education and teaching

(责任编辑: 杨娟 责任校对: 杨娟 孙志军)