



新世纪高等学校规划教材·数学教育系列

数学考试 理论与方法

主编◎朱文芳

SHUXUE KAOSHI
LILUN YU FANGFA



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社



新世纪高等学校规划教材·数学教育系列

数学考试 理论与方法

主编◎朱文芳

SHUXUE KAOSHI
LILUN YU FANGFA



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

图书在版编目(CIP)数据

数学考试理论与方法/朱文芳主编. —北京:北京师范大学出版社, 2017. 6

新世纪高等学校规划教材·数学教育系列

ISBN 978-7-303-22304-6

I. ①数… II. ①朱… III. ①教育数学-考试-高等学校-教材 IV. ①O1-49

中国版本图书馆 CIP 数据核字(2017)第 100675 号

营销中心电话 010-62978190 62979006
北师大出版社科技与经管分社 www.jswsbook.com
电子信箱 jswsbook@163.com

出版发行: 北京师范大学出版社 www.bnup.com
北京市海淀区新街口外大街 19 号
邮政编码: 100875

印刷: 北京京师印务有限公司
经销: 全国新华书店
开本: 787 mm×1092 mm 1/16
印张: 15.75
字数: 327 千字
版次: 2017 年 6 月第 1 版
印次: 2017 年 6 月第 1 次印刷
定价: 32.00 元

策划编辑: 刘风娟	责任编辑: 刘风娟
美术编辑: 刘超	装帧设计: 刘超
责任校对: 赵非非	责任印制: 赵非非

版权所有 侵权必究

反盗版、反侵权举报电话: 010-62978190

北京读者服务部电话: 010-62979006-8021

外埠邮购电话: 010-62978190

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 010-62979006-8006

内容简介

随着对数学教育专业实践性要求的不断提升,我国一些高等院校数学教育专业,为研究生陆续开设了“数学教育测量与评价”的课程。开设初期,由于这门课程还没有形成自己的学科体系,基本上借用教育测量或心理测验的名词、术语、理论和方法,来解释、说明数学教育过程中产生的一些问题。

众所周知,数学教育实践中,数学教育的测量与评价活动是最经常发生的,并且内容最为丰富。特别是数学考试,是我们最熟悉不过的。事实上,从近几十年来,国际上各种大规模的教育测量与评价活动中可以发现,其中最先关注的就是数学测试。同时,“数学教育测量与评价”这门课程,最独具特色的也是数学考试。为此,在建设“数学教育测量与评价”课程时,我们首先完成关于“数学考试”的理论探索。由于学校教育中所暴露出的数学教与学对考试的严重依赖,还要求我们结合数学考试的实际情况,来满足数学教育实际的需要。

本书梳理了数学考试简史,概述了数学考试的功能与类型,详细介绍了数学考试的基本理论,包括怎样科学地命制数学试题,以及数学考试的计算机化发展,数学考试分数的意义等,最后简单介绍了对学生具有重要影响的数学考试。

本书可以作为高等师范院校、教学研究机构、教师进修学校中数学专业本科生、研究生的数学教育测量与评价课程教材,也可以作为关注数学考试的教师、学生、家长等相关人员的参考书。

前言

数学考试是一个影响了我们现代人成长的重要事情。虽然，现在已有许多关于数学考试的研究成果，但是大多数是从应试的角度加以研究的。对数学考试系统化、科学化研究的成果还比较少。

1991年，根据《中国数学奥林匹克教练员评审条例》，我被中国数学会奥林匹克委员会授予贰级教练员称号。从那时起，我开始在北京市数学奥林匹克学校(高中部)任教练。1993年，我参与编写了《奥林匹克中小学系列教材 高中数学》(北京师范大学出版社)和《最新初中数学奥林匹克专题讲座与解题技巧》(中国科技大学出版社)。2000年，我又参与编写了《小学数学竞赛辅导》(北京大学出版社)。从此我开始关注数学考试中的试题。

2005年，我开始给北京师范大学数学科学学院数学教育专业的研究生，开设“数学教育测量与评价”的课程，其中，主要探讨数学考试科学化的问题，并指导本科生、研究生，重点研究中考数学与高考数学等相关问题。

2007年春天，我开始参与教育部基础教育质量监测中心的筹备、建立，以及对中小学数学学习质量进行监测研究等工作。这一年的11月30日，教育部基础教育质量监测中心正式成立；我参与完成了国家对3个省的数学监测试点工作。从此，我将研究的工作重点全面转向数学考试。但是，由于数学考试结果的敏感性问题，从2007年一进入教育部基础教育质量监测中心，就被纪律要求填写了保密协议。随后，2008年，参与完成了国家对8个省市的数学监测试点工作。2012年、2014年，参与完成了对全国大部分地区的数学监测试点工作。2015年，我带领一个研究团队代表北京师范大学数学科学学院，参与了国家正式对8年级学生的数学学习质量监测工作。

经过了这20多年来对数学考试的理论与实践积累，我觉得可以总结一下对数学考试的认识，抛砖引玉，为提升我国数学考试的科学化水平尽微薄之力。下面的拙作，就是基于这种想法的一个总结。特别要感谢曾经与我一起为教育部基础教育质量监测中心做研究的数学老师，选上我开设的“数学教育测量与评价”课程的研究生们。感谢我所指导的研究中考或高考数学的学生们，是他们的工作与论文极大地丰富了本书的内容。

数学考试的理论与方法是一门比较新的学科，有许多问题尚处于研究阶段。所以本书并非高度抽象而概括的论著，只是多年来我们一起探究的想法与结果。希望读者把本书当作一个起点，通过思考数学考试中的诸多问题，从此走上一条探索之路。同时，希望对从事数学教育测量与评价课程建设的老师，以及想了解数学考试理论基础的人能有一些帮助。但能否如愿还有待实践的检验。当然，由于作者水平有限，书中难免存在各种疏漏，敬请各位专家、学者和广大读者提出宝贵的意见和建议。

2016年12月6日于北京师范大学

目 录

第1章 绪 论 /1

1.1 数学考试简史	1
1.1.1 科举制度(606年以前)之前的数学考试	1
1.1.2 科举时期(606—1905)的数学考试	2
1.1.3 西方的数学考试	4
1.1.4 数学考试方法	5
1.2 智力测验与数学考试	6
1.2.1 对人的记忆的测量	6
1.2.2 智力测验	7
1.2.3 莱斯的拼字测验与选择题的发明	8
1.3 大型国际测评中的数学考试	8
1.3.1 国际数学与科学研究趋势 TIMSS 中的数学测试	8
1.3.2 国际学生评价项目 PISA 中的数学测试	12
1.4 我国中小学数学考试的现状	23
1.4.1 我国现有的中小学数学考试	23
1.4.2 对我国现有中小学数学考试的分析	24
1.4.3 世界各国纷纷进行基础教育质量监测	24
1.4.4 构建具有中国特色的数学教育质量评价体系	26

第2章 数学考试概述 /30

2.1 数学考试的功能	30
2.1.1 选拔	30
2.1.2 评估	30
2.1.3 诊断	31
2.1.4 研究	31
2.2 数学考试的类型	31
2.2.1 数学常模参照性考试	31

2.2.2 数学标准参照性考试	32
2.2.3 数学常模参照性考试与数学标准参照性考试的比较	34
2.3 测量与教育评价的概念	35
2.3.1 测量的概念	35
2.3.2 测量的种类	36
2.3.3 教育评价	38
2.4 数学能力与考试难度、区分度	39
2.4.1 数学考试的测量与评价含义	40
2.4.2 关于数学能力的构想	40
2.4.3 数学能力与数学考试的关系	41
2.4.4 数学考试测量的究竟是什么	42
2.4.5 数学考试怎样实现可测	43

第3章 数学考试的基础理论 /52

3.1 数学考试的误差	52
3.1.1 数学考试的误差	52
3.1.2 数学考试的效度	54
3.2 数学考试的信度	56
3.2.1 信度的含义	56
3.2.2 信度的求法	56
3.2.3 关于信度系数的解释	60
3.2.4 SPSS 软件中信度分析实例	61
3.3 数学考试的经典测量理论	64
3.3.1 数学考试的真分数模型假设	64
3.3.2 经典测量理论的相关概念	65
3.3.3 使用经典测量理论判断数学考试质量的局限性	66
3.4 数学考试的项目反应理论	67
3.4.1 基本概念与数学模型	67
3.4.2 项目反应理论的基本假设	70
3.4.3 项目反应理论中的参数问题	71
3.4.4 项目反应理论在数学考试中的应用	72
3.5 数学考试的概化理论	74
3.5.1 概化理论的相关概念	74
3.5.2 概化理论的研究过程	75
3.5.3 概化分析中的统计方法	77

第4章 数学试题的命制方法 /82

4.1 数学选择题的命制方法	82
4.1.1 命制数学选择题语言上的要求	83
4.1.2 题干的编写	83
4.1.3 选项的编写	85
4.1.4 选择题的质量分析	92
4.2 数学填空题的命制方法	93
4.2.1 命制数学填空题的语言要求	95
4.2.2 数学填空题的空白数及位置	97
4.2.3 数学填空题的质量分析	99
4.3 数学解答题的命制方法	102
4.3.1 数学解答题题设的命制要求	105
4.3.2 数学解答题要求的命制	106
4.3.3 数学综合解答题的设问命制	114
4.3.4 数学综合解答题的质量分析	118
4.4 数学开放题的命制方法	121
4.4.1 数学开放题的特点	121
4.4.2 命制数学开放题的要求	124
4.4.3 几种数学试题类型的简单比较	126

第5章 数学试卷与题库建设 /129

5.1 数学试卷的编制依据	129
5.1.1 什么是数学试卷	129
5.1.2 数学试卷编制的一般程序	134
5.2 数学试题组拼合成试卷	140
5.2.1 确定数学试卷结构	141
5.2.2 数学试卷的题量(试卷长度)的确定	142
5.2.3 数学试卷要组合搭配各种题型	143
5.2.4 数学试卷的难度分布	144
5.2.5 数学试卷的参考答案和评分标准的编制	148
5.3 数学试题库的建设	148
5.3.1 数学试题库的含义	149
5.3.2 建立数学试题库的工作内容	149
5.3.3 数学试题库的功能	152
5.3.4 建设基于项目反应理论的数学试题库的步骤	152
5.4 数学考试的实施方式	154
5.4.1 纸笔方式的数学考试	154

5.4.2	计算机在数学考试中的应用	155
5.4.3	计算机自适应性数学考试	157

第6章 数学考试分数的意义 /164

6.1	数学考试的原始分数	164
6.1.1	数学选择题的计分方法	164
6.1.2	数学填空题的计分方法	169
6.1.3	数学解答题的计分方法	169
6.1.4	开放性数学试题的计分方法	177
6.2	数学考试分数的合成与转换	182
6.2.1	数学考试原始分数的获得	182
6.2.2	数学考试原始分数的转换	184
6.3	不同数学试卷的等值问题	185
6.3.1	不同数学试卷得分转换的意义	186
6.3.2	不同数学考试等值的问题	186
6.3.3	不同数学试卷等值的条件	187
6.3.4	怎样编制等值的数学试卷	187
6.4	数学考试分数的价值	189
6.4.1	数学考试分数的等级制	190
6.4.2	数学考试分数线的设定方法	191
6.4.3	数学考试的百分制分数	192
6.4.4	数学考试满分设置问题	193
6.4.5	数学考试分数的置信区间	194
6.5	数学考试成绩报告	195
6.5.1	数学考试分数与等级转化	195
6.5.2	PISA 中的数学能力分数	196
6.5.3	向考生报告的数学成绩	196

第7章 对学生有重要影响的数学考试 /199

7.1	中考数学简介	199
7.1.1	中考数学的性质	199
7.1.2	中考数学命题依据	200
7.1.3	中考数学试卷结构	200
7.2	高考数学简介	204
7.2.1	高考数学的性质	204
7.2.2	高考数学命题依据	205
7.2.3	高考数学试卷结构	207

7.3 数学竞赛简介	216
7.3.1 数学竞赛的性质	216
7.3.2 数学竞赛的命题原则与方法	217
7.3.3 国际数学奥林匹克竞赛简述	225
7.3.4 我国数学竞赛的概况	227
7.4 数学考试的发展	235
7.4.1 扭转数学考试对数学教学的过度影响	235
7.4.2 加大对数学考试的反馈研究	237
7.4.3 数学考试要关注国家考试制度的政策变化	237
7.4.4 提高数学考试的科学化水平	238

现代社会里，人们都很熟悉数学考试。一个学生，从小学到大学，要经历很多次的数学考试。数学考试是不是自古就有的？它是人们追求事业发展与参加社会工作不可缺少的吗？数学考试的现状是怎样的？

第1章 绪论

现代社会中，由于资源、岗位等稀缺，使得竞争成为大多数人生活的重要组成部分。因此，如何分配进一步学习，以及就业的机会，成为社会稳定发展的关键因素。为了实现公平竞争，考试成为一种手段。虽然学校教育也在不断变革，但现实仍旧是“考考考，老师的法宝”，“分分分，学生的命根”。数学由于是主要基础课程，因此从小学开始，各种各样的数学考试就进入学生的学习，甚至是伴随着学生一起成长。

1.1 数学考试简史

正规、系统的考试体系起源于中国。但在西汉以前，“考”与“试”是两个词，都是用来表明考查的功绩，没有将二者合并，使用“考试”一词。我国古籍中较早出现的“考试”名称，是汉代董仲舒的《春秋繁露·考功名》，其中有“考试之法，合其爵禄，并其秩，积其日，陈其实，计功量罪，以多除少，以为名定实，先内第之”^[1]。可见，“考试”的含义仍旧是考察政绩之用的。下面先以中国的科举制度为界，简述一下数学考试的演变。

1.1.1 科举制度(606年以前)之前的数学考试

早期数学知识的传承是一种父对子或师带徒的个体教育。《礼记·学记》中记载：“比年入学，中年考校，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。”这事实上就是一套系统的对学生教学和考核的要求。

秦汉时期，数学教育开始从生活、实践中分离出来，出现了数学专著，但还没有数学考试。汉代建立了古代人才评价、选拔制度的雏形——选士制度，包括乡里选士、诸侯选士、学校选士。选拔标准规定为德行和道艺，具体内容：六德、六行、六艺。其中“六德”即明于事理、热爱人和万物、通达有见识、善于决断、言行一致和刚柔相济；“六行”即孝敬父母、友爱兄弟、和睦家庭、善处外亲、信于朋友和赈济贫乏；“六艺”即礼仪、音乐、射技、驾车、书法和算术。

考试通过“对策”“射策”方式出题。“对策”是以政事、经义等设问，把问题写在简策上，让考生作答；《后汉书》解释说：“射策者，谓为难问疑义，书之于策，量其大小，署为甲乙之科，列而置之，不使彰显。有欲射者，随其所取得而释之，以知优劣。”也就是抽签笔答，这是我国首次将笔试用于考试。在我国封建社会中，人们为了打江山，长期重视

选拔军事上勇猛的战士、武将，通过笔试选拔人才是一个非常大的进步。因为笔试突出的不是“体能”而是“智能”，同时，笔试在技术上注意了统一评分标准，这使得人才选拔更具客观性。

1.1.2 科举时期(606—1905)的数学考试

到了魏晋南北朝时期，中国逐渐形成了一套比较完善的管理人才体系，这就是通过选拔，任命官员，所有管理者实行中的“九品中正制”。品级越小，官位越大，到九品时，只是个芝麻大的小官儿。电影《九品芝麻官》说的正是此意。为了更好地选拔管理人员，隋炀帝大业二年(606)初置进士科，策试诸士。教育史认为，由此中国进入到科举制度的时期。可见，科举制度的建立，从一开始就是为国家选拔和任用官吏服务的，这就是我们常言所道的“学而优则仕”的缘故。从那时起，我国大约持续实施了1300年的科举制度。

1. 科举制度概述

所谓科举制度，是指国家采取举荐和考试相结合的方式选拔人才的制度。由于这个制度，每个人只要努力学习，总会有改变命运的机会，中国社会形成了崇尚读书，勤奋好学，积极进取，“万般皆下品，唯有读书高”的社会风气。

科举考试中的科目共有进士、明经、明法、明书、明算五种，其中的明算，就是数学考试。科举考试主要有帖经、问义、策论和诗赋四种题型。

据《通典·选举三》记载：“帖经者，以所试(习)经掩其两端，中间开试一行，载纸为帖，凡帖三字，随时增损，可否不一。”这也就是我们今天在各种考试中常见的“填空题”，因此填空题是中国人的发明。

问义，是就经书提出问题，要求考生回答典籍中有关史实与大义，及上下文的连缀。考生以笔作答称为“墨义”，口答时，称为“口义”。这实际上就是我们今天很多科目考试中使用的“简答题”。

帖经、问义，这两种题型侧重考查考生的熟读经典，记忆能力。

策论，不仅要求考生熟读经史，还要善于观察思考社会问题而通晓时务，作答时要针对所问写出既有一定思想、主张、见解，又要有文采技巧的论文。这与我们今天一些考试中使用的论述题类似。策论考查的是考生理解与应用知识的能力，分析问题与解决问题的能力，考查的是人的智力方面。

诗赋，要求考生须按题目的要求和规定的格律进行创作。诗有格律体裁严格规定，赋是骈体文的诗化，二者都十分注重文辞华丽典雅，追求形式。诗赋具有文艺性，注重的是个人主观思想情感的表达，因此考查的是人的情意方面。

由此可见，科举制度目的明确，它实际上是创造了一套行之有效的考试内容与方法体系。这套完备的考试系统，与我们现代的考试制度相差无几。因此，科举制度，又被人看作是与中国的四大发明一样重要的“第五大发明”。科举制度对世界的重大意义，或者说科举制度的先进性更在于它不问出身，学子通过寒窗苦读，求取功名，这是普通百姓社会地位上升的唯一途径(比印度国民的种姓等级制进步)，虽然有时会出现“上品无寒门，下品无世族”的局面，但它至少给出身低下的有为人才提供了改变自己命运的机会，是一种社会进步的表现。

光绪年间,清朝廷于1905年颁布诏书“著即自丙午科为始,所有乡、会试一律停止,各省岁科考试亦即停止。”经历了约1300年之久的科举制度,从此正式被废除。科举制度被废除其根本原因在于,考试内容陈腐(独尊儒术),专注人伦关系等社会科学的内容,不涉及生产经验和科技进步的发展成果,缺少当时已经对世界各国都具有重大影响的自然科学内容,而不在考试制度本身。

2. 科举制度中的数学考试

隋文帝时期(589—604),国家在国子监设“算学”,这是中国最早的学习数学的专门学校。设算学博士1人,助教2人,学生8人,定期进行考试。因此独立的数学考试,实际上是出现于这个时期。

科举制度中的数学考试,即明算。考试方法是闭卷笔试为主,考试内容以算经十书:《九章算术》《海岛算经》《孙子算经》《五曹算经》《张邱建算经》《夏侯阳算经》《周髀算经》《五经算》《记遗术》《三等数》为主(内容丰富)。明算科考试要求是:明数造术,详明术理。这简明扼要的八个字的意思是:考生要给出试题的正确得数(明数),同时要有得到该数的方法(造术),这就是我们今天数学考试时要求的既有结果又要有过程;考生还要详细地说明解题时所用方法的理由(详明术理)。

到了唐朝:“明算”作为常设考试科目每年举行。“明算”考试与“算学”考试内容相似。据史书记载,“凡算学,录大义本条为问答者,明数造术,详明术理,然后为通。试《九章算术》三条,《海岛算经》《孙子算经》《五曹算经》《张邱建算经》《夏侯阳算经》《周髀算经》《五经算》各一条,十通六;《记遗术》《三等数》帖读十得九,为第。试《缀术》《辑古算经》录大义为问答者,明数造术,详明术理,无注者合数造术,不失义理,然后为通。《缀术》七条,《辑古算经》三条,十通六;《记遗术》《三等数》帖读十得九,为第。落经者,虽通六,不第。”^[2]

北宋时期,直到元丰七年(1084)才有算学考试之举。《宋史·职官志》记载“算学元丰七年,诏四选命官通算学者,许于吏部就试,其合格者,上等除博士,中、次为学谕”。志第一百一十七职官四^[3],宋本《数术记遗》后附“算学源流”中记载了一些数学考试的情况。与唐代相比,并无多大变化。

元代(1206—1368)科举考试将数学内容完全砍去,明代(1368—1644)官学不设算学,虽然科举考试更加完善,但考试内容不设“明算”。

清代康熙五十二年(1713)“设算学馆于畅春之蒙养斋”,康熙“钦定”编写的《数学精蕴》是算学馆的主要教材。《清史稿·选举志》记载,乾隆四年(1739)“算学隶国子监,称国子监算学,遵《御制数理精蕴》,分线、面、体三部,部限一年,通晓七政限二年。有季考,岁考”^[2]。

可见,明算科的考试与我们现今常见的数学考试很像。科举制度使得我国很早就有数学考试的形式上,已经具备现代考试的雏形。值得注意的是,科举制度中的数学考试,却没有像今天高考数学那样,是最重要的科目。从上面的简介中可知,明算科在中国各朝代历史发展过程中是时有时无,断断续续的。这说明了在古代,作为考试内容(科目)的数学并非不可或缺。

1.1.3 西方的数学考试

说到西方古代的数学,就有言必希腊的说法.毕达哥拉斯学校创建的,一直到中世纪都是西方受教育的人必修的“七艺”,包括:算术、几何、音乐、天文、文法、逻辑和修辞.算术、几何属于数学的内容无可置疑;逻辑与数学密不可分,很长一段时期,教育上就是用欧氏几何来训练人的逻辑思维;同时,古希腊人将“音乐视为流动的数”,“天文被看作是运动的形”,因此,音乐与天文也是数学家研究的内容.这样一来,“七艺”中就有“五艺”与数学相关.可见,古希腊对数学的重视程度之高.

1. 古希腊的数学考试

古希腊时期,从某种意义上说,受教育属于奢侈、享受.只有少数有闲的贵族阶层才能获得受教育的机会,而且数学教育实际上是一种精英教育.由于教育基本上受苏格拉底(Socrates)教学的影响很大,通过面对面口授的方式进行教学,采用问答谈话,施行“产婆术”,因此,对人的数学水平考查也是口头形式的,通过口试了解教学效果和促进学生学习.

古希腊贵族生活方式的一个核心价值观就是竞争与竞赛.这不仅仅表现在竞技场上的体育竞争与竞赛,还体现在对宇宙和自然认识上的智力竞争与竞赛.在数学史上,许多数学成果的诞生就源于数学爱好者们求解数学难题的竞争与竞赛.例如,著名的一元三次方程、四次方程求解公式的发现,就是爱好数学的人们辩论和竞赛的成果.

同时,进入中世纪以后,西方的教育主要由宗教组织负责.教会的僧侣,在学校中数学教育并不是最主要的科目,教学与学习方式大多都是死记硬背,考试采用口试方式.在很长的时间里,古代西方国家并无中国式的科举制度.

2. 中国科举考试在西方的发展

第一个详细介绍中国科举制度的西方学者,是16世纪末的意大利传教士利玛窦(Matteo Ricci, 1552—1610).所以,西方先是教会学校建立了考试制度,随后,从教会学校推广到普通学校.直到1702年,在英国剑桥大学才开始采用笔试考核学生,这应该是欧洲学校笔试之肇始^[4].

19世纪中叶,英国在剑桥和牛津考试中取得优异成绩的人,差不多日后都有较高的社会地位,是所在专业的杰出人才.1853年,斯坦利(Stanley)爵士说“本国很陌生的一个原理,据说在中国却很流行,所以应该把它叫做中国原理,即通过智力的竞争来选拔文官”^[5].虽然,当时人们对“年轻人的早慧”就决定了他未来表示怀疑,但这总比赐官要强得多.因此,竞争性考试逐渐成了一个民主但又是选择精英的工具.

法国,从1747年开始,在巴黎的公立中学设立了会考,为的是奖励竞争.1776年,开始用竞争性的学衔考试来选拔大、中学教师.法国大革命以后,考试的作用日益重要,它不仅被看作是控制大众的手段,而且也是受过教育的杰出者接近权力的途径.1795年,法国师范学院入学考试出现.1808年法国建立了学士学位考试,这个考试面向学校的全体教师和学生,也面向政府部门的工作人员.拿破仑战争后,考试体系成了使教育中央集权化的手段.

考试在德国教育的系统化和集权化过程中也起了重要作用。1810年,普鲁士引进了教师需要国家认证的资格,这提升了教师的教学水平。1812年,普鲁士开始设置中等学校的升学考试(这在1788年已被发明);到了1871年,整个德意志帝国都接受了这种制度。随后,德国通过各种考试,不仅用来控制学校,也用来选拔政府公务员。

1845年,美国初等学校普及,学生数激增,对毕业生一一口试已不可能。为提高效率,美国学校也开始逐渐采用统一笔试^[6]。

到19世纪末,在西方考试地位基本稳固了。它主要服务于三个目的:第一,在教学中,是一种“刺激物”,激励学生更好地表现;第二,作为一种“侵入性力量”,它允许外部实体(大学或政府机构)对教育体系内部的运作施加影响,于是教育体系变得越来越复杂;第三,作为一种“碰运气的办法”,虽有争议,但在选择公务员或证明职业资格方面,考试还是有长处的。

就如孙中山先生曾经说过的“考试制度在英国实行最早,美国实行考试不过是二三十年,现在各国的考试制度,差不多都是学英国的。穷流溯源,英国的考试制度原来是从我们中国学去的。所以中国的考试制度就是世界中最古最好的制度”^[5]。

1.1.4 数学考试方法

长期以来,西方对考试的批评并没有主要针对其目的,而是主要针对考试技术。人们注意到,试题缺乏稳定性,不同评分者之间缺乏一致性。因此,考试存在误差,就像对时间、距离、重量等物理量的测量中存在误差一样,考试只是一个粗略的测量。1888年,埃奇沃斯(Edgeworth)在一篇讨论考试误差的文章中说,任何测量都以一种规则的方式表现出对精确值的偏离,因此小的离差就比大的离差更为普遍。把这种现象画成图,就呈现出正态曲线。这种“不可避免的不确定性”使人们担忧,将考试作为一种测量工具为许多人事决策提供依据时,它的科学性能否得到保证。为矫正弊端,力求考试的客观化,西方人开始系统研究考试的科学方法。

1. 口试

口试是通过考生口头回答问题,对其数学学习进行测试的一种方式。口试时,考官与考生是面对面地进行的,因此常可以清晰地观察考生回答问题时的情绪变化;同时,口试也可以不受事先规定的考试题目的限制,就考生的作答情况,深入追问下去,以便获取考生更多、更深层次的心理倾向。口试在考查考生的学习态度、学习动机等非智力因素方面,更具有优势。

过去很长时间,由于没有录音、录像设备,口试无法保留考生作答的情况,以便作为日后评价的依据。现在,由于信息技术手段的发达,这个障碍已经不存在了。我们可以借助录音、录像设备,记录下考生作答的行为表现,作为比较、评价的依据。然而,即使到今天我们对口试进行标准化、量化也很困难。

2. 笔试

笔试是通过考生进行书写回答事先设计好的一系列问题,对其数学学习进行测试的方式。笔试有诸多优点,如节约时间,内容覆盖面更宽;作答结果留有书面资料,评分更公

平精确，容易免除教师的偏袒和成见；结果还可公之于众，试题难易可由公众人士评审等。

笔试虽然客观性、可靠性比口试高，节约时间，测试结果优于口试，但评分也会受到主观偏见影响，且试题太少，不足以反映考生所获数学知识与能力的全貌。

事实上，我国的科举考试是首先传入我们的近邻，比如日本、韩国，随后才进入西方的。由于世界大战的影响，我国对科举制度的废除，使得我们很长一段时期，对考试的使用与研究进展不大，倒是传到欧美的科举制度，使得西方开始了对考试制度、方法体系的系统研究，并形成了一整套现代教育测量与评价的理论与方法。

1.2 智力测验与数学考试

现代教育科学中，数学考试是归属于教育测量与评价这个学科的研究范畴。中国古代教育测量与评价的思想中包含典型的东方文化特点：即定性描述及带有道德判断色彩。例如，我国古代“七巧板”是很常见的一种儿童玩具。现实生活中，人们实际上把它看作测量儿童智力的工具。在西方也有数学是智力游戏的说法。为此，我们阐述一下数学考试与智力测验的关系。

由于自然科学的成就，人们发明了各种工具来测量物理现象，如时间、长度、重量等。我们能否测量像智力这样的属于人脑的机能呢？德国心理学家艾宾浩斯·赫尔曼(Ebbinghaus, Hermann 1850—1909)对记忆这种心理现象的研究，为人们进行智力测量的研究奠定了基础。

1.2.1 对人的记忆的测量

1885年，艾宾浩斯为纯粹研究记忆，避免联想的干扰，他发明了无意义音节。利用这种材料，艾宾浩斯推想，学习材料的难度，可以用学习材料时所需要重复的次数来测量它，而计算重复的次数可以作为完全再现的标准。他出版的《论记忆》一书中，介绍了对人的记忆进行测量的结果，如下：

(1)无意义材料的学习与有意义材料的学习的差异是，识记无意义音节的速度是有意义材料的9倍。

(2)完全再现较长的材料需要次数较多的重复，也即需要较多的学习时间，而增加所学音节的数量，学习每一音节的平均时间也显著增加。结论是，学习的材料数量愈多，所用的时间就愈长。

(3)学习和回忆之间的时间历程是，材料在学习后头几小时遗忘最快，随着时间的推延，材料遗忘越来越少，如图1-1所示的艾宾浩斯遗忘曲线。

除此之外，艾宾浩斯还研究了一些能影响记忆

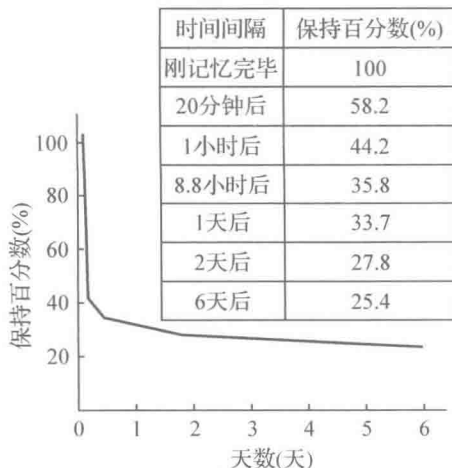


图 1-1 艾宾浩斯遗忘曲线