

# 由一道高考历史共同题看考生群体的差异

——以定量分析为主

教育部考试中心 徐奉先

近年来,一些研究者对高考考生群体的差异进行了研究和论证,这些研究大多集中在定性方面,缺少量化分析。<sup>①</sup>2016年高考,山东、安徽、重庆、四川、福建、广东共6省市首次使用文科综合全国卷,为更好地了解不同地域考生的实际学习水平,历史学科在该年度文综全国卷第I、II卷中设计了一道共同题第28题,可在一定程度上反映使用不同卷种的考生在历史学习能力方面的差异。试题描述如下:

19世纪中期以后,中国市场上的洋货日益增多,火柴、洋布等用品“虽穷乡僻壤,求之于市,必有所供”。这种状况表明

- A. 中国关税主权开始丧失
- B. 商品经济基本取代自然经济
- C. 民众生活与世界市场联系日趋密切
- D. 中国市场由被动开放转为主动开放

本题考点是高中历史主干知识“晚清中国经济结构的变化”,主要考查考生准确理解、获取历史材料信息,并运用所学知识说明和解释历史现象、认识历史事物的本质的能力。题干材料描述的历史现象是:19世纪中期以后,洋货大量涌入,即使偏僻的农村地区,火柴、洋布等日用品在市场上都有供应。出现这种现象的历史背景,是近代以来欧美列强以枪炮不断打开亚非拉国家的国门,倾销商品和掠夺原材料,在此过程中资本主义世界市场逐步形成。近代中国亦不例外,中国被迫打开国门后列强对华商品输出不断扩大和深入,城乡市场日益被纳入到世界市场体系中,市场上的洋货无论数量和种类都不断增多,对民众的日常生活产生了深刻影响。这种

状况一方面给中国传统自然经济带来巨大的冲击,另一方面使民众的日常生活与世界市场的联系日趋密切。因此,本题的正确答案是C项。

## 一、考生作答结果分析

从考生作答情况来看,第28题在I卷和II卷使用省份的测试效果均比较好。使用全国I卷的省份为河北、河南、山西、山东、安徽、湖北、湖南、江西、福建、广东共10个,其中山东、安徽、福建、广东四省为首次使用全国卷,抽样人数共24958人,实测难度值0.633,即63.3%的考生能够答对(客观题的难度,以正确回答该题的人数占参加测验的总人数的比率来表示)。一般来说难度值在0.5~0.55的试题是测试效果比较理想的试题,根据I卷使用地区考生的作答情况来看,该题显然是一道偏易的试题。II卷的使用省份为内蒙古、黑龙江、吉林、辽宁、陕西、甘肃、宁夏、新疆、青海、西藏、重庆共10省市,其中重庆市是首次使用全国卷,陕西省是首次使用全国II卷。从实测数据来看,该卷使用省份共抽样24820人。与I卷抽样人数基本持平,实测难度值0.558,即55.8%的考生能够答对,属于中等难度试题。相较I卷的测试效果,该题对于II卷使用地区的考生而言,明显难于I卷考生。

从试题的区分度来看(试题的区分度是指测验题目对考生实际水平的区分程度或鉴别能力),区分度通常在-1~+1之间,数值越大区分度越好。一般而言,试题的区分度在0.4以上表明此题的区分度很

作者简介:徐奉先,教育部考试中心助理研究员(邮编100080)。

①任子朝、符华均、黄正正、张心、陈昂:《高中文理科学生教学水平比较研究》,《课程·教材·教法》2015年第6期。

好,0.3~0.39表明此题的区分度较好,0.3以下则表明此题的区分度不好。据表1、表2可知,第28题在I卷中的区分度为0.475,在II卷中的区分度为0.404,均在0.4以上,是一道区分度非常好的选择题。

表1 课标文综I卷第28题得分率

科目: 文综合(全国I)		题号: 1 . 28					
选项	组一	组二	组三	组四	组五	比率	区分度
未选	8	5	1	0	0	0.001	-0.0285
A	813	831	637	312	88	0.107	-0.1698
B	1093	845	482	226	59	0.106	-0.2393
*C	1539	2233	3240	4081	4714	0.633	0.4747
D	1538	1078	632	373	130	0.150	-0.2830
样本量: 24958 满分: 4.0 平均分: 2.5 标准差: 1.93 区分度: 0.475 难度: 0.633							

表2 课标文综II卷第28题得分率

科目: 文综合(全国II)		题号: 1 . 28					
选项	组一	组二	组三	组四	组五	比率	区分度
未选	8	4	3	2	1	0.001	-0.0196
A	1520	1086	840	554	186	0.169	-0.2458
B	1023	982	789	491	191	0.138	-0.1726
*C	1512	1988	2805	3390	4367	0.558	0.4043
D	901	926	747	527	230	0.134	-0.1427
样本量: 24820 满分: 4.0 平均分: 2.2 标准差: 1.99 区分度: 0.404 难度: 0.558							

图1、图2使用坐标系分别描述了在文科综合I卷和II卷中第28题的难度分布,在图示坐标中,用横轴表示参加文科综合考试的考生从0到300的各个分数段,纵轴表示第28题的得分率,由图可见,总分越高的考生,在该题的得分越高;得分率呈正态分布,有效区分了高、中、低三个分数段的考生。观察图1和图2的差异,可以看出使用I卷的考生在该题的得分率分布,总分180以上者得分率在80%以上,而使用II卷的考生得分率分布,总分200以上者得分率才能达到80%以上,说明该题对于使用I卷的考生而言更容易,这与上文中难度的分析结果是一致的。

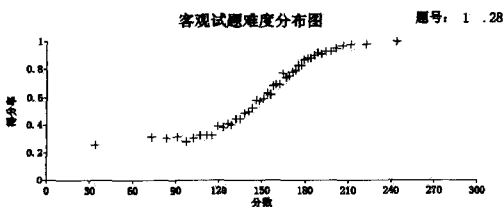


图1 课标文综I卷第28题难度分布图

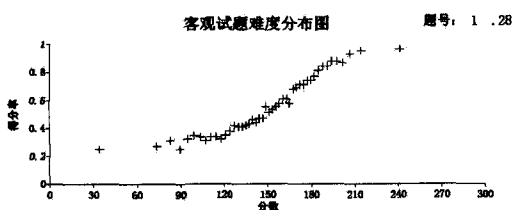


图2 课标文综II卷第28题难度分布图

题目的区分度与难度有密切的关系,一般而言,题目的难度过大或过小,区分度会比较低。题目的区分度与测验的信度也有密切的关系,整个测验中题目的区分度平均数越大,测验的信度越高。与题目的难度一样,区分度也是以某一群体为被试计算的,受所选群体的同质性影响。如果被试的差异很小,一个区分能力很强的题目的区分度指标可能很低;如果被试的差异很大,一个区分能力很差的题目的区分度指标也可能很高。就高考而言,考生的能力水平差异较大,因此偏难或偏易的试题都会影响试题的区分度。但正是由于高考针对的是水平差异较大的受试群体,在设计试题时就需要考虑到受试者的不同水平设计具有针对性的试题。例如就历史学科而言,难度值在0.65以上的试题,能够比较有效地区分低水平和中等水平的考生;难度值在0.3左右的试题,对于区分高水平和中等偏上水平的考生非常有效;而难度值在0.5~0.6之间的试题,对于低水平和高水平的考生区分效果良好。因此,在一张试卷中,并不是所有试题的难度值都保持在同一区间就好,这样并不能将不同水平的考生区分开,只有难、中、易各层级的试题合理搭配,才能保证整张试卷具有良好的区分效果,而通过在不同受试群体的不同试卷中设置共同题,则更能看出考生群体之间的差异。

## 二、考生差异分析

共同题的作用在于能够将使用不同试卷的考生间的差异和使用相同试卷的考生间的差异有机联系起来。一般来说,将分值较大的主观题设计为共同题的效果更好,可以直观地了解和比较考生的作答、得分情况和思维过程,但由于各省在阅卷的操作过程中对主观题评分标准的把握存在差异,如果直接将不同省市的数据进行比较,就降低了比较的科学性;而且就高考这样高利害高风险的大规模考试而言,将主观题设置为共同题的泄密风险也更大。而将选择题设置为共同题的好处是,选择题答案唯一,不存在测不准的情况,分值较小泄密风险也随之下降。

就2016年文综I卷和II卷的共同题第28题而言,I卷平均分为2.5分,II卷平均分为2.2分,考生群体在该题的表现,I卷考生优于II卷考生。以该题为参照,I卷中难度值在0.633以下的试题(共10题),如果用于II卷,难度值势必会低于0.558。而II卷选择题中难度低于0.558的试题共8题(参看表3)。如果I卷中的选择题由II卷考生作答,那么II卷

考生作答的情况,难度值很可能会低于0.438,这就证明了I、II卷考生之间的差异,I卷考生在获取和

解读信息、调动和运用知识方面的水平整体高于II卷考生(选择题主要考查以上两方面能力)。<sup>①</sup>

表3 2016年高考文综全国I、II卷历史选择题难度

卷种	题号												选择题难度
	24	25	26	27	28	29	30	31	32	33	34	35	
I卷	0.18	0.336	0.614	0.291	0.633	0.515	0.663	0.318	0.492	0.559	0.327	0.308	0.438
II卷	0.37	0.463	0.693	0.402	0.558	0.585	0.283	0.683	0.387	0.371	0.46	0.489	0.479

基于2016年的考后难度值数据,笔者对历史试卷各题型之间的相关性进行了统计。将历史试卷的整体难度系数设定为1,第41题难度和第40题难度与整卷之间的相关系数分别为0.901、0.869,说明这两个题型(即主观必答题)的难度对整卷难度的影

响最大。如果以客观题为错题来讨论主观题的难度差异,I卷主观题难度值为0.556,II卷主观题难度值为0.385,说明II卷考生在描述和阐释历史事物、论证和探讨问题方面的能力尚有待加强。

目前全国卷的使用情况,是若干省使用同一套

表4 2016年高考文综全国卷历史科各题型相关性分析

		历史难度	选择题难度	材料题难度	开放题难度	选考题难度
历史难度	Pearson 相关性	1	0.580	0.869**	0.901**	0.651
	显著性(双侧)		0.132	0.005	0.002	0.080
		8	8	8	8	8
选择题难度	Pearson 相关性	0.580	1	0.302	0.225	0.140
	显著性(双侧)	0.132		0.468	0.593	0.740
		8	8	8	8	8
材料题难度	Pearson 相关性	0.869**	0.302	1	0.852**	0.379
	显著性(双侧)	0.005	0.468		0.007	0.354
		8	8	8	8	8
开放题难度	Pearson 相关性	0.901**	0.225	0.852**	1	0.648
	显著性(双侧)	0.002	0.593	0.007		0.082
		8	8	8	8	8
选考题难度	Pearson 相关性	0.651	0.140	0.379	0.648	1
	显著性(双侧)	0.080	0.740	0.354	0.082	
		8	8	8	8	8

\*\*在0.01水平(双侧)上显著相关。

试卷,在一定范围内有一个统一的衡量标准,这对比较各省教育水平、衡量新生的教育质量都提供了有益的参考。<sup>②</sup>2016年全国卷分为I、II、III、IV卷。上文分析了使用不同试卷的考生群体之间的具体差异,而实际上使用同一套试卷的不同省份考生之间也同样存在差异。表5和表6分别是2016年文综I卷、II卷历史试题各题型的平均分与难度统计,由表可知,使用I卷的省份中,客观题(即选择题,总分48分)平均分的分差为2.19分,差距较小,说明考生

在该题型表现出的能力水平大体相当;但主观题(总分52分)平均分的分差达到14.31分,差距较大,其中不排除各省阅卷的评分误差因素。客观题和主观题一致性较好的是江西、湖南、广东等省,考生的成绩比较能真实反映实际水平。而在使用II卷的省份中,客观题平均分的分差为5.98分,差距较大;主观题平均分的分差达到22.21分,差距非常大。这一方面反映出各省考生之间确实存在能力差异,教育水平分布不均衡;另一方面也反映出各省对评分标准

<sup>①</sup>教育部考试中心:《2017年普通高等学校招生全国统一考试大纲》,2016年,第133页。

<sup>②</sup>李立峰:《高考分省命题不是改革的必然方向》,《粤海风》2011年第1期。

把握尺度的不同造成了人为的误差。一些省的客观题平均分比较低,但主观题平均分却远远高出其他省,造成考生成绩的虚高;而一些省的客观题平均分较高,主观题平均分却偏低,也没有反映考生的实际水平,应当在今后的高考阅卷过程中加强对评卷教师的培训。

表5 文综I卷历史试题各省各题型平均分与难度统计表

省份	客观题 平均分	客观题 难度	主观题 平均分	主观题 难度	总分 平均分	总分 难度
A省	19.57	0.41	32.79	0.63	52.35	0.52
B省	20.23	0.42	20.77	0.40	41.00	0.41
C省	21.70	0.45	35.08	0.67	56.77	0.57
D省	20.92	0.44	23.35	0.45	44.27	0.44
E省	21.15	0.44	28.61	0.55	49.76	0.50
F省	20.64	0.43	22.93	0.44	43.56	0.44
G省	20.62	0.43	30.57	0.59	51.19	0.51
H省	19.51	0.41	27.96	0.54	47.47	0.47
I省	21.02	0.44	34.82	0.67	55.84	0.56
J省	20.73	0.43	30.83	0.59	51.56	0.52

表6 文综II卷历史试题各省各题型平均分与难度统计表

省份	客观题 平均分	客观题 难度	主观题 平均分	主观题 难度	总分 平均分	总分 难度
K省	18.11	0.38	11.27	0.22	29.37	0.29
L省	23.65	0.49	25.59	0.49	49.25	0.49
M省	22.96	0.48	26.11	0.50	49.07	0.49
N省	21.78	0.45	13.51	0.26	35.28	0.35
O省	24.09	0.50	20.18	0.39	44.28	0.44
P省	18.55	0.39	27.76	0.53	46.30	0.46
Q省	22.43	0.47	25.24	0.49	47.67	0.48
R省	21.92	0.46	27.96	0.54	49.71	0.50
S省	22.15	0.46	33.48	0.64	55.63	0.56
T省	18.99	0.40	18.71	0.36	37.69	0.38

### 三、高考依照考生群体分卷的必要性

鉴于恢复高考后全国一张卷的高泄密风险,自2004年起,全国卷采用一纲多卷的形式供各省考生使用,最初主要是出于试题信息安全保密的需要。使用全国卷并不意味着各省份使用同一张试卷,而是实行“一纲多卷”。“一纲”是保证全国统一高考的

基础,通过“一纲”保证各类型试卷遵循全国统一的考试要求,充分体现国家人才选拔要求的一致性和公平性。“多卷”是根据不同省份的教育发展水平提供不同的试卷。由于各省经济社会发展水平存在差异、推进高中课程改革进度不一,高中教育教学及学生实际水平也存在差异。通过本文对2016年考后具体数据的采用和分析,可以看出各省的考生群体之间既有学科能力或素养方面的差异,也存在一些共性。针对考生的具体情况,使用不同类型的试卷能够使试题更好地与学生水平相拟合,提高试卷的区分度,实现更好的选拔效果。此外,实行“多卷”还有利于全国统一考试安全。当前高考作为一项高利害考试,组考的外部环境复杂严峻,不同地区使用不同的试卷,有效降低了全国统一高考的安全事故风险。

随着考试和课程改革的不断深入,经过10余年的实践与探索,全国卷的卷种和考试内容已经发生了极大的变化。《国务院关于深化考试招生制度改革的实施意见》明确指出,新一轮高考改革要从“2015年起增加使用全国统一命题试卷的省份”,总体目标是“到2020年基本建立中国特色现代教育考试招生制度,形成分类考试、综合评价、多元录取的考试招生模式”,可见一纲多卷依然是未来高考的主要形式。但分类依据已然不仅仅是出于安全保密的需要,而是更多地考虑了考生群体的差异和需求。面对命题格局的重大调整,考试中心曾多次赴相关省份开展调研工作,了解各地教育教学情况,指导做好命题调整和复习备考的准备工作,确保命题调整工作平稳顺利。在坚持统一考试大纲的基础上,2016年全国卷又增加了1套,共命制3套“3+文综/理综”试卷、1套“3+3”试卷。各套试卷的命题依据、考试标准、试卷形式等方面均保持稳定。

目前北京、上海、天津、江苏、浙江等省份仍继续依据全国统一《考试大纲》实行分省命题,形成了全国统一命题和少数省市自主命题相结合的新格局,这一安排既是对考试招生制度改革及高中课程改革的保障,也有利于促进高考命题水平提升和推进考试机构专业化建设。“高考命题在坚持正确的改革方向的同时,需适当考虑众多学校的实际情况和考生的实际水平,慎重把握改革的力度和步伐,稳步推进,使新高考与新课改实现良性互动”。<sup>①</sup>

(责任编辑:李月琴)

①毛竞飞、盛兰芳、李金波:《高考成绩群体差异性分析》,《现代教育管理》2011年第4期。