雷新勇/著

大规模教育考试：命题与评价

Large-Scale Educational Test Development and Evaluation

华东师范大学出版社

雷新勇/著

大规模教育考试：命题与评价

Large-Scale Educational Test Development and Evaluation

华东师范大学出版社

图书在版编目(CIP)数据

大规模教育考试：命题与评价/雷新勇著. —上海：华东师范大学出版社，2006.3
ISBN 7-5617-4617-2

Ⅰ.大… Ⅱ.雷… Ⅲ.①中学—考试—命题—研究②中学—考试—试卷—评价 Ⅳ.G632.474

中国版本图书馆 CIP 数据核字(2006)第 017480 号

大规模教育考试：命题与评价

撰　　著　雷新勇
项目编辑　张继红
文字编辑　林　敏
责任校对　王丽平
封面设计　黄惠敏
版式设计　高　克

出版发行　华东师范大学出版社
社　　址　上海市中山北路 3663 号　邮编 200062
电　　话　021－62450163 转各部　行政传真 021－62572105
网　　址　www.ecnupress.com.cn　www.hdsdbook.com.cn
市 场 部　传真 021－62860410　021－62602316
邮购零售　电话 021－62869887　021－54340188

印 刷 者　江苏宜兴市德胜印刷有限公司
开　　本　787×1092　16 开
印　　张　21.25
字　　数　346 千字
版　　次　2006 年 4 月第一版
印　　次　2006 年 4 月第一次
印　　数　4100
书　　号　ISBN 7-5617-4617-2/G·2694
定　　价　36.00 元

出 版 人　朱杰人

作 者 简 介：

雷新勇 男 博士 现为上海市教育考试院命题与研究办公室主任，主要从事大规模教育考试命题研究和管理、考试研究和评价。

近年来，在我国教育考试和研究理论刊物上发表的主要论文有：

◆ RASCH评分量表模型及其在作文评分中的应用. 考试研究与信息，2003，（2）

◆论常模参照考试的质量分析. 中国考试，2003，（2）

◆ 用多元概化理论研究综合能力测试（上海卷）改革的必要性.中国考试，2005，（1）

◆ 高考英语（上海卷）多元概化理论研究. 考试研究，2003，（2）

◆ 上海市高考“3+1”科目组测量误差研究.考试研究，2004，（2）

◆ 2004高考（上海卷）地理科考试评价. 考试研究，2005，（1）

内 容 介 绍

本书以定性和定量方式，系统地回答了大规模教育考试命题管理人员、广大命题教师关切的一系列命题和评价的基本问题，介绍了相关的技术和方法。考试机构命题管理、考试研究和评价人员，参与大规模教育考试命题的教师，在高等学校学习教育测量和教育考试的研究生，教育招生考试机构领导，一般教育问题研究者都可以在书中发现自己需要的内容。

本书阐述了大规模教育考试与学校教育教学的关系、现代教育和心理测量的一些概念和理论以及大规模教育考试的局限性；阐述了大规模教育考试设计、开发的基本问题；运用大量的试题实例，讨论了客观题和主观题的主要测量功能，命制客观题和主观题应该注意的主要问题，主观题评分标准问题等，以及在网络阅卷的背景下，主观题评分的误差控制理论和方法；介绍了考试评价的理论和技术，包括试题分析，考试结果的信度和误差分析，效度和效度检验；阐述了大规模教育考试影响学校教学的机制，以及研究这一问题的方法和改进考试对学校教学影响的策略。

书中包含了作者近年发表和未发表的对上海教育考试的大量研究成果和实例，作者企图以此将我国，尤其是上海大规模教育考试命题和评价的实践经验，融合到现代教育考试理论之中。

谨以此书献给

上海市教育考试院成立十周年

上海市高校招生考试单独命题二十周年

序一

异常兴奋地阅读了雷新勇博士的近作《大规模教育考试：命题与评价》，尤其此书完稿在上海高考单独命题20周年和上海市教育考试院建院10周年的有纪念意义时刻，更激起我思绪联翩。

20年前，为了促进上海地区中学教育的整体改革和教育质量的提高，经国家教育部授权，上海试点高考单独命题，并伴以相应的考试制度改革，即率先建立高中会考制度，在此基础上减少高考科目，实现“3＋1”六个科目组设置方案。10年前，又为了推进政事分开的政府机构改革，以便更科学、规范、高效地实施各类教育招生考试，市委与市府又组建了上海市教育考试院。这两项涉及招生考试领域的改革举措，得到全国的回响，现在“分省命题”已在16个省（市）铺开，建有教育考试院机构的省市也逐渐增多。但就高校招生考试制度改革的目标而言，我们还远未达到。中国是考试的故乡，每年的高考牵动着千百万家庭的心，这成为当今中国社会生活的独特现象。当我们在构建社会主义和谐社会时，不得不审视一下在和谐社会下应有的高考，其中有一个如何科学地看待考试的问题。

现实社会中，我们不时听到对“高考指挥棒”诛伐的言辞，但人们从生活中直感，又不能没有考试，那么不受诛伐的考试该是什么模样的呢？人们不时谴责现实教学中“考什么，教什么，学什么”违反教学活动规律的怪圈，但换以“教什么，考什么”就科学了吗？有考试，就有命题，一般人认为，所谓命题就是找一些专家，拿出一些题目，组合一下就成了试卷，但并不清楚作为面向公众的大规模教育考试，其命题有着特殊的规范，有科学的要求，惟此才能成为公平、公正的评价工具。诸如上述疑问，使我意识到只有实事求是地认识考试，对待考试，普及考试科学的基本点，我们才能避免浮躁，接近招生考试制度改革的目标，找到现实的可持续发展的改革方案。考试的客观评价功能不能抹杀，但考试功能是有限的，不能无限拓展、加

载、拔高，终而毁了考试。总之，还是要以科学发展观统领考试，提倡科学考试，学一点考试科学。

很欣慰，雷新勇博士的近作，给我们提供了一本科学认识考试与了解什么是考试科学的好教材。雷博士在自然科学、教育与外语方面有着扎实的基础，先后在大学教学、中学校长、考试机构命题实践与命题管理的岗位上勤奋工作，因而本书的产生有着厚实的土壤。该书所述的问题是上海高考单独命题20年来工作中所感受的实际问题的总结，其中有经验，也有教训。该书又以简明而不失系统的篇幅介绍了国际上关于命题和评价方面的理论与方法上的成果，特别是上世纪末以来最新的国际研究成果，同时也赋以上海单独命题实际的体会和例子，使我们读来不觉得乏味。

神州大地上，高考"分省命题"的试点实践正在展开，我认为本书的出版，对命题教师，对考试机构的命题专业人员，以及一切对考试科学关心的教育工作者、行政管理者，无疑是个喜讯。作为一个在上世纪九十年代由于工作的需要，闯入招生考试领域的教育工作者，我愿为此书的出版鼓与呼！

胡启迪

（原上海市教育考试院院长，教授）

于2006年元月

序二

2005年是废除科举制度100周年，同时也是科举制度诞生1400周年。我国是考试的发源地，同时又是一个考试大国，仅高等学校入学考试目前全国一年的考生数即将突破1000万。在我们对考试的发明而沾沾自喜的同时，又不得不为我国考试理论发展的滞后而汗颜，在我们进行大规模考试实践的同时，也不得不为缺乏系统而实用的理论指导而困惑。

今年是上海市高考自主命题20周年，也是上海市教育考试院成立10周年纪念。我们从事了多年的考试实践，但却缺乏从理论上系统的总结和分析，因而历年命题、命题管理和分析仍然较多地停留在经验上。

在今年这一值得纪念的年份里，我们欣喜地看到雷新勇博士《大规模教育考试：命题与评价》一书的完稿。它不仅是一部上海市高考自主命题20周年和上海市教育考试院成立10周年纪念的献礼之作，更是一部补缺之作，因为本书不仅是讨论大规模教育考试命题和评价问题的一本实用性理论著作，也是国内第一本关于大规模教育考试命题和评价理论及技术的著作。

同时要提及的是本书作者雷新勇博士不但具有扎实的数理统计功底和教育测量理论基础，而且具有较长时期的高考和中考命题管理的实践经验，作者的背景无疑将凝聚而形成本书的一大特色——理论紧密联系实际。

全书分为四部分：第一部分主要讨论大规模教育考试与学校教育教学的关系，以及如何研究大规模教育考试对学校教育教学的影响。在这一部分中作者明确指出课程标准与考试大纲的关系、教育考试命题与教材的关系、教育考试的形式和方法与学校教学活动的关系、教育考试的内容与学校教学内容的关系，介绍了考试影响学校教学的机制，并以实例介绍了研究这一问题的方法，以及改进考试对学校教学影响的策略。第二部分介绍

现代教育和心理测量的一些概念和理论，包括测量、考试和评价的概念；基本的测量学质量指标；测量量表的概念以及教育考试的局限性；重点介绍了经典的真分数理论中的均质考试和单因素模型，及均质考试的信度问题；介绍了多元概化理论以及试题反应理论。第三部分讨论大规模教育考试设计、开发的基本问题，包括考试的测量目标问题、考试的内容规范和试题规范的设计、试卷的结构、试题背景材料选择、大规模常模参照考试和标准参照考试的难度等问题。同时，运用大量的试题实例，介绍了客观题和主观题的主要测量功能、命制客观题和主观题应该注意的主要问题、主观题评分标准问题等，以及在网络阅卷的背景下，主观题评分的误差控制理论和方法。第四部分主要介绍考试评价的理论和技术，包括试题分析、考试结果的信度和误差分析、效度和效度检验以及大规模教育考试对学校教学的影响。试题分析中不但讨论了经典真分数理论框架下的试题分析，也讨论了在试题反应理论框架下的试题分析的理论和方法。信度和误差分析中，讨论了在经典的真分数理论、多元概化理论、试题反应理论框架下信度和测量误差估计的理论和方法，并且侧重分析误差来源及不同来源误差对考试信度、效度的影响。效度和效度检验中，讨论了大规模教育考试分数解释和适用的有效性证据，阐述了收集内容方面证据、考试内部结构方面证据的理论和方法，阐述了与标准相关的证据的问题，尤其值得提出的是作者从考试分数使用和考试评价的角度，讨论了考试的单维性问题。

本书第一部分为作者对大规模教育考试与学校教育教学的基本认识，第三和第四部分包含了作者近年发表和未发表的对上海教育考试的大量研究成果和实例，作者企图以此将我国，尤其是上海大规模教育考试命题和评价的实践经验，融合到现代教育考试理论之中。

2004 年教育部推进高考分省命题工作，当年，继上海、北京自主进行高考命题后，又有九个省市实行自主命题，2005 年自主命题省市扩大到十三个，2006 年将扩大到十五个。与此同时，各省市招生考试机构的考试命题管理人员、广大的命题教师，越来越感到大规模教育考试理论和实践经验的缺乏，迫切需要了解大规模教育考试命题和评价的理论和技术。本书回答了大规模教育考试命题管理人员、广大的命题教师关切的一系列命题和评价的基本问题，介绍了相关的技术和方法。内容丰富，资料翔实，观点明确，可读性强。

本书不但适合于考试机构命题管理、考试研究和评价人员，也适合于

参与大规模教育考试命题的教师，也可以作为高等学校学习教育测量和教育考试的研究生之教材，还可以作为教育招生考试机构领导、一般教育问题研究者和教育行政机构高级公务员的案头参考书。

李瑞阳

（上海市教育考试院院长，教授）

于2005年岁末

前　言

大规模教育考试的目的是为对考生进行教育决策提供决策依据，如普通高等学校招生考试是为高校选择新生提供依据；高中阶段学校招生考试是为高中选择新生提供依据；初中毕业生学业考试是检查学生是否达到了课程标准规定的学习目标，决定是否准予学生初中毕业；高等教育自学考试检查考生是否达到学科课程的要求，决定是否能够授予考生学科合格证书，并最终决定是否准予考生毕业，承认其学历。显然，大规模教育考试都是高利害考试，它与考生及考生家庭的利益，甚至学校的利益密切相关。

大规模教育考试的目的和性质决定了考试结果误差必须尽可能少，具有高度的可靠性——信度；我们可以将考试结果解释为考生学科素养的标志，以考试分数为依据对考生进行决策，因此要求考试结果具有高度的有效性——效度。信度和效度是大规模教育考试最重要的质量标准。提高大规模教育考试的信度和效度是教育考试机构的工作者以及命题教师、考试研究和评价人员的重要职责。

大规模教育考试的信度和效度取决于很多因素，如考试的测量目标是否准确、测量目标的行为目标及其表现水平标准是否明确；试卷的长度及试卷的结构；考试试题的类型、试题的难度、试题撰写的质量等。考试的设计者、开发者、命题教师以及考试研究和评价人员应该知道并理解这些因素如何影响考试的信度和效度，在考试设计、开发的各个阶段应该采取何种措施，最大限度地克服影响考试信度和效度的不利因素；应该知道考试以后如何通过对试题和考试结果的定性、定量分析，对试题和考试结果作出恰如其分的评价，确定考试结果的可靠性，以及考试结果解释和使用的有效性；应该发现试题和考试结果中存在的问题，找出解决这些问题的对策和方法。

大规模教育考试对中学的教育教学有着巨大的影响，它不但影响学校

的教学内容、教学和训练方法，还影响到学校的教育教学目标，甚至影响到学校校长、教师、学生的价值判断。这种影响既有正面的，也有负面的。考试的设计者、开发者、命题教师以及考试研究和评价人员应该研究考试对学校教育教学的影响，尽可能降低考试对学校教育教学的负面影响。

作者怀着上述目的撰写此书，希望与国内同仁分享上海市教育考试院命题工作者近二十年在命题工作方面的认识、经验和教训。

本书共十八章，大致可分为四个部分。第一部分主要讨论大规模教育考试与学校教育教学的关系，以及如何研究大规模教育考试对学校教育教学的影响，包括第一章和第十八章。第二部分考虑到国内考试机构考试工作者队伍的现状，扼要地介绍一些基本的教育测量概念和理论，包括第二章到第五章。第三部分主要介绍考试设计、开发的基本问题，客观题和主观题的功能，命制客观题、主观题时应该注意的基本问题，以及主观题评分误差的控制，从第六章到第十四章。第四部分主要介绍考试评价的理论和技术，包括试题分析，考试结果的信度和误差分析以及效度和效度检验。从第十五章到第十七章，书中许多实例均为作者对历年上海高考、中考试题及考试结果的研究成果。本书没有包含教育考试中试题功能偏差方面的内容，原因有二：一是这个问题到目前为止并没有引起教育考试工作者、命题教师的普遍重视，没有引起考试机构的普遍重视，更没有引起教育行政部门的重视，其实这个问题在我国的大规模教育考试中是普遍存在的；二是由于众所周知的原因，目前对这个问题的研究，还仅限于内部，成果尚不能公开。

本书不但适合于考试机构的工作人员，尤其是命题管理、考试研究和评价人员，也适合于参与大规模教育考试命题的教师，还可以作为高等学校教育测量和教育考试方向研究生之参考材料。

由于时间仓促及作者的能力所限，书中难免有文字上的疏漏和理论认识上的不全面，敬请广大同行不吝赐教。

本书撰写过程中，得到上海市教育考试院院长李瑞阳教授、前院长胡启迪教授的关心和大力支持，得到上海市教育考试院其他领导的大力支持。在此谨致由衷的谢意。

雷新勇

2005年于上海

目　录

第四章　多元概化理论基础

第五章　试题反应理论模型

第六章　大规模教育考试开发的基本问题

第七章　教育考试的测量目标

第八章　考试内容规范和试题规范的设计

第九章　教育考试试卷结构

第十章　试题背景材料的选择

第十一章　考试的难度

第十二章　客观性试题(选择题)的编撰

第十三章　主观题的编撰

第十四章　主观题评分误差的控制

第十五章　试题分析

第十六章　考试的信度和测量误差