

历史

教育考试研究

姚锦祥

著

LISHI
JIAOYU
KAOSHI
YANJIU



NORTHEAST NORMAL UNIVERSITY PRESS

东北师范大学出版社

WWW.NENUP.COM

历史

教育考试研究

著

姚锦祥

LISHI
JIAOYU
KAOSHI
YANJIU

东北师范大学出版社
长 春

图书在版编目 (C I P) 数据

历史教育考试研究 / 姚锦祥著. — 长春: 东北师范大学出版社, 2008. 12
ISBN 978 - 7 - 5602 - 5517 - 0

I. 历… II. 姚… III. 历史课—教学研究—高中
IV. G633.512

中国版本图书馆 CIP 数据核字 (2008) 第 195615 号

☐ 责任编辑: 沙铁成 ☐ 封面设计: 张 然
☐ 责任校对: 孙维石 ☐ 责任印制: 张允豪

东北师范大学出版社出版发行
长春市人民大街 5268 号 (邮政编码: 130024)
销售热线电话: 0431—85687213
传真: 0431—85691969
网址: <http://www.nenup.com>
电子函件: sdcbs@mail.jl.cn
东北师范大学出版社激光照排中心制版
吉林省吉育印业有限公司印刷
长春市经济技术开发区深圳街 935 号 (130033)
2008 年 12 月第 1 版 2008 年 12 月第 1 次印刷
幅面尺寸: 185 mm×260 mm 印张: 14.25 字数: 332 千

定价: 25.00 元

【前 言】

高度决定视野 角度改变观念 尺度把握人生

改革之路从来都是不平坦的，没有明确方向的道路更是举步维艰。高考改革需要勇气和力量，也需要理论的指导。只有站得高才能看得远，所谓“高度决定视野”；改革的进程就是解决问题的过程，但问题形成的原因是多方面的，从不同的角度可以发现不同的问题，所谓“角度改变观念”；多一把尺子便多一批好学生，科学的测量方法和评价理念能够催人奋进，所谓“尺度把握人生”。

大学通过考试遴选申请入学者，在我国已有相当长的一段时间。“文化大革命”期间停止高考改用推荐的做法，不仅使大学教学无法正常进行，而且也使基础教育难以为继。1977年恢复高考制度是教育界拨乱反正的重要举措。它改变的不仅仅是个人的命运，更带来了我们国家与民族的复苏和新生。在当今和以后相当长的时间内，由于社会诚信度不高和优质教育资源有限等原因，高考因其严密的规范程序和公开统一的标准仍应是高校选拔新生的基本方法。

近年来，高考的问题及其负面影响日益暴露，其批评之声也越来越响。其中，有的涉及命题目标、内容和方法等测量方面的问题，有的涉及评价方面的问题。最为重要的莫过于对分数决定论的批评，即高考成绩对能否为高校录取、为怎样的高校录取起着决定性作用，而考生的道德修养、综合素质和个性特征则被忽略。唯考分是论的评价标准还进一步扩展到了其他领域，如考生学业的优劣、教师工作的勤惰、学校办学的业绩等全系于考分的高低。

高考过分强调甄别的作用对于中学教育产生了深刻影响。按分数排队，以考试成绩取代其他评价内容和手段等，对学生的发展造成了不良压力，不利于学生多样性、个性化的发展。产生这些问题的主要原因是未能建立科学的评价体系。

我国自1990年起建立了高中毕业会考制度。该制度着眼于评价学生是否达到合格高中毕业生的标准，有助于普通高中学业水平测试制度和评价制度的建立。高中毕业会考制度实施的效果并不理想。因此，高考不仅承担着为高校选拔合格新生的重任，还需承担评价中学教育的功能。

课程改革带动了高考制度的改革。从已经出台的广东、山东、宁夏、海南、江苏和浙江等省区的高考方案来看，尽管具体的措施和方法有所不同，但在指导思想、科目设置等方面基本上体现了对素质教育的积极导向作用，展示了会考与高考改革的新方向和新趋势。

一是将综合素质评价引入高考的范畴。新课程实验省区的高考方案将综合素质评价作

为高考录取的依据之一，其实质是改革高校录取新生中“唯分是举”的弊端，形成录取标准多元化的格局，使中学教育阶段的过程性评价成为高考评价的一部分，有利于推动高中新课程的评价改革。二是采取高中学业水平考试与高考入学考试相结合的方法。这可以有效地避免学生的偏科现象，促进高中课程方案的全面实施，又可以防止过多科目挤入高考增加考生的负担。同时，承袭高中毕业会考基础上的高考制度设计框架有利于解决高中毕业会考实施中的不尽人意之处，促使其应有功能的正常发挥。

考试是教学系统中不可或缺的一个环节，是检测教学质量、衡量教学效果的重要手段，也是检验和评定学生学业情况的重要方法。目前，高考还是我国高校选拔人才的基本手段，它将长期存在。

高考是对基础教育成果的一种评价，对中学教育具有重要的导向作用。既然高考的指挥棒作用无法避免，我们就需要考虑如何使这一指挥棒向积极的方向（也就是向素质教育的方向）导引。目前最为重要的是要借助这一指挥棒为高中新课程的顺利推进起保驾护航的作用。

高中课程改革正逐步深化，与之相适应的高考命题改革也在紧锣密鼓地进行。考试、评价应该成为新课程实施的助力。实施课程标准条件下的高考，如何进一步完善其目标、内容、方法和评价问题，并解决好考试与教学协调发展的问题，已经成为高中教育考试研究的重要课题。

2007年至2008年，广东、山东、海南、宁夏和江苏等省区相继进行了课程标准条件下的高考命题改革。其考试大纲和考试说明及其试卷和试题体现了高中历史新课程的基本理念，在一定程度上代表了高考命题改革的特点和方向。由此，以下一些问题自然进入了人们的视野。

新课程高考是怎样的？命题方式与旧课程有什么区别？新课程高考的考试目标和要求、考试内容与知识范围、试卷设计与试题编制等问题究竟发生了哪些变化？这些变化的理论基础是什么？它们对新课程的教学和复习有怎样的要求？采用哪些教学方式和方法才能与新高考接轨？新课程高考改革是否适应高中历史课程改革、教材改革和教学改革的要求？新课程高考改革取得了哪些成果和经验？还有哪些存在的问题？

这些是已经进入和即将进入高中新课程实验区的教师以及研究者十分关注的问题。

本书结合高中历史课程标准及其考试大纲、考试说明和高考试题，就相关问题展开了深入的分析和研究，力图以新的视角审视高中历史课程标准条件下的高考命题改革和高中教学改革，并对其运行过程和结果进行一些反思。其内容和特点主要表现在以下几个方面：

第一，对主题内容的阐释集中而全面。本书由六章组成：“考试改革的理论视野”、“考试目标和要求”、“考试内容和知识范围”、“试卷设计和试题编制”、“试卷分析和试题评价”以及“考试与教学和复习”。它们涉及考试目标、内容、方法、评价以及考试和教学的关系等领域，并围绕着高考改革这个主题，分别从理论视野、目标改革、内容和方法改革以及评价改革等方面展开，对主题的阐述可谓集中而全面。

第二，对主题内容的表述重过程也重结论。本书采用章节体方式阐释问题。每章和每节的内容均在起首段提出要阐述的问题，并据此进行分析和论证。每章以“总结和讨论”为题，就历史高考改革中取得的成果和经验进行总结，对存在的问题进行讨论，以期引起

各方面的注意。

第三，对主题的认识具有宏阔的视野和新颖的视角。本书既有考试学和心理学等方面的理论阐释，也有命题和教学问题的实践探讨；既有包括大陆和台湾在内的国内考试评价制度及其命题改革的分析，也有发达国家相关问题的比较；既有高考命题内容及其规律的梳理，也有教学和复习内容及其规律的总结。从理论到实践，从国内到国外，从考试到教学，可谓高屋建瓴，视野开阔。

第四，选择的材料丰富而鲜活。“考试改革的理论视野”一章中，有关目标分类学的材料取自近年的研究成果。考试目标和内容的研究对象，以2007年和2008年我国各地区的考试说明为主体。相关试题中包括山东和宁夏的文科综合试题，广东、江苏和海南的历史单科试题，以及全国文综甲卷和乙卷的试题。

第五，注重实际。本书无论是理论的阐释和实践经验的总结，还是命题和教学问题的探讨，都从实际出发，具有很强的操作性。这从相关章节的标题上就可看出。如第四章主要涉及试卷设计和试题编制的问题，各节的标题依次表述为：以考试说明为依据的试卷设计、历史科目试题体现的考核目标和设计方法、文科综合科目历史题的命题方法与特点、不同版本教材背景下的试题编制、专题型历史内容的命题方式。最后两章在考试评价和教学复习方面 also 具有很强的实践性，如试卷和试题信息的采集与评析、答卷答题信息的采集与分析、试卷与试题分析示例、以历史专题为单元的教学和复习、以历史时期为单元的教学和复习等。它们均从方法和策略的角度对相关问题进行了系统的总结和提示，应当对命题和评价以及教学和复习具有重要的指导作用。

本书系教育部“十一五”重点课题研究成果之一。课题名称：中国百年高考的历史回顾及其现代启示。项目批准号：DRR070172。负责人：胡金平。主要参加人：姚锦祥和胡平等。

本书系2005年江苏省高校哲学社会科学项目研究成果之一。课题名称：新课程条件下教师教育课堂教育方式的变革。负责人：姚锦祥。项目批准号：05SJB770003。

本书着眼于高中历史教育考试测量的基本理论和实践问题的探索，读者以高中教师和研究人員为主。编写者力戒空洞说教和资料堆砌，既高屋建瓴，视野开阔，又密切关注新课程、新高考，拥有鲜活的材料，进行了生动的表述。

需要说明的是，本书作为本人的一项研究成果，编写过程中得到了许多同行的热情鼓励，并参考了他们的大量研究成果。书稿的写作也得到了东北师范大学出版社吴东范老师的支持，正是她的鞭策和指导才使本书有了现在的面貌。对此，我深表谢意，并热诚地期待同仁的批评和指正。

姚锦祥

2008年8月写于随园

【目 录】

第一章 考试改革的理论视野	1
第一节 教育测评学的视野	2
第二节 目标分类学的视野	9
第三节 信息加工心理学的视野	19
第四节 高中学业水平考试与大学入学考试制度的改革	25
总结和讨论	31
第二章 考试目标和要求	33
第一节 历史科目考试的能力目标和要求	34
第二节 文科综合科目考试的能力目标和要求	44
第三节 历史考试目标与课程目标的关系	51
总结和讨论	63
第三章 考试内容和知识范围	64
第一节 历史学科考试内容的确定	65
第二节 考试内容的组织方式和特点	70
第三节 考试内容的主题体现了文明史观的价值取向	76
第四节 文科综合卷历史学科考试内容的分析	82
总结和讨论	88
第四章 试卷设计和试题编制	90
第一节 以考试说明为依据的试卷设计	91
第二节 历史科目试题体现的考核目标和设计方法	100
第三节 文科综合科目历史试题的命制方法和特点	107
第四节 不同版本教材背景下的试题编制	117
第五节 专题型历史内容的命题方式	125
总结和讨论	129

第五章 试卷分析和试题评价	131
第一节 试卷和试题信息的采集与评析	132
第二节 答卷答题信息的采集与分析	147
第三节 试卷和试题分析示例	157
总结和讨论	176
 第六章 考试与教学和复习	 178
第一节 课程标准是考试与教学的联结点	179
第二节 以历史专题为单元的教学和复习	188
第三节 以历史时期为单元的教学和复习	202
总结和讨论	217

【第一章】

考试改革的理论视野

具有很高公信力和很强科学性的考试都是在心理学和测量学基本原则的指导下进行的。到目前为止，我国心理测量学还没有真正融入到考试的实践中去，所以在考试的目标、内容与方法以及评价方面产生了一系列的问题。相关问题的解决，有必要关注支撑考试实践的测量学与心理学以及考试理论研究的进展，从中寻找解决问题的可能途径。

历史高考改革的理论基础主要包括教育测评学、目标分类学和信息加工心理学三个方面的内容。“教育测评学的视野”从测量学和评价学的层面探讨包括历史学科在内的高考现象的基本规律；“目标分类学的视野”从考试目标分类的层面思考历史高考应该考什么的问题；“信息加工心理学的视野”从历史教育心理学的层面分析历史高考应该怎样考的问题。在此基础上归纳和概括中外高中学业水平考试和大学入学考试改革的现状，容易对高考改革的方向和趋势形成清醒的认识。

第一节 教育测评学的视野

教育测量与评价的实践和理论是各种考试制度形成和发展的基础。教育测量在社会公认的价值标准下,其本身就隐含着价值分析和价值判断。高考既担负着教育测量的职能,也被认为是一种重要的教育评价。

一、教育测量和评价的历史发展

教育测量和评价的产生与对学生学力和学业成就的检测活动的发展有着紧密的联系,其历史源远流长。一般认为,现代的教育测量与评价发端于20世纪30年代美国教育家泰勒的“八年研究”,而系统的测评活动则萌生于古代的中国。

1. 中国古代教育测评的发展历程

以科举制度的产生为界,古代中国的教育测评活动可以分为前、后两个时期。此前为萌生期,包括先秦至魏晋南北朝时期的选士测评活动;此后为科举时期,经历了从隋唐至清末我国古代教育测评的发展历程。

早在西周时期,随着学校教育制度的初步建立,我国的教育测评活动也随之展开^①。我国古代最早的一部教学论专著《礼记·学记》中记载:西周的大学,“比年入学,中年考校”。其内容为:“一年视离经辨志,三年视敬业乐群,五年视博习亲师,七年视论学取友,谓之小成。九年知类通达,强立而不返,谓之大成。”学生在第七年和第九年的考试中达到了标准,就是“小成”和“大成”。西周官制规定^②,在“大成”的学生中进一步选取优秀者,并经过考核选拔,就可授予其官位和爵禄。可见,西周学校的选士制度已有一定的测评程序和规程。

从公元606年隋炀帝设进士科开始,我国的科举考试制度由此形成。科举制将考试、教育与做官联系在一起,使教育测量与教育评价有机结合,开创了古代教育测评学的先河。科举考试的方法对我国的教育测量和评价有着重大的影响,有的甚至流传至今。例如,以下一些考试方式在今天的教育活动中仍然可见其踪影。

帖经,将所考经书的某页遮掩两端,仅露一行,再用纸盖住此行的三五字,让考生填写这几个字。此法即为填空题的前身,主要考查记忆能力。

墨义,根据经书出题,要求用原文回答,有口试和笔试两种。它相当于现今的简答题,主要检查背诵经书的情况。

策问,按照题目的要求写出论证文章。与现今的论述题相仿,主要考查知识的理解和运用能力以及文字表达能力。

诗赋,按照题目的要求和规定的格律创作,可以看成作文题的前身。

① 西周,中央有国学,地方有乡庠、州序、党校、闾塾,皆为官办。校内有完整的教学、考核、升降制度。国学分为小学和大学两个阶段。大学生源主要来自贵族胄子,选士在大学进行,与考试紧密结合。

② 据《礼记·王制》记载:“……论进士之贤者,已告于王,而定其论。论定然后官之,任官,然后爵之;位定,然后禄之。”

中国古代的教育测评一开始强调的就是其筛选和管理功能。从西周学校的选士到隋代开始的由官府组织的科举,都是一种选拔性考试。特别是科举制,它以科举考试为主要手段,以封建礼教为基本内容,对个人学识德能进行测量评价,从中选拔治理国家的人才。

科举制发展至明、清时期,由于其陈腐的内容和一成不变的形式已经成为阻碍中国社会变革的桎梏,最终在1905年被废除。但其按照成绩高低定取舍的准则、分科考试的方法和严格的管理制度引起了国外有识之士的注意并得以新生。明末清初时期,科举制度被介绍到欧美国家。18世纪以后,英、法、美等国开始学习中国的考试方法,并将之改造成文官考试制度。

2. 现代教育测评的发展历程

现代教育测评制度是工业革命和近代学校教育制度的产物。只有大学毕业才能进入政府部门工作的规定引起了大学入学的竞争。19世纪末,在工业革命的深入发展中,西欧国家中等教育考试制度纷纷建立。著名的有英国的十一岁考试和中等学校证书考试、德国的完全中学毕业证书考试以及法国的初中和高中毕业会考。20世纪初期,在教育实践和心理测量理论发展的推动下,欧美国家尤其是美国开始了现代教育测量的实践活动和理论探索。现代教育测量理论在此基础上建立起来,并经历了前、后两个发展阶段。

20世纪50~60年代之前为古典测量理论阶段。其主要特点是以真值理论为基础,以信度、效度、难度和区分度为指标。它认为考生的分数并非一个绝对的值,而是其真实能力或技能的反应与测量误差的和,所以称为真值理论。在这一假设的基础上延伸出信度、效度、难度和区分度等概念。同时,人们逐步建立并完善了测验发展的标准化程序,以提高所编制测验的信度和效度。古典测量理论对于考试实践曾经起过非常大的指导作用。

1901年,美国举办了第一次跨州的统一考试,称为“学业成就考试”(ACT)。1926年,受比纳—西蒙智力测验量表的影响,美国大学入学考试委员会举行了学术性向测验(SAT),并逐步成为一种标准化的测验。这种考试的目的是为了测验学生的书本知识,而是看其是否具有在大学学习的能力。试卷分为言语和数字两部分,通过若干子测验及收集到的分数制成剖面图,了解学生的强项和弱项所在,从而为学习、教育和就业选择提供指导。标准化测验也曾遭到许多批评。研究者认为,那种全部为选择题的测验难以对多数学科的教育目标和教学内容作出全面的测量和评价。

20世纪60年代以来,项目反应理论逐渐形成。其主要特点是采用严格的数学模型来刻画不同能力水平的测试者对测试题目的反应模式。相对于古典测量理论来说,项目反应理论取得了巨大的进步。例如,古典测量理论的难度概念往往是一个平均的概念。但对于不同能力水平的考生来说,难度概念的意义是不一样的。能力低的考生觉得难的题目能力高的考生并不觉得难,所以古典测量理论难以反映出题目与被测试者之间一种真正的交互关系。项目反应理论对此有所拓展,并与信息技术进行了有效的结合,其应用领域日益广泛。目前正在蓬勃兴起的计算机自适应测试就是一个有力的证明。计算机自适应测试(CAT)是由计算机自主呈现测验题目。如果学生正确回答一个问题,则呈现较易的问题;如果学生不能正确回答一个问题,计算机则自动估计其能力水平,然后呈现更容易的问题,直到计算机对其能力的估计稳定在某一水平为止。CAT之所以引起人们的高度关注,是因为它表现出一些非常诱人的长处:一是大约节省50%的考试时间;二是不同的学生将接受不同的考试内容;三是对学生能力的估计更为准确;四是对学生不同能力水平的估计

放在同一尺度上便于报告和分析。

教育测量理论与实践的发展也推动了教育评价的进展。在1929年至1933年经济危机的影响下,教育的社会效果问题越来越多地受到关注,这就推动了以教育的社会价值为依据的教育评价活动的发展。

1934年至1942年,美国俄亥俄州立大学教授泰勒(R. W. Tyler)受卡耐基金会的资助,进行了教育评价发展史上著名的“八年研究”。在对以往的课程和测验设计提出尖锐批评的基础上,泰勒提出了一套以教育目标为核心的课程和测验编制原则,试图以此把社会的要求、学生的需要反映在课程和测验中。他认为,“在本质上,评价过程乃是一种测量课程和教学方案在多大程度上达到了教育目标的过程”。换言之,教育评价就是衡量教育活动达到教育目标程度的一种活动,测量只是其手段。1949年后,随着泰勒《课程与教学的基本原理》一书的出版,教育目标与教育评价之间的紧密联系性逐渐为人们所接受。泰勒提出的四个基本原理也成为教育改革的经典性命题。实际上,确定目标、选择学习经验、组织学习经验和评价结果,既是教学的过程,也是评价的过程。50~60年代,他的学生布卢姆发展了他的思想,把目标与评价完好地结合在一起。

泰勒理论的核心是目标,但由于教育目标概念与教育结果的概念紧密相关,这就容易使人们将更多的注意力集中到对教育结果的评价上。因此,泰勒评价更多地被看成一种对结果的总结性评价。但形成性评价可能有着更为重要的作用。此外,人们认识到,目标本身的合理性也需要评价,对那些预期目标以外的教育效果也不能排除在评价之外。

在对泰勒评价观进行分析和批判的基础上,有人提出了新的教育评价观。克龙巴赫(L. J. Cronbach)修改了泰勒的评价定义。他认为,评价就是“为作出关于教育方案的决策,收集和使用信息”。斯塔佛比姆(D. L. Stufflebeam)给出的评价定义则是“为决策提供有用信息的过程”,并强调“评价最重要的意图不是为了证明,而是为了改进”。评价对教学的反馈作用受到注意。人们由以往只注意教学过程对评价过程的单向作用,开始注意到教学过程与评价过程的交互作用,并利用测评的结果诊断教学中出现的问题,由此拓宽了评价的功能。

随着质性研究方法的兴起,人们对标准化评价方法进行了反思。只从最后的结果、只看学习数量的评价方法受到质疑,如何评价教育的质的方面引起了研究者的关注。在此基础上,人们提出了形成性评价的概念。形成性评价强调的是学习过程中每一步骤的结果与目标的对照,评价工具已从单一的试卷扩大到各种调查量表和问卷。

20世纪80年代前后,现代认知心理学的发展促进了现代教育测评理论的进一步发展。研究者认为,学习的质量不仅反映在学习的效果上,也反映在学习的过程中。面对学习的任务,学习的动机及其策略(即学习方式)不仅反映了学习者如何经历学习过程的途径,还决定了对学习的期望和最后可能获取的结果,反映了学习的质量。因此,评价既要关注学习结果也要关注学习的动机和策略过程。过程性评价的概念由此从形成性评价的概念中分离出来,并凸显出以学生发展为本的特色。

过程性评价不仅关注阶段性和过程性的成果,同时关注学习方式。由于评价的回流功能,过程性评价导向对学习过程的不断修正,成为促进学习和发展的媒介。这就使得评价的过程与学习的过程交互作用,相互融合。这种倾向于“过程”和“发展”的价值取向成为新的评价观的理念基础。80年代以后,过程性评价的技术发展迅速,影响最大的就是

所谓的“档案袋评价”（即成长记录）的评价方法，表现性评价和真实性评价的方法也有一定的发展。

二、教育测量和评价的要素和特点

从教育测量和评价的产生和发展历史来看，首先是对学生学力和学业成就的评定需要，导致了教育测量的产生。随着社会对学生的全面发展的需要，教育测量开始走向教育评价。

1. 教育测量和评价的概念

通常人们所说的测量，是用一定的工具检定某一物体的物理性质的活动，即根据某些法则和程序，用数字对事物在量上的规定性予以确定和描述的过程。教育学和心理学的不断发展使人的知识水平、思维品质和能力倾向等心理属性也能够通过一定的方法和手段进行量化研究。

教育测量就是对教育领域内的事物或现象，根据一定的客观标准进行考核，并依据一定的规则将考核的结果予以数量的描述。评价的实质在于价值判断。评价活动的过程就是对人或事物的价值进行分析、衡量和判断的过程。当把“评价”一词特别地用于学校教育领域或课堂教学情境时，在一些情况下，“评价”就是“教育评价”一词的简称。教育评价是以教学目标为依据，运用可操作的科学手段，通过系统地收集有关教育的信息，对教育活动的过程和结果作出价值上的判断，并为被评价者的自我完善和有关部门的科学决策提供依据的过程。

教育测量和评价是两个既紧密联系又相互区别的概念。测量和评价是同一活动的两个不同阶段。测量是一种事实描述过程，只对事物的属性赋值；评价是一种价值判断过程，是在测量的基础上对事物的好与差、对与错、优与劣的判断。不过，测量只能是在评价需要的指导下的事实描述，否则测量没有任何意义；评价必须是在测量基础上的价值判断，否则评价结果就是空中楼阁，毫不可信。例如，西周时期的“小成”与“大成”的考校和隋唐时期的科举考试，如果不是为了选士和选官，其考试就没有意义。反之，不经过严格的、公平的考试程序，就难以选出需要的人才。二者相辅相成。正是由于教育测量与教育评价之间的密切联系，我们将二者合在一起，统称为教育测量和评价，简称为教育测评。

可以认为，现代教育测量和评价是在对教育目标、教育过程、教育结果以及影响教育的各种因素作出量（或质）的描述的基础上进行的一种价值判断活动。教育系统中的一切因素都是测评的对象；教育测量就是对评价对象的属性和规律作出客观的、定性或定量的描述，这种描述的结果是对测评对象的某些属性分配数值；教育评价就是在教育测量的基础上对评价对象作出价值判断。

我国的高考有广义和狭义之分。狭义的高考是一种教育测量。它是根据考试目标选择有代表性的内容和问题，按照一定的方式对考生的知识、能力和思想素养等进行测量的过程。广义的高考是指考试部门从统一组织命题和考试到统一组织阅卷和录取的全过程，即在考试过程的基础上，根据招生的计划和要求，通过对考生成绩等资料信息的分析和推断，评估其知识和能力水平以及入学的资格，并作出是否被高校录取的决定。可见，高考既承担了教育测量的功能，也发挥着教育评价的作用。

2. 教育测量的要素和特点

无论哪一种测量都离不开测量工具、参照点和单位这三项要素,教育测量也不例外。但在内容和功能作用上,教育测量与物理测量相比具有明显的不同。

被测对象的测量结果是通过测量工具的使用确定的,其精确性由测量工具决定。物体的测量工具,长度可以用米尺,重量可以用天平秤,温度可以用温度计。教育测量的常用工具为试卷。与米尺、天平秤和温度计等量具的客观性相比较,试卷这一测量工具由许多人为的因素决定,具有一定的主观色彩。

要确定事物的量,必须有一个计量起点,这样才能使测量结果具有相互比较的基础。测量物体的参照点可以分为绝对零点和相对零点两种。长度和重量都以绝对零点为起点。温度的起点则是相对的,一般以水在标准状况下的冰点作为参照点,记为零度。教育测量的特质很模糊,考生的知识和技能等不能说是在丝毫没有的基础上获得的。其参照点只能是相对的,并随着参照的标准而变化。参照的标准一般有常模参照和目标参照两种。

有了统一的单位,才能将被测对象的量与数字对应起来,从而精确地表示数量的多少。“克”、“千克”以及“厘米”和“米”等就是计算物体轻重和长短的测量单位。测量单位应当是公认的,即不会引起歧义;测量单位的距离是不变的,即在量尺上单位的实际价值处处相等。例如,10米和9米的差距与90米和89米之间的差距是一致的。但在教育测量中往往是不等距的或等距不等值的。例如,同样是5分的两道试题,识记水平与运用水平的5分的测量单位是不等值的。

由于测量要素的具体内容和价值作用的差异,教育测量不同于物理测量的特点主要表现在以下方面:

第一,测量方法的间接性和推断性。物理测量是对物体的直接测量,教育测量测量的不是外在的物理属性,而是考生内在的心理属性,如知识、能力和思维品质等。它们看不见,摸不着。当今的科学技术还无法支持教育测量科学发展到能用某种量具直接进入人脑或人体内去测量人的精神方面的属性。高考只能通过考生对试题的外显行为的刺激作出的反应结果,对其知识水平和能力水平等作出推断性的检定。

第二,测量对象的模糊性和测量误差的不可避免性。从测量结果来看,物理测量通过采用精良的测量工具,按照科学的测量法则和规范的测量程序,其测量误差完全有可能控制在一定的范围之内,但是教育测量做不到。一方面,知识与能力水平等内容的测量不像桌子的长度或人的重量等物理特性那样明确,容易受人的心理活动的影响;另一方面,这样的测量只能是对考生外显行为与反应的取样分析加以推断而成,这就使测量的对象显得模糊和不确定。除了随机误差和系统误差之外,还存在抽样误差。高考考查的内容只是考生所学内容中的一小部分,不一定真实反映考生的实际水平。

教育测量的结果不可能获得与物理测量一样的精确性,其原因是多方面的,最主要的是测量工具。物理测量结果的客观性来自工具的可靠性。如用一把尺子度量某一物体的长度,无论何时、何地、何人进行,其结果应该是一致的,因为这把尺子是可靠的。欲使教育测量也具有客观性和可靠性,作为测量工具的试卷应当具有两个基本的指标:一是效度,即测量的结果符合测量的目的;二是信度,即测量的结果无论何时、何地、何人使用都是基本一致的。

历史教育测量是教育测量在具体学科中的应用,其概念也可由此导出,即:依据一定

的法则侧重从量的规定性上对历史教育的效果（主要是历史知识和历史学科能力）予以确定和描述的过程。从定义出发，影响历史教育测量的因素主要包括“一定的法则”（主要是考试标准）和历史知识的特点与历史学习的认知特点。

任何有效的测量工具都要符合被测事物发展的规律，都是建立在一定的法则基础之上的。天平秤能有效地测量人的体重是因为它遵循了力学杠杆原理的法则，温度计能测量物体的温度体现了事物热胀冷缩的规律。教育测量工具同样需要遵循一定的规则。在历史教育测量中，考试标准、历史知识的特点和历史学习的认知特点就是所要依据的规则，也是影响其测量工具有效性和可靠性的最为重要的因素。

三、高中学业评价的基本类型和要求

在学校系统中，教育评价包括学生评价、教师评价和学校评价三种类型。在学生评价中，与学生学习状况及其成就水平有关的学业评价占有重要的地位。对学业评价类型及其功能和作用的了解是其正确使用的前提。

学业评价从不同的角度划分有不同的分类方法。从评价目的和教学时机的角度可以分为诊断性评价、形成性和过程性评价以及总结性评价；从测量标准和考试分数解释的角度可以分为常模参照评价和标准参照评价；从测量内容的角度可以分为学业成就评价和学习能力评价等。

诊断性评价又称准备性评价，是教学活动开始前进行的评价。它主要是对教学背景和学生的各方面情况作出评价，并据此进行教学设计。诊断性评价的方式灵活多样，可以是发放问卷的方式、制成试卷前测的方式，也可以是课堂上提问和谈话的方式。通过测量和评价可以了解学生对将要学习的内容的优势和不足。对教学的意义在于：如果是学生的薄弱环节，也就是教学的重点；如果是学生的优势，也就是教学中不必花工夫的地方。同时，还可以对学生进行有针对性的指导。了解学生是否具有新的教学目标所必需的知识和能力水平，把不同程度的学生分置在最有利的教学序列中，有利于教学计划和教学内容的安排。

形成性评价和过程性评价都是在教学活动的进程中实施的。其目的是及时了解教学进展情况，发现教学方法，促使教学不断完善。形成性评价常用提问和谈话反馈的方式进行。通过交流可以评判学生的学习状态。如果所有学生的进展都很理想，教学就不需要作任何的调整，反之则需要大幅度调整教学方法。通过形成性评价监控学生的进步，有时可以调整教学计划。假设有的单元内容大部分学生都已经掌握，原先安排两周的时间可以作出缩减的计划。与形成性评价相比，过程性评价采取目标与过程并重的价值取向，对学习效果和过程等各方面因素进行全面评价，主张内外结合的开放的评价方式，主张评价过程和教学过程的相互作用。其功能涉及学习水平的判断、成绩的肯定、对学习过程的反思、评价方法的掌握等方面。

总结性评价又称终结性评价，是在一相对完整的教学阶段结束时对教学目标实现的程度作出的结论性评价。终结性评价的作用在于通过考试收集证据信息，以给学生一个合理的分数等级。学生的学习成绩构成了他们人生履历中重要的最初的成就记录，教师的评价决不能草率行事。高中毕业学科的学业水平测试（俗称“会考”）和大学入学考试（俗称“高考”）也可以看成一种终结性评价。

同时,学业评价对教学效果的评定、教师的评估、教学目标的确定和调整具有重要的意义。高中毕业考试和升学考试等高利害的评价结果对教师和学校的评价具有极其重要的影响。

目标参照性评价又可称为绝对评价,是以某一预定的目标为客观参照点(如课程目标),寻求评价对象达到客观标准的绝对位置的评价。它不以区分个体之间的差异为目的,而是评价学生是否达到了课程目标规定的要求以及达成的程度。我国高中毕业学科的学业水平测试或高中会考属于目标参照评价。

常模参照性评价又称相对评价,是以对象群体的平均水平为参照点,确定评价对象在群体中相对位置的一种评价方式。目的在于判断考生在该群体中的相对位置,以区分其学习的水平。它常被应用于评定学生的优劣和选拔优秀的人才。高考就是一种常模参照性考试。

学业成就评价(也叫学习成绩评价)旨在对经过一定阶段和形式的学习活动以后学生在学习内容上所达到的学业水平的评价。主要为评价学习成绩而设计,可以是形成性评价,也可以是终结性评价。

学习能力评价旨在对个体的比较稳定的表现在认知水平方面的心理特质的评价,与学生所学的课程和教材内容关系不大,如观察力、记忆力、理解力、想象力和判断推理能力等。这种能力不因外界的影响而轻易改变。学习能力评价又分为一般能力评价和特殊能力评价。

高中学业评价是教学评价的重要组成部分,具有反馈、调控教学并促进学生全面发展的重要功能。2001年6月我国教育部颁布的《基础教育课程改革纲要(试行)》对高中学业评价提出了基本的要求:

改变课程评价过分强调甄别与选拔的功能,发挥评价促进学生发展、教师提高和改进教学实践的功能。

.....

建立促进学生全面发展的评价体系。评价不仅要关注学生的学业成绩,而且要发现和发展学生多方面的潜能,了解学生发展中的需求,帮助学生认识自我,建立自信。发挥评价的教育功能,促进学生在原有水平上的发展。

.....

高中毕业会考改革方案由省级教育行政部门制定,继续实行会考的地区应突出水平考试的性质,减轻学生考试的负担。

高等院校招生考试制度改革,应与基础教育课程改革相衔接。要按照有助于高等学校选拔人才、有助于中学实施素质教育、有助于扩大高等学校办学自主权的原则,加强对学生能力和素质的考查,改革高等学校招生考试内容,探索多次机会、双向选择、综合评价的考试选拔方式。

《纲要》同时指出,在教学过程中,要及时对学生的进行学习进行客观、有效的评价,不断激励学生的学习,及时获取反馈信息,更好地改进教学。学业评价必须以学科的课程标准为依据,遵循既注重结果也注重过程的基本原则,灵活运用各种科学有效的评价手段,对学生在知识与能力、过程与方法、情感态度与价值观方面的表现和水平作出定量和定性相结合的评价。在其实施过程中,应调动学校、教师、学生、家长以及社会各界的积极

性，共同参与对有效学习评价方法的探索。

新世纪之初，我国政府已经启动了包括学业评价内容在内的基础教育改革。这为新一轮的高考改革规划了大体的思路，并指明了发展方向。基于测量和评价理论的发展提供的启示，思考目前在基础教育学业评价中的问题，新课程背景下学业评价的改革应当遵循以下几个原则。

一是能力立意的原则。试题的命制，重在知识获得的过程、方法和能力，而非传统的知识。二是促进学生发展的原则，即如何让考试为基础教育服务的问题。考试要体现对学生发展的促进，这是新一轮课程改革强调的一个核心概念。三是规范性原则，即反映学业评价的规范化。如果考试命题能够形成一种规范化的程序，那么通过这种程序就能最大程度地降低考试中的误差。

第二节 目标分类学的视野

教育目标是预设的教育结果，是教学工作的出发点和归宿，也是教育测量和评价的依据。20世纪50年代提出的认知领域的目标分类学经过半个世纪的运用和完善，经历了从单向的目标分类体系向双向的目标分类体系的发展过程。分类学的产生和发展为考试目标的确立和实施奠定了基础。

一、单向的目标分类体系

1956年，由美国学者布卢姆（B. S. Bloom）等人编著的《教育目标分类学》第一分册《认知领域》正式出版，标志着教育目标分类学的研究拉开了序幕。到1966年，情感和心理动作领域也初步完成了教育目标的分类研究。50年来，布卢姆的教育目标分类学产生了巨大的影响，至少被译介成22种文字。美国全国教育研究学会1994年年鉴曾专门对布卢姆的教育目标分类学40年历程及其贡献作了回顾。

布卢姆的教育目标分类学，以外显行为作为教育目标分类的基点，以行为复杂程度作为分类的依据，把认知领域的学习水平由低到高分六类，即知识、领会、运用、分析、综合和评价；把情感领域的学习水平由低到高分五类，即接受、反应、价值评价、组织和性格化。

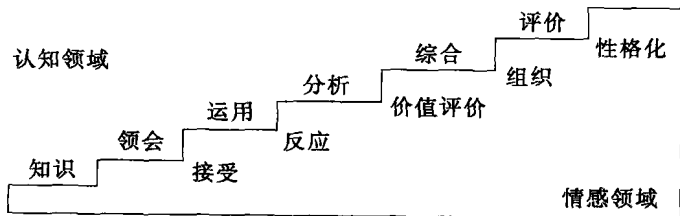


图 1-1 认知领域和情感领域目标层次图

1956年版认知领域的目标由“知识”和“智慧能力与技能”（简称能力和技能）两部分组成。其中，知识目标处于最底层，是最为基础的目标。能力和技能目标的实现都与知识的整理和应用有着密切的联系，都以知识目标为基础。