

《考试对策》丛书

考试方法

王建新
姜秀英
董

编写

《考试对策》丛书之三
考 试 方 法

王建新 姜秀英 董 冀 编著

黑龙江科学技术出版社

责任编辑：张锦新

封面设计：步利君

版式设计：邹旭昌

· 考 试 方 法

主建新

姜秀英 编

董 冀

黑龙江科学技术出版社出版

(哈尔滨市南岗区建设街35号)

双鹤印刷厂印刷·黑龙江省新华书店发行

787×1092毫米32开本5.625印张115千字

1990年8月第1版·1990年8月第一次印刷

印数：00001—10.000册 定价：1.90元

ISBN 7-5388-1037-4/G·203

《考试对策》丛书序

考试，这是一个令人生畏的字眼，因为它有生杀予夺之力，可在很大程度上决定人们的命运。世界上不怕考试的人是不多的。

考生，这是我们每个人都体验过的角色，从小学、中学到大学。可谓身经百战，应考无数。过关斩将，金榜题名之喜；考场失意，名落孙山之忧，酸甜苦辣、一言难尽。

考试的历史悠久，源远流长。虽然考试的内容和形式随着人类的进步不断发展，但考试作为评价教学效果和选拔人才的方法，其职能却从未改变。各行各业都离不开考试，取消考试的苦果我们已经尝过了。

考试弊端象其功能一样明显。古有50中举，喜极发疯的范进；今有高考落榜，饮恨黄泉的青年。古往今来，陷入“考试地狱”的人难以数计。

为了更好地发挥考试的积极作用，克服其消极影响，我们编辑了这套《考试对策》丛书。

如何正确对待考试，如何有效准备考试，这是考生关心的热点，也是本丛书所要回答的问题。

诚然，一个人的考试分数主要取决于他对应备知识掌握的程度；但会不会考，无疑也能对成绩产生影响。两个知识水平相当的人，一个受过应试技能的训练，临场镇定自若，另

一个应试经验不足，临场焦虑万分，二人的成绩便可能大有出入。把应试技能与克服焦虑的方法公之于众，使考生都能掌握，大家在同等条件下竞争，会提高考试的可靠性和有效性。

本丛书由国家教委考试管理中心兼职研究员、全国教育统计与测量学会理论研究部部长、北师大心理系副主任郑日昌教授领衔主编，其他编委和执笔人员亦多是有关学科的优秀教师、研究人员。在写作过程中，他们根据国家教委颁发的教学大纲，悉心钻研教材，对近几年来高考、中考试题作了深入分析，并结合考生易出现的问题，有的放矢地提出了应考对策。此外，一些书中还对目前我国刚刚兴起的标准化考试的理论和方法作了概略介绍。

“丛书”注重科学性、实用性、针对性，通过传授考试对策，加强对考生分析问题、解决问题能力的培养。经过诸位同仁的携手努力，目前虽不尽如人意，但它毕竟作为一份答卷奉献在读者面前，如果对广大中学生朋友和教育界同行有些微帮助，便是“丛书”作者、编者的莫大安慰，

是为序。

《考试对策》丛书编委会

前 言

近年来我国对于考试的研究已十分重视，考试方法也在不断改进，逐渐实行较为科学的标准化考试，令人瞩目的高考的标准化程度提高较快，受其影响，其它的考试也在逐步改革。但考试方法改革并非仅几个人或某个部门的事情，而是涉及到全体教育工作者的事情，可是，有些教育工作者对科学的考试方法并不掌握，同时，考试方法中的许多问题还需进一步研究，本书将对考试方法中的有关问题加以讨论，并侧重对标准化考试方法加以介绍。

对于考生来说，考试关系到能否升级、毕业、升学或被录用等等。考生需要不断适应，了解考试方法，以考出自己的真实成绩。作者期望本书不仅有益于广大教育工作者的，也同时有益于广大考生。

编 者

《考试对策》丛书编委会

主 编：郑日昌

副主编：董 冀、李 哲、包瑞全、孙玉峰
邹长海、王建新、张 岑

编 委：（按姓氏笔划）

于 华、于福贵、张吉兰、冉德润、
孙万鹏、毕 进、季春山、武静环、
陈友新、徐荣聚、王海全、乔元春、
曹大庆。

目 录

第一章 考试方法概述	(1)
第一节 考试概述.....	(1)
第二节 考试方法的类别.....	(6)
第三节 考试的科学化与标准化.....	(8)
第二章 考试目标	(12)
第一节 考试目标的定义.....	(12)
第二节 制定考试目标的意义.....	(14)
第三节 确定考试目标的依据.....	(16)
第四节 表述好的考试目标的特征.....	(24)
第五节 制定考试目标的方法.....	(29)
第三章 试题编制技术	(34)
第一节 试题选材与设计蓝图.....	(34)
第二节 命题原则及常用的试题形式.....	(37)
第三节 试题质量分析.....	(55)
第四章 题库的建立	(61)
第一节 题库的结构.....	(62)
第二节 建立题库的方法.....	(64)
第五章 试卷的编制	(72)
第一节 编制试卷步骤及要点.....	(72)
第二节 评分标准的确定.....	(75)
第六章 考试的实施与评分	(88)

第一节	考试与答卷纸型	(88)
第二节	考试时间的安排	(94)
第三节	考场设置及考试制度	(95)
第四节	监考人员	(97)
第五节	如何防止和处理舞弊行为	(99)
第六节	评分方法	(102)
第七节	开卷考试与口试的实施	(111)
第七章	考试质量分析	(114)
第一节	考试误差的来源	(114)
第二节	考试的信度	(118)
第三节	考试的效度	(134)
第八章	考试分数的解释	(147)
第一节	原始分数和导出分数	(147)
第二节	怎样看个人分数	(154)
第三节	怎样看班级分数	(165)

第一章 考试方法概述

第一节 考试概述

考试在我国有着悠久的历史。远的不谈，仅从隋炀帝606年设置进士科，创科举取仕考试制度开始，至1905年清朝推行新政前的1300年中，选拔人才大都是以考试结果为准。

16世纪后期，我国的考试制度传入西方，1791年法国参照我国的科举制度建立了自己的文官考试制度。随后，英美两国也相继效仿建立考试制度，打破了门第或名人推荐的受官任职制度。

考试作为检查学生学习质量的手段也由来已久。早在1219年，古代波兰大学法学考试就开始运用口试。1599年耶稣会颁布了笔试条例(在中学实行)。1787年普鲁士毕业考试采用论文式的考试。到19世纪，欧美的发达国家逐步建立了学校考试制度，并把考试同升学、就业、奖学金等联系起来。有的国家还专门制定法规，例如，法国政府1933年9月立法规定六年级(11岁)学生一律参加统一考试，使小学和中学预备班接近。1936年8月立法规定义务教育延长到14岁，小学毕业可参加竞争性升学考试，及格即升入中学，否则进补习班或职业学校。又如苏联为改进学生的学业成绩和操行评定，1944年1月开始在学校实行五级记分制(保留至今)。同年6月苏联人民委员会公布了《关于提高学校教

学工作质量的措施》的决定。按照这个决定，初等学校和七年制学校以及中学毕业生必须参加毕业考试。毕业考试包括学生在校学习的全部课程，对于品学兼优的中学毕业生颁发金质奖章和银质奖章。1977年苏联高等和中等专业教育部又详细制定了学生考试评分的标准，1981年又对这个问题发了专门指示信。美国本来中小学多采取一律升级制度，为保证教育质量，1975年佛罗里达州率先以州立法形式规定全州中学生升学或毕业的最低学力标准，特别是要求举行严格的高中毕业统考。这种做法以后陆续为其它30多个州所仿效。

在我国考试仍是学校教育活动的的重要组成部分，以考试为手段，检查学生的学习质量和教育效果，决定升留级、毕业、结业和升入高一级学校。除一般学校中规定的小考（平时测验）、期中考、期末考外，国家还举行全国统一招生考试。

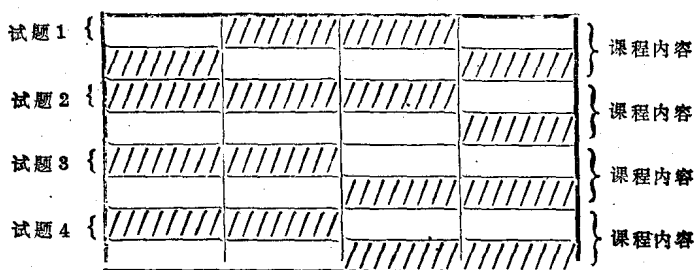
另外，当前还出现了一些大规模的考试专业机构，最新科技成果（如电脑、光电阅读器等）也在考试中得到应用。最著名的考试专业机构是美国的教育考试服务社，它成立于1947年，主要举办高等学校入学考试（SAT），研究生资格考试（GRE）和作为外语的英语测验即“托福”（TOFEL）考试等。

在较长的时期内，考试方法并不科学。例如我古代科举考试中的代圣贤立言的八股文考试，就是十分落后的考试方法。自19世纪以来，欧美各国考试内容已发展成为学校学生所学的全部或部分课程，考试方式以笔试为主，有的还采用口试、操作等方式，我国目前在学校教育活动中所采取的考试方法等，仍以传统考试方法为主，这存在许多问题，主

要是:

(1) 试题量小, 不能覆盖全部课程内容。传统考试一般是以极少几个题目进行, 即使这几个题目必须以课程内容知识作答, 也只需其中部分知识, 如果学生这些部分知识恰巧掌握得好, 则无疑可得高分, 反之则得低分。知识掌握相同的学生, 则可能由于掌握点不同, 得分不同。如图1—1—1 A、B、C、D 4个掌握教材内容都为50%的学生得分情况

图 1—1—1



得分: A、75分 B、100分 C、50分 D、0分

图中四个部分代表A、B、C、D四个学生掌握教材的状况, 阴影部分是已掌握的课程内容, 空白部分是尚未掌握的。4个学生都掌握了课程的50%, 但由于4道试题不能覆盖全部课程内容, 则使得4个学生测验所得分数截然不同, 以百分制计, A得75分, B得100分, C得50分, D得0分。这样对于一个群体来说, 要看平均掌握情况还问题不大, 但对每一个学生来说, 如果考察他的学习质量情况误差就太大了。另外这种考试也会引起“题海战术”和“押题猜题”等影响双基教学目标培养的恶果。

(2) 传统考试方法中许多试题往往是答案和评分没有固定的标准，或者正确答案可以有多种表达形式。例如几何证明题，被试者可以通过不同的证明途径而获得正确解答，一个作文题目，被试做出来更会各有千秋。这类题目，就难以客观地按某个标准评分，而是评分者主观地评分。同一份卷子答案落在不同的评分者手中，即使评分者水平一致，他们所作的评分也会有很大差别。例如国外曾有许多教授同评一份历史学答卷，但分数却为40分至90分不等，差别甚大。又如，斯达奇与伊立欧特曾将同一份英文试卷请142位英文教师评阅分数，结果这一份试卷所得分数从50到98分不等。共得35种分数。后来它们又将同一份几何试卷请107位中学数学教师评分，结果评出42个分数，最低为28分，最高为89分，相差61分之多。我国也曾有人在北京、广州、上海等地作过类似实验，其结果也都能看出主观评定造成很大误差。

(3) 试题的随意性。各类考试考什么，是考知识还是考能力？考哪些知识和能力？该各占多少分？怎么考？这些往往由命题者主观决定，随意性很大，难以与其它同类考试结果进行比较。

(4) 偏重对知识的考试，而缺少对能力的考试。目前各类考试，主要限于笔试，难以考察到非书面形式才能表现的各种能力。如语言的听、说能力，理化或生物的实验操作技能，工具和仪器的使用或制作能力，等等。

(5) 分数合成不合理。分数合成是考试的一项重要工作，一次考试往往要涉及几科，这几科在全部课程中所处的地位或所占比重是不同的，学生对各科所需投放的精力和时间也不同。但考试分数却往往是每科都按百分制计，最后合

成总分，作为学生的总成绩。实际上这样的合成方法是不合理的。一门学科的考试，涉及的各部分知识内容和要求程度不同，比重也不同，但一般考试时给各题分值却往往不考虑这一点，而是习惯地按题目在卷中所占量的大小来确定，与所考察的知识内容并不完全对应。这样，一份由若干个试题组成的卷子，各题得分合成该卷总分，也难以合理。科学的作法是不论多科考试合总分，还是一科考试赋给每题分值，都应考虑“权重”的分配问题。然而这一点在考试中往往被忽视。

(6) 浪费人力和时间。学生答题时要写大量的文字，而这些文字中又可能大部分是多余的。这就增加了考试的时间和不必要的工作，使考试题量受到了限制。同时，学生作的答案文字过多也给教师阅卷增加了工作量，加上学生有时对自己不能确切回答的问题采用模棱两可或含糊不清的语言掩饰，也给教师评分造成困难。再者，这样的卷子只能用人工评卷，无法利用现代化技术设备。

从本世纪初以来，心理与教育测验的理论和技術有了很大发展，这对考试方法产生了直接的影响。随着标准化测验的出现，标准化考试也开始应用于教育活动之中。所谓标准化考试，即是根据教育测验原理和方法进行的考试。由于这种考试方法基本上能克服传统考试方法中存在的问题，所以倍受人们的重视。我国从1977年恢复高考后，教育出现了生机，教育测量学也获得新生。近年来，教育测量学的研究有了很大发展，随之考试方法改革也在进行，国家对考试改革十分重视，在高考逐步推行标准化考试。1985年开始在广东省进行了数学和英语两科的标准化考试试验，1986年英语标

准化考试扩大到广东、辽宁、山东和山西省，同时在广东省增加了物理学科的标准化考试试验，87年高考在广东、山东、四川、辽宁、广西、湖北、陕西和北京等省市对部分课程实行了标准化考试，并用“光电阅读器”阅卷。关于考试的科学研究，从1981年以来，伴随着教育统计、教育测量和教育评价的研究也活跃起来，尤其对于标准化考试的研究，更为广泛和深入。还出现了一些专门研究考试的报刊，如《考试报》和《考试研究》以及涉及考试研究的《教育统计与测量文选》等。

第二节 考试方法的类别

考试方法可按不同的分类标准进行分类。根据我国当前的考试方法看，可作如下划分：

(1) 按考试形式可划分为口试、笔试和操作考试等。

口试，一般用于考察被试的语言表达能力等，常为外语考试和选拔性考试所用。口试前，主试要准备好大量试题，根据试题的性质、难易程度适当搭配，做成很多考签。考试开始后，由考生抽签后略做准备后作答。口试可以比较深入而确切地了解被试的一些有关情况。

笔试是要求考生在规定的时间内在试卷上作出书面回答的一种方法。由于这种方法简单、省时省力，可同时进行大规模的考试，便于比较准确地评价和比较考生的成绩，所以应用得最多、最为广泛。但目前看，笔试在实际应用中的偏重记忆、主观性、内容覆盖差、以及处理上的一些缺陷还远未得到纠正。要使笔试科学化，须根据测验原理和方法改进，

使笔试客观化和标准化，即采用标准化考试方法。

操作考试是通过考生进行操作而进行的一种考试方法。考试步骤与口试大致相同，它主要用于检查考生掌握操作技能及有关工具 and 仪器的使用和制造能力和理论联系实际的能力。

(2) 按考试目的可划分为达标性考试、选拔性考试、诊断性考试等。

达标性考试也称为合格性考试，是用于检查被试是否达到规定标准的考试。如学生的毕业或结业考试、体育锻炼达标考试、技术级别定级考试等。这种考试由于是以一定标准决定考生是否合格，一般不存在考生间的竞争问题。

选拔性考试也称为择优性考试，是用以选拔合格或优秀人才的。例如各类高、中等专业学校的招生考试，招聘职员或专业技术人员的考试等。这种考试由于只能取其中优胜者，所以竞争激烈。这类考试不仅要有考察考生现有知识和技能水平的内容，还要有能够预测考生是否具备未来在某方面成功的可能性的考试内容。

诊断性考试是用以检查被试某些方面的特殊优点和缺点，偏重于发现学生在学习上的困难之处，作为改进教学方法或补救的依据。模拟考试亦属此类。

(3) 按命题范围可划分为学科考试和综合考试。

学科考试是用以检查学生学习某学科后的知识和技能掌握情况。学科考试是教育活动的主要考试。

综合考试的试卷题目涉及几门学科的内容，用以检查考生知识掌握情况和综合运用各学科知识的能力。这类考试在教育活动中用得较少，近来硕士研究生招生考试中有时使

用。

(4) 按考试规模可划分为全国性统考、地区性统考、校内班(年)级考试等。

全国性统考有组织严密,有统一的考试大纲、复习范围、试卷和考试时间,评卷时按统一标准进行。

地区性统考是指按计划局局部区域性的统一考试,这种考试用于地方学校招生、达标验收或者检查地方、学校教育质量。

校内班(年)级考试是学校对在校生所学的各学科进行的阶段性考试。这类考试是由任课教师或教研室(组)命题和评卷,如学校进行的期中、期末考试,基本上属这一类。

(5) 按考试方法可划分为传统考试和标准化考试。

第三节 考试的科学化与标准化

科学的考试,首先应当是试题能较好覆盖考查内容,其次是具有较高的信度、效度和区分度,以及科学的实施过程等。所谓信度,是反映考题稳定性和可靠性的指标。当同一组被试用同一试卷考试两次后,所得分数一致,或者同一组被试经过一次考试后用另一个等值试卷再考一次,两次所得的分数一致,则这样的考试都是可靠性或可信程度很高的,即信度高。所谓效度,是反映试题的准确性和有效性的指标,它反映着对所测量的东西能够测量到的程度。例如要检查被试的计算能力,则必须用被试必须具有这种计算能力才能解答的试题来考,而不能用与计算能力无关的试题。区分度是指试题对不同考生具有的鉴别能力的程度。区分度高是说