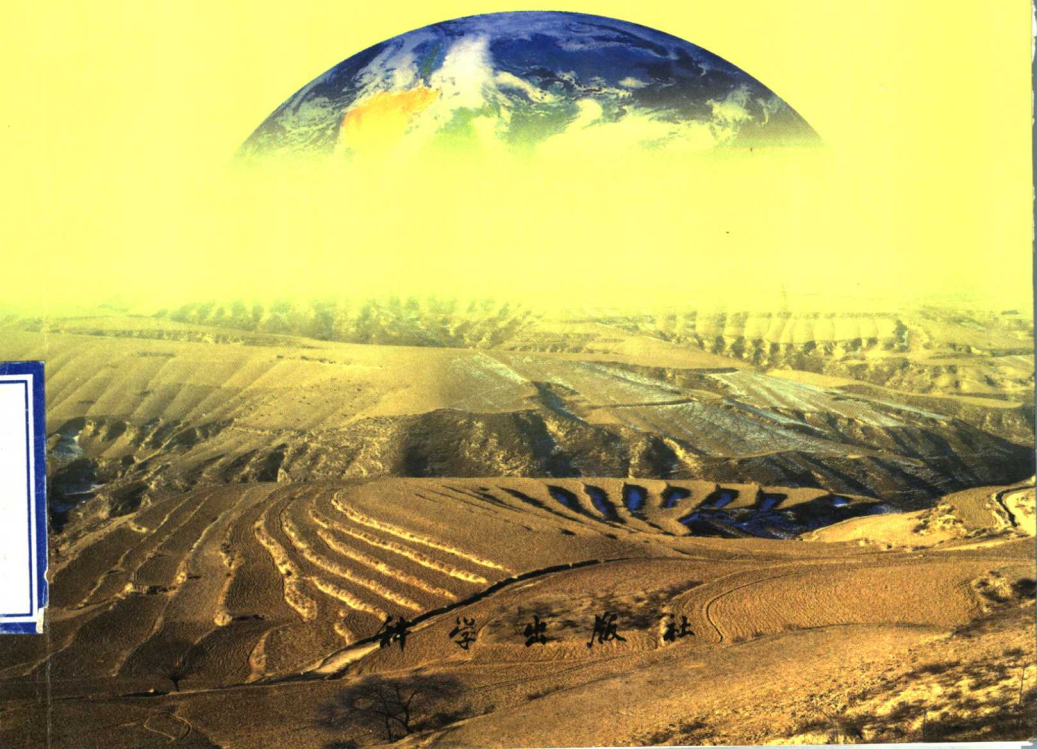


地理教育测量

——理论及其在 **地理高考命题** 评价中的应用

张亚南 著



地理教育测量

——理论及其在地理高考
命题评价中的应用

张亚南 著

内 容 简 介

本书由国家教育部考试中心张亚南著。作者积10余年来高考地理命题的实践经验,系统地阐述了地理教育测量的概念与基本理论,论述了地理教育目标、测量目标、评价目标及其相互关系,阐明了高考地理命题的指导思想、基本方法和命题技术,解释了地理教育测量的评价方法及其应用等。书中列附录6个,分别介绍了1998年和1999年两年的高考地理试题与答案,1999年高考地理试卷抽样统计数据和试题评价报告。是地理教师命题原则、方法和考生学好地理科目、把握高考的重要指导性著作。

本书可供地理教育工作者、地理教育研究人员、中学地理教师,以及广大准备参加地理高考的考生和师范院校地理系学生等参考使用。

图书在版编目(CIP)数据

地理教育测量/张亚南著. —北京:科学出版社,2000.

ISBN 7-03-008322-9

I. 地… II. 张… III. 地理课-高中-升学参考资料

IV. G634.553

中国版本图书馆 CIP 数据核字(2000)第 03891 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

西禄印刷厂 印刷

科学出版社发行 各地新华书店经销

2000年5月第 一 版 开本:850×1168 1/32

2000年5月第一次印刷 印张:6 7/8

印数:1~4 000 字数:178 000

定价:15.00 元

(如有印装质量问题,我社负责调换(新欣))

前 言

教育测量学是具有理论和实践意义的研究领域。它是以现代教育学、心理学和统计学作为基础,运用各种测试方法和手段,对教育过程、教育结果、学业成就及受教育者的能力、品格、学习能力倾向等方面进行科学测定的学科。虽然,教育测量的精确度永远达不到物理测量的水准,但准确、科学、客观、公平的测量永远是教育测量努力追求的目标。为不断接近这一目标,教育测量必须深入研究、探索测量技术及工具的科学性和标准化。而教育测量的主要工具便是测验或考试。

本书力图运用教育测量学的理论和方法,结合地理教育学的理论,阐释地理教育测量的基本理论和原则,论述地理考试特别是地理高考命题的理念、方法和评价。旨在教育测量研究和地理教育研究领域之间建立起联系,并努力探求这种联系的性质、意义及价值。

地理学是近年来发展、变化最为显著的科学领域之一,而教育测量学作为科学的重要研究领域也取得了长足的进展。两个学科都在理论观念和操作方式方面获得了新的研究成果。要将这些成果结合起来并形成一个系统的认识,作者实感学力不足。本书只是抛砖引玉,诚望识者教正。

作 者

2000年1月

目 录

前 言	
第一章 绪论	1
第一节 地理教育测量研究导论	1
第二节 我国地理教育测量的历史、现状和展望	12
第二章 地理教育测量的基本理论	15
第一节 地理教育测量的基本特征	15
第二节 地理教育测量的基本原则	21
第三节 地理教育测量的依据	25
第四节 地理高考评价目标	36
第三章 地理试题的编制	69
第一节 试题构成的要素	70
第二节 题型的分类及其功能	75
第三节 组卷	123
第四章 地理教育测量的评价	129
第一节 试题质量分析与评价	129
第二节 测验质量评析	145
第三节 地理高考评价及其意义	151
附 录	159
一、1998 年普通高等学校招生全国统一地理考试测 试题	159
二、1998 年普通高等学校招生全国统一考试地理试题 参考答案及评分标准	172
三、1999 年普通高等学校招生全国统一地理考试试卷 (广东省)	176
四、1999 年普通高等学校招生全国统一考试地理试 题参考答案及评分标准	189
五、1999 年普通高等学校招生全国统一考试地理考试	

(广东省)抽样统计数据	192
六、1999 年普通高等学校招生全国统一考试地理科试 题(广东省)评价报告	199
参考文献	212

第一章 绪 论

第一节 地理教育测量研究导论

一、教育测量及教育测量学的历史发展概述

1. 教育测量

测量通常是指用量具或仪器来测定物体的尺寸、角度、几何形状或表面相互位置的过程,也包括用仪表来测定各种物理量的过程。简而言之,测量就是用一定的工具来检定某一物体的物理性质的活动。其实质就是用特定的法则确定事物的性质。即如 S. S. 史蒂文斯(S. S. Stevens)所说:“广义而言,测量是根据法则给事物赋予数量。”同理,人们也可以用一定的方法和手段来测定人的知识、思维品质、能力、学术水平、成就等心理特征,对人的心理属性进行定量化研究,即心理测量。在教育过程中,根据一定的法则用数字对教育效果或教育过程加以评量,即所谓教育测量。

教育测量有广义、狭义之分。从广义上说,教育测量泛指运用测量手段对教育活动所进行的量的测定。它涉及的范围很广,凡是需要并能够测量的与教育有关的活动均在研究之列,如教育投入、教育过程各要素、教育的效果等等。从狭义上讲,教育测量专指按一定规则对学生的知识、智能、个性发展等所进行的量的测定。我们通常所提及的多是狭义上的教育测量。本书所讨论的内容也主要涉及狭义的教育测量。

教育测量的本质和原理与物理测量、心理测量是一样的,但在测量的对象、工具、目的、方法、过程、结果等方面则不同。与物理

测量相比较,教育测量的对象是人,这比测量任何物体都复杂,因为世界上不存在两个完全相同的人。所以教育测量是无法直接测量的测量,它只能通过检测心理现象的外显行为或外在表现特征来推知个体的心理能力和个性特点。同时,教育测量很难排除一些无关因素的影响,诸如知识水平、教学条件、师资水平、情绪、健康状况、主试人导向等多方面因素都或多或少地影响到教育测量的结果,并使之出现系统性误差和随机性误差。进行测量必须要有相应的测量工具。物理测量的工具可以有量尺、温度计、计时器、磅秤等等,不同的工具又有不同的量纲单位,使测量结果用以确定测量对象的属性和使测量结果具有可比性或可加性等性质。而教育测量旨在对教育效果进行科学的测定,因而教育测量的主要工具是测验或考试。科学的测量应具有可靠性、客观性。如物理测量中长度的度量,用一把尺子度量某一物体的长度,那么,无论何人、何地、何时进行这项测量工作,其结果应该是一致的,惟此这把尺子才是可靠的。教育测量亦同此理。一个成功的测验或考试,应当是在不同的人使用、不同的时间、地点施测时,其结果仍具有一致性。在分析和评价教育测量的结果时,有两个指标不可缺少:一是表示教育测量的可靠性和客观性指标,即测验或考试的信度;一是反映测量目的的有效性指标,即测验或考试的效度。然而,教育测量的间接性、多元性和随机性注定它比物理测量要复杂困难得多,教育测量的精确度永远达不到物理测量的水准,但准确、科学、客观、公平的测量永远是教育测量追求的目标,要求教育测量必须不断研究、探索测量技术和工具的科学性和标准化。

心理测量是指按照一定规则对人的心理活动所进行的量的测定。它主要侧重于对人的心理现象的测量,反映人的心理活动的差异,主要包括智力测量,个性测量等等。教育测量与心理测量有着密切的内在联系。一方面,教育测量的某些理论与方法借鉴了心理测量的有关内容;另一方面,教育测量在内容上有些就是测定人的心理活动状态。例如对学生的智能发展、以及对兴趣、态度、性格的状况及变化的测定便是借助心理测量的方法和工具来实现

的。教育测量主要是测定学生在教育影响下的心理活动的变化,着眼于教育效果的测量。

2. 教育测量学历史发展概述

19 世纪末 20 世纪初,随着心理测验理论的发展,西方的教育测量学逐渐兴起。这是社会历史发展的必然要求。资本主义大生产和大规模的商品经济,以及与此相适应的迅速发展的近代教育,对客观、准确、高效率地评估人的心理特质与教育成就水平,提出了迫切的要求。近代科学与技术的发展,包括心理学、教育学、数学、统计学以及实验技术等的发展,也为逐步满足这种要求提供了一定的现实条件。心理与教育测量学以比纳(Binet)的智力量表和 E. L. 桑代克(E. L. Thorndike)的书法、拼字、作文、图画、算术等标准化的教育测验量表,以及他们的理论研究的问世为标志,而正式地确立起来。

应该说,对人的心智的评价是人类自古以来的一个愿望。早在西方社会以近代科学手段对人进行测量之前,我国就开始了这方面的探索,只不过由于历史条件的限制,只能凭借政治制度去实现,我国从公元前 11 世纪西周的“试射”迄今,对心理测量的探求已有 3000 年的历史。战国时期的孟子说:“权然后知轻重,度然后知长短。物皆然,心为甚。”在一定程度上表达了心理与教育测量的可能及测验实施的原理。三国时刘劭在《人物志》中提出“观其感变,以审长度”的观点,即说根据一个人的行为变化便可推测他的一般的心理特点。公元 606 年(隋炀帝大业二年),我国始置进士科,从此有了在我国通行长达 1300 多年的科举考试制度。这是一种以选拔为目的、大型统一的纸笔测验,不但对我国古代政治、经济、文化发展产生过深刻影响,而且对近、现代西方文官考试制度的产生,具有借鉴与启发作用。

我国科举制的创立,在测量与评价史上具有划时代的意义,它的积极作用在于建立了完备的考试制度和创造了一整套行之有效的考试方法。科举的考试方法包括口试、帖经(用纸遮盖经书中一

页的大部、仅留一行,再帖上三、五字,让考生按原文填写)、墨义(根据经书出题,要求考生用原文回答)、策论(按题目的要求,写出论证文章)和诗赋(按题目的要求和规定的格律进行创作)。这些方法正是今天的填空、简答、论述和作文等考试题型的渊源。但科举制度也存在着严重的缺陷,其中最突出的是考试内容陈腐且经年不变。它一直延用儒家经学作为考试内容。诚然,在儒家学说经典中存有许多有价值的思想文化,对中华民族乃至世界文化思想都产生了深刻影响。但是,随着社会的不断发展,时代的更替变革,其中若干内容已变成阻碍社会进步的教条,成为影响新的思想文化产生和发展的桎梏,终于在1905年(清光绪三十一年)遭废除。

17世纪中叶,科举考试方法作为评价人才的手段开始介绍到欧美,并得到进一步发展。18世纪以后,英、法、美等许多发达国家开始学用中国的考试方法;至19世纪末20世纪初,实验心理学和心理测验的研究发展推动了教育测量的发展。教育实践的需要和心理测量的发展相结合,促使教育测量作为教育科学的一个领域发展起来。这一时期的测量理论成为了经典测量理论。

辛亥革命后,教育测量学随着西方的科学技术而被引入我国。1918年,我国早期教育理论工作者俞子夷根据E. L. 桑代克(E. L. Thorndike)书法量表的编制程序,编制了《小学国文毛笔书法量表》,开我国教育测量编制之先河。之后,师范高校开始设置测验课程。1931年成立了中国测量学会,使关于教育测量的理论研究得到加强并取得了进展。

1949年以后直到1978年十一届三中全会之前的30年间,由于受原苏联教育理论的影响,教育测量学遭到否定。1978年中国共产党的十一届三中全会以后,教育和心理测量重新开始受到重视,我国的教育测量与评价活动得到空前发展。首先是我国对教育测量与评价的需求日益迫切,党和国家领导人高度重视,社会各界普遍关注;其次是教育测量理论工作者与广大实际工作者日趋结合,教育测量与评价的理论迅速普及,成为指导考试、教学改革

的重要理论依据。1977 年我国恢复了中断十余年的高考制度,被赋予了新的内涵的公平、竞争的理念重又纳入社会生活。目前,如何使高考科学合理,可靠有效,成为社会和学校共同关心的问题。我国的心理和教育测量理论工作者积极探索符合我国国情的标准化考试的道路,使教育测量学理论的研究不断深入和发展,也为我国高等学校入学考试步入科学化的轨道提供了有力的理论支持。

综上所述,不难看出:

(1)教育测量的本质是根据一定的规则对教育本身和受教育者进行测定与评价。测定与评价的目标、方法与社会的发展及当时的社会需求息息相关。如果教育测量的主要工具——考试的实施仅靠社会制度维系,那么当考试的运作与社会的发展发生了难以协调的矛盾时,就会遭到废止。相反,如果注重教育测量的全方位研究,保证其测量工具、方法的科学性和测量目标的时代性,即能体现时代的要求和价值观念,那么考试就是有意义和具有生命力的。

(2)系统的教育测量理论在我国起步较晚。虽然在 20 世纪初西方的教育测量理论就引入我国,但是尚未形成研究规模,更谈不上建立多层次、多种类的测量系统。20 世纪中叶,当经典测量理论在我国起步之时,建立在潜在特质理论基础之上的项目反应理论(IRT)在发达国家尤其是美国迅速发展起来。目前发达国家基本上是将经典测量理论与项目反应理论相结合,形成了一套系统的教育测量体系,并日益显示出其重要作用。从这个意义上说,我国的教育测量理论研究与实践还面临许多工作,要吸收经典测量理论和项目反应理论的优点,根据中国实际,创造出一套自己的教育测量体系。

二、地理教育测量的概念

地理教育测量是教育测量学的研究理论在具体学科中的应用。因此,地理教育测量的概念可以根据教育测量的概念引出,

即:根据一定的法则对地理教育的效果包括学生的地理基本知识、基本技能、地理思维品质等或地理教育过程所进行的量的测定。

这一定义包含三个元素:

1. 地理教育测定的对象或目标

即学生学习或完成地理课程后所具有的地理素质,包括地理基本知识、基本技能和地理思维方法等。通常用地理测验或考试的成绩来表现。

2. 法则

即地理教育测量所依据的规则和方法。这是地理教育测量过程的关键。使用好的法则可以得到可靠的测量,反之可能得到不可靠的甚至是错误的测量。法则的好与差取决于它是否符合客观事物属性的规律以及是否易于制定和便于操作。具体说,作为地理教育测量形式之一的全国普通高等学校招生考试地理科考试,所依据的法则是《地理科考试说明》。其中规定了地理考试的评价目标即能力要求,规定了考查的内容及各部分比例,规定了试题类型、难易题比例、考试时间及答题形式等。从而保证了大学入学考试地理科考试(简称地理高考)考试结果的一致性和可靠性。

3. 量

表示事物的属性。如对学生的地理学业成就、学习能力水平给予数字化的描述,这些数字就显示了学业的差异。

地理教育测量的主要工具是地理测验和地理考试。而测验和考试的涵义并不完全叠合。

目前为大多数心理测验学家所接受的关于测验的定义是:“测验实质上是行为样本的客观的和标准化的测量。”其含义为:首先,测验是对行为样本进行测量。通常,学生的知识、智能、个性等是通过其行为表现出来的。理想的作法是通过一定的刺激,将学生与所测项目有关的知识、技能、观念与思维活动全部诱发出来,引

起他们的行为,然后用一定的标尺衡量。但这将面对浩瀚庞杂的具体知识和技能,使测验难以进行。相反,若仅仅凭借对某一行为的测定而轻易下结论,又会发生很大的偏差,也是不足取的。因此,切实可行的作法是,借助对一组有代表性的行为样本的测定以达目的。例如要测量学生的学业成就,学生的行为是对题目刺激所作出的反应和解答,学生的行为样本赖以题目样本来激发。一次测验的题目就是所测项目的全部题目的一个样本。因此,测验的成功与否,首先取决于题目的代表性。这就要求题目的数量适当,并有较高的质量,能够较全面地反映测量目的和要求。

其次,标准化。是指测验在编制、实施、记分及分数解释方面依靠一套系统的程序。对一个成功的测验来说,除具备高质量的测试题外,还要依据一定的标准和规定组织实施和评分,并尽量控制无关因素的影响,为所有考生提供相同的测验条件,力求准确地测得所要测量的东西。所以从整体上看,构成测验的各个环节彼此有着密切的联系,是一种统一的系统程序。

第三,客观测量的评价指标。测验的标准化是测验的客观而科学的测量所要求的。但绝对的标准化是不存在的。于是,人们设计了测验客观性程度的几个评价指标,以使测验的客观性、科学性尽量接近标准化测验的要求。指标是:①题目的质量分析,包括难度和区分度;②信度,指测验结果的可靠性程度;③效度,指测验结果的有效性程度,这是评价测验质量最重要的指标。

考试作为一种教育测量与评价的术语,其意义与测验很相近。所谓考试是由主试者根据一定的社会要求,采取一定的方式方法,选择适当的内容,对应试者的才学、德识、技能等所进行的测度或甄别活动。也是对人的知识、智能、品德等所进行的一种测量。在很多情况下,考试与测验可以彼此互相替代,特别是在测量与评价学业成就时尤其如此。但是,严格地说两者是有一定差异的。主要表现在:

(1)测验的形式应用范围较为广泛,如非正规场合下的平时教学小测验、单元测验、时事小测验等;和用于智力、个性等方面的测

量,如心理测验、各种常识测验、智力测验、能力倾向测验、人格测验等。而考试一般指用于比较正式场合下的测验,如期中或期末考试、毕业考试、自学考试、升学考试(中考、高考)等。

(2)测验的目标一经确定,测验的内容一般变化在一定的范围,且测验可以经过相当一段时间的使用,逐步达到标准化的过程;而考试则是目的性较强的测验,常会受到非智力性因素如时间、目的、地域等的影响,且考试的标准化进程比较艰难。

本书将主要讨论地理考试,具体说是大学入学考试地理科考试试题的编制及其考试结果的评价。

三、地理教育测量的功能

地理教育测量的功能既反映了一般教育测量的性质和特点,同时又具有学科测量自身的功能性特点。

1. 因材施教

教育的一条基本原则就是“因材施教”。实际的教育过程欲体现这一原则,则要求教师必须了解自己的学生。其依据有:①基于经验之上的主观直觉;②借助于测验或考试的客观测量。前者是前科学的方法,有时并不可靠,常常在很大程度上受到主体状态的影响。例如对学生的地理能力、知识水平等方面的评量,主观直感可以在一定范围内使教师了解学生的特点,但这并不完全可靠,很容易受到教师主观意识的支配,如好恶、成见、情绪等多方面因素的影响。而使用地理测验并依据测验或考试结果来确认学生的地理学习效果才是客观、准确的,这是了解学生真实水平的重要方法和渠道。唯如此,教师才有可能针对学生的具体情况做出相应的合理安排,依据学生能力和已有地理知识水平的个别差异做出适当的教学决策,如补习、分组、开特殊课、课后个别辅导等等。有针对性、目的性的教学应该是效果最好的教学。

2. 区分差异

测验和考试是有意识地测量考生个体和群体(如学校、地区、全国、各国)间差异的活动。测验和考试的区分功能是指通过测验和考试区分被试在知识、能力等方面的相对高低,使不同水平和能力的考生的测验、考试成绩拉开距离。

体现在日常的地理教学过程中,就是通过地理测验、考试把学生的地理学业成绩用百分制、五分制或等级制,将学生成绩划分为优秀、良好、及格、不及格等等级,以区别学生在地理知识的获得和地理能力方面的差别,给优者以鼓励,给劣者以警示,给教者以信息。

考试的区分功能在升学考试中体现得更为充分。考生在接受相同时间、相同内容的教育,并在同一时日、回答同一套试题、采用统一评分标准阅卷评分的条件下,获得不同的考试成绩。其分数差异可以区分考生的学习效果、思维品质和是否具有进入上一级学校学习的能力及潜力,并以此来决定考生进入不同类别、级别的高一级学校。如参加高考的学生可根据成绩分别进入全国重点院校或一般院校;进入本科院校或专科院校。初中毕业参加升学考试的考生通过考试成绩来区分考生是升入重点高中,还是一般高中或职业高中。总之,考试的区分功能为各级毕业生的有序分流提供了公平可靠的依据。

3. 选拔人才

教育测量是一种选拔人才的手段。为了保证教育质量,培养出高层次高水平的人才,并最高效益地使用好有限的教育经费和教育设施,必须甄选出最有可能成功的学生。随着社会与科技的发展,凭借个人经验的选拔已无法满足实际的需要。而测验和考试这种的大规模、公平、竞争的特性及其能高效、准确地检定被试水平、划分被试等第等优势使其天然成为选拔人才的手段和工具。地理考试亦同样具有此项功能。如:全国普通高等学校入学统一

考试地理科考试、全国成人高等学校入学考试地理科考试、研究生入学考试地理专业考试等,都是为选拔、造就有地理专业学科研究志向和潜力的或具有完备知识结构及文化底蕴的高级优秀人才服务的。

4. 诊断补救

诊断本是临床医学术语,指医生根据就医者的症状判断,确认病情的活动或过程。测验作为诊断疾病的一种手段,始于本世纪初,目的是为了了解那些心理上有缺陷的儿童,以对他们进行早期矫治。

测验与考试的诊断功能可以如临床诊断一样体现在教育过程中。影响和造成学生在校适应不良、学习困难的因素会很多,如表现在地理学习过程中的缺乏地理学习动力;数、理、化基础薄弱,理解掌握自然地理原理有困难;某章教材难点集中;缺乏地理基本技能训练等等。地理测验和考试能够帮助教师找出问题症结之所在,为有效地解决问题提供较可靠的信息。教师可以依据这些信息,有针对性地采取补救措施,对症下药,解决问题,达到最佳教学效果。

5. 评价教学

地理测验和考试的评价功能在教学评估中既可面向教师与教学方法,亦可面向学生与学业成就。测验和考试可以选拔与评定优秀的地理教师,不断提高师资水平;还可以评价和鉴定地理教材与教学方法,从而提高地理教学质量。根据地理科单元考试,期中、期末考试,会考、高考等,可以对学生的地理学业成就作出评定,决定学生的等级成绩和升学就业。通过测验和考试,教师得以了解学生,学生得以自我了解和自我评价。

以上是地理教育测量的一般性功能,但在我国目前的教育环境和条件下,地理考试,确切地说是大学入学考试地理科考试被赋予了一个独特功能,即:

6. 保证中学地理教学秩序,提高中学地理教学质量

多年来,学校中的地理教育由于地位不明确而常常受到忽视,甚至非难。这种情况普遍存在于世界各国包括发达国家的学校中。德国著名地理学家阿尔芙雷得·赫特纳曾形象地把学校的地理科比喻为“小媳妇”和“杂役”,以说明其低下的地位和不被重视的景况。

近几年来,我国因在高等学校入学考试中取消了地理科而使中学地理教育陷入低谷。地理课时常常要让位给列入高考的学科;许多优秀地理教师因无法实现工作价值而流失;学生们则在应付完会考之后基本放弃了地理科的学习,地理素质普遍下降。由此不难看出,高考科目对中学学科教学的重要影响,也应当认识到,考试施于教育乃至社会的作用。

事实上,影响学生学习的因素有许多,激发学生的学习兴趣,鼓励学生主动、积极地获取知识的手段也很多。心理学的研究表明,个人学习态度的好坏,获取知识的多少(学习成绩),主要的的是由学习动机推动的。学习动机不仅决定或支配着学习行为活动,而且影响着学习者思维的发展。据调查,目前青少年学习的主导动机仍是为获取较好的分数、为升学、为未来的出路和幸福而勤奋学习。因而许多教育学家提出以考试来促动教学的思想(有些国家已付诸实施)。尽管这种提法多少带有些功利性,尽管考试和考试制度的弊端能被轻而易举地罗列和受到谴责,然而通过考试这一手段可有效地影响和改变学习方法及学习过程则是不争的事实。如果我们能公正客观地评价、接受考试的作用,从而科学地研究考试的理论和方法,使考试真正成为提高教学质量、发展思维、培养能力的手段,那么,在客观上实难以扭转的“应试”则将被赋予积极的意义。学校正常的教学秩序会得到充分保证,教学质量也会不断提高。

随着高考改革和基础教育课程改革的不断深入,我国中学地理教育受到忽视的情形有望改观,希望赫特纳早在本世纪 20 年代