

谢小庆 罗 莲 / 主编

考试研究文集

（第5辑）



经济科学出版社

教育部《学位与研究生教育》
2003年第1期

考试研究文集

(第4辑)

教育部学位与研究生教育发展中心
2003年10月

考试研究文集

(第5辑)

谢小庆 罗 莲 主编

经济科学出版社

图书在版编目 (CIP) 数据

考试研究文集 (第5辑)/谢小庆, 罗莲主编. —北京:
经济科学出版社, 2009. 11

ISBN 978 - 7 - 5058 - 8653 - 7

I. 考… II. ①谢…②罗… III. 考试学 - 文集
IV. G424. 74 - 53

中国版本图书馆 CIP 数据核字 (2009) 第 181064 号

责任编辑: 卢元孝

责任校对: 杨海韩宇

版式设计: 代小卫

技术编辑: 潘泽新

考试研究文集 (第5辑)

谢小庆 罗莲 主编

经济科学出版社出版、发行 新华书店经销

社址: 北京市海淀区阜成路甲 28 号 邮编: 100142

总编辑部电话: 88191217 发行部电话: 88191540

网址: [www. esp. com. cn](http://www.esp.com.cn)

电子邮件: [esp@ esp. com. cn](mailto:esp@esp.com.cn)

北京密兴印刷厂印装

880 × 1230 32 开 14. 25 印张 380000 字

2009 年 11 月第 1 版 2009 年 11 月第 1 次印刷

ISBN 978 - 7 - 5058 - 8653 - 7 定价: 33. 00 元

(图书出现印装问题, 本社负责调换)

(版权所有 翻印必究)

前言

教学与评价是教育赖以发展的两个车轮，相辅相成，互为因果，缺一不可。好的教学是基于评价之上的、有的放矢的、因材施教的教学；好的评价是可以为教学提供反馈信息和改进信息、可以促进教学的评价。评价的科学化，直接影响到教育的发展。

从6世纪科举制度形成以来，评价的意义就不仅仅体现在人才选拔上，而且体现在对于教育的导向（wash back）功能，评价的方式直接影响到教育的发展方向，发挥着“指挥棒”的作用。近代以来，人们进一步清楚地认识到，评价的结果是教育问责（accountability）的依据。近几年，人们更加清楚地认识到，好的评价不仅仅是“关于学习的评价（assessment of learning）”，更是“为了学习的评价（assessment for learning）”。因此，关于评价问题的思考，就需要包含哲学、政策和技术三个层面。

杜威在《民主主义与教育》一书中说：“哲学乃是教育的最一般方法的理论”，而“教育乃是使哲学上的分歧具体化并受到检验的实验室”。^①教育的对象是人，对人的

① 杜威：《民主主义与教育》，人民教育出版社2001年版。

本质的认识直接关系到我们怎样理解教育的目的和任务，怎样理解教育对于一个自然人以及对于社会所要担当的责任。教师的责任是什么？我们为什么办学？我们需要培养什么样的人才？我们需要思考这些教育的核心问题。在哲学层面上，我们同时还要思考评价考试质量的标准（criterion）是什么？科学的考试赖以立足的逻辑基础是什么？脱离了对这些基本哲学问题的思考，考试研究就迷失了方向，考试研究就成为无本之木。

在政策层面，我们需要思考考试的形式、内容和题型，以及考试分数的报告方式，等等。在考试领域，很少见“有百利而无一害”的政策选择，我们需要考虑每一项政策抉择可能产生的社会效应，需要对这项政策抉择的得失利弊进行权衡。

在技术层面，我们需要尽可能地利用现代科学技术已经提供给我们的技术手段，最大限度地降低考试的负面效应，不断使一些“不可能”的事情成为“可能”。恰如毛泽东所说：“我们不但要提出任务，而且要解决完成任务的方法问题。我们的任务是过河，但是没有桥或没有船就不能过。不解决桥或船的问题，过河就是一句空话。不解决方法问题，任务也只是瞎说一顿。”^①

遗憾的是，考试问题，尤其是高考改革问题，今天已经成为一个“公共话题”，随便什么人都可以对考试问题发一通议论，都可以仅仅凭借“眉头一皱、计上心来”而谈论“考试改革”。随便什么人，都可以在不做任何文献调研和不做任何实证研究的情况下，对考试提出一系列的

① 《毛泽东选集》第1卷，人民出版社1991年版，第139页。

批评和提出一系列的建議。在考試領域，到處可見毛澤東同志所批評的“空話”和“瞎說一頓”。這些“空話”，大多停留在考試的政策層面，鮮有人進行哲學層面的思考，更鮮有人能夠在技術層面提出切實可行的實施方案。

本輯收集了我們近兩年來的一些研究成果，反映出我們在哲學、政策 and 技術三個層面對考試問題的思考和探索。在哲學層面，我們追問“為什麼要進行高考改革”，我們思考科學考試的效度依據；在政策層面，我們討論了筆試與面試、選拔考試與資格考試、多元化選拔、檔案式評估、描述性記分、教師評價方式等問題；在技術層面，我們討論了測驗等值技術、核等值、IRT 和 CTT 的題目分析，IRT 中的數據與模型擬合、作弊甄別技術、平滑技術等問題。這些研究涉及的考試包括漢語水平考試（HSK）、中國少數民族漢語水平等級考試（MHK）、職業漢語能力測試（ZHC）、漢字應用水平測試（HZC）、跨文化交際能力測試（ICCT）、中學生語文學習能力考試（YNC）、國際漢語教師能力考試、公務員考試等。

甲型 H1N1 病毒對每一個人都構成威脅，但很少有醫學界以外的人對甲型 H1N1 的流行和防治議論紛紛，喋喋不休。考試問題成為“公共話題”，折射出考試科學研究的尷尬，折射出考試研究尚缺乏物理學、生物學、醫學研究那樣的专业地位。我們希望與考試研究領域的同行們共同努力，逐漸形成考試研究的科學範式（paradigm），逐漸建立起考試科學研究的专业地位。

編者

2009 年 8 月 24 日

◎ 概 论 ◎

- 谢小庆：为什么要进行高考改革 1
- 谢小庆：再谈为什么要进行高考改革 5
- 谢小庆：高校招生实现多元评价论析 9
- 谢小庆：大学招生中应该给高中教师发言权 21
- 谢小庆：从科学测试写作能力的角度看高考作文题 30
- 谢小庆：《民族汉考（三级）优秀作文评析》序言 32
- 谢小庆：《职业汉语教程》绪论 36
- 谢小庆：职业汉语不同于文学汉语 48
- 杨 洋：标准化考试不能承受之重 53
- 杨 洋：关于进行跨文化交际能力测试的构想 61
- 张 洁：高校招生制度改革中的教师评价 72
- 刘雪莹：初中生的语文能力测试需要改革 79

◎ 文献综述 ◎

- 谢小庆：考试分数等值的新框架 93

- 陆 敏: 甄别答案抄袭的 K 指数方法 110
罗 莲: 一种新的等值方法
——核等值法 125
景利波: 从构念模型到论证模型
——《教育测量》(第四版) 的新效度观 140
景利波: 基于效度证据的分数解释 154
李靖华: 一种新的评价形式
——档案式评估 171
孔 娜: 对 IRT 模型与数据拟合度的研究综述 182
黄海峰: 浅谈语言测试中的语言交际能力和表现 190
齐 桀: 描述性记分介绍及其在语言测试中的应用 201
罗 莲: 第二语言测试研究的新趋势 216
杨 洋: 美国教育面临的问题及对中国的启示 225

◎ 实证研究 ◎

- 张晋军
邱 宁
张 洁: 汉语水平考试 (HSK) 与《国际汉语能力标准》挂钩
研究报告 233
张晋军
李桂梅: 汉语作为第二语言测试的听力理解题型设计新思考 249
景利波: 等百分位等值中平滑法的应用及等值误差估计
——对 HSK (初中等) 等值的一项实证研究 260
陈俊丽: 核等值法与其他等值方法的比较研究 290
李靖华: 论汉字应用水平测试的试卷结构 307
杜红玉: 合成分数与 HSK 信度检验 316
马洪超: 对 IRT 和 CTT 题目及能力分析差别的考查 331

何卫革：	HSK（高等）题型效度研究 342
------	-------------------

◎ 考试探讨 ◎

张晋军：	关于商务汉语考试分级问题的分析 357
许义强：	试论汉语作为第二语言测试命题中的敏感性审查 364
陈作宏	
邓秀均：	外国留学生汉语进修班分班测试初探 374
李勉东	
孙洪威：	MHK 作文中的偏误简析 389

◎ 其 他 ◎

谢小庆：	在欢迎 2008 年新研究生时的发言 399
谢小庆：	应试辅导教师们，请爱惜自己的学术生命 413
谢小庆：	提请教育官员注意两条报道 415
谢小庆：	致敬爱的幼儿教师们 417
谢小庆：	大学校长的“两难”背后 419
谢小庆：	教育改革与外汇保值 422
谢小庆：	需要开辟第二教育渠道 425
谢小庆：	公务员笔试是不得已的办法 428
谢小庆：	将公务员笔试改为“资格考试”的困难在哪儿 430
谢小庆：	公务员考试为什么没有统一的外语要求 432
罗 莲：	谈谈考试的人性化 435

概 论

为什么要进行高考改革^{*}

谢小庆

30年前，高考的恢复，受到了社会各界人士由衷的欢呼。我本人也于1978年通过高考进入大学，开始了自己的学术生涯。30年后的今天，高考却受到越来越多的批评，改革高考的声音不绝于耳，甚至“取消高考的呼声此起彼伏”。有人说“高考已经到了彻底变革的时候了”，有人说高考是“地狱”，有人说高考是“悬在中小学头上的剑”，有人说高考伤害了中小学生的健康，有人说高考“偷”走了中国人的创造力，甚至有人主张将高考送进坟墓。^①

今天，高考牵动着成千上万家庭，受到社会各界的关注，许多人都在对高考改革发出议论。我发现，许多谈论高考改革的人，并没有认真思考过“我们为什么要进行高考改革”的问题。

^{*} 原刊于《中国教师》2007年第4期。

^① 以上关于高考的批评均引自舒云的长篇报告文学《高考殇》，载《北京文学》2005年第10期。

世界闻名的科举考试在中国延续了1300年。科举考试是一种形式上非常公平的考试，许多保证考试公平性的措施今天都很难做到。例如，科举的“单间”施测，进场搜身，誊录（为了避免阅卷人根据笔迹作弊将所有答卷全部重新抄录），等等。17世纪传教士将中国的“考试取人”方法介绍到西方，得到西方进步思想家的高度赞扬。今天在世界许多国家实行的公务员录用考试，大都源于中国的科举。

但是，形式非常公平的科举却遭到近代几乎所有进步思想家的切齿痛恨，严词挞伐。例如，明末的进步思想家顾炎武痛斥“八股之害，甚于焚书，而败坏人才，有甚于咸阳之郊之所坑者”。康有为认为“中国之割地败兵，非他为之，而八股致之也”。梁启超讲“救弊之法归之于废科举”。谭嗣同讲“变科举诚为旋乾转坤转移风会之大权而根本之尤要者也”。在朝野的一片批评声中，清廷终于在1906年取消了科举考试。

科举考试为什么遭到进步人士猛烈攻击？科举为什么最终被“改革”掉了？并不是由于科举不够公平。科举被取消的原因是由于它既没有为中国带来政府人员素质的提高，也没有带来国富民强和国家尊严。科举失败的原因在于考试脱离实际，引导人死读书，读死书。科举不是以“德才”取人，而是以“背书”取人。清廷是迫于“文无经世致用之臣、武无统兵御敌之将”的局面，迫于“圣阙涂炭、祖陵震动”压力，才取消了科举。

从20世纪80年代初开始，我本人已经专门从事考试研究27年。作为一个职业的考试研究者，我清楚地知道，考试是追求效率的工具，考试所追求的并不是公平。多数情况下，考试都会造成社会的不平等。今天，国家政策明确规定不允许在小学和初中入学中采用考试。为什么？就是因为采用了考试就损害了每个儿童与生俱来的平等的受教育权利，就是因为采用了考试就损害了每个儿童都应平等分享优质教育资源的权利。尽管我本人是靠考试吃饭的，这种禁止考试的政策一定程度上会影响到我的“饭碗”，但我一直是

旗帜鲜明、毫不含糊地坚决支持教育部禁止小学、初中入学考试的政策。为什么禁止考试，就是由于考试导致不平等。

我反对在小学、初中采用入学考试，却支持在高中、大学和研究生招生中采用考试。支持采用考试，主要是出于提高教育效率的考虑。借助考试，可以更好地实现“因材施教”，可以提高高中、大学和研究生教育的效率。

今天，招生制度改革的一项重要任务是降低高考在大学招生中的比重。从2006年起，复旦大学、上海交通大学在大学招生中开始尝试的面试方式，就是降低招生中考试比重的一项措施。反对在大学招生中降低考试比重的人说，这样做会损害大学招生的公平性。这些人需要认真思考，我们究竟为什么要进行高考改革？

我们今天进行高考改革，主要是为了改变青少年的普遍厌学状况，主要是为了保护青少年的好奇心和青少年的创新能力。2005年5月在巴黎进行的中欧贸易谈判中，商务部薄熙来部长曾说：“为了买回一架空中客车A380，中国需要出口8亿件衬衣”。这句话，突出地体现了中国经济发展中缺乏自主创新的问题。因此，在中共中央十六届五中全会上，“自主创新”成为核心话题之一。在《全会公报》中提到要“着力自主创新”，提到“必须提高自主创新能力”，提到“把增强自主创新能力作为调整产业结构、转变增长方式的中心环节”。

今天，中国对人类在科技和文化艺术方面的贡献，与得到世界公认的中国人的“聪明与勤劳”并不相称。虽然导致这种局面的原因很多，但不能说与“学以应考、教以应考”的“应试教育”没有关系。这种“应试教育”，从童年起就挫伤了中国儿童的好奇心和创造力，影响到从基础教育到高等教育的整个国民教育体系，影响到我国的国民素质，摧残着中华民族的自主创新能力。这种“应试教育”，造成普遍的“读一本书、教一本书、背一本书、考一本书”的局面。我们改革高考，主要是为了改变这种“应试教育”的局面，主要是为了改变目前这种“8亿件换1架”的局面，

主要是为了改变我们跟在空中客车、波音、英特尔、微软、朗讯、松下、索尼、夏普后面喝一点残汤、吃一点剩饭的局面。

我们进行高考改革的目的是主要不是为了追求公平，而是为了保护儿童的好奇心和学习兴趣，而是为了追求教育的效率。今天高考中确实也存在一些公平方面的问题，例如地区间录取率的差异问题，地区间名额分配的不平衡问题，等等。这些都属于高考改革的任務，都需要认真予以解决。但是，这些并不是高考改革所面临的主要问题。

想清楚了我们究竟为什么要进行高考改革，才可能正确看待今天降低大学招生中考试比重的种种努力，包括扩大自主招生比例，增加面试，给高中教师一定的发言权，更多地参考平时成绩和成长记录，等等。

再谈为什么要进行高考改革^{*}

谢小庆

2007年年初，中国青年报社会调查中心完成了一项有2165人参加的关于高考改革的调查。结果显示，92.8%的人认为，目前的高考制度有必要改革。35.2%的人认为“只需小改”，57.6%的人则认为需要“动大手术”。^①这项调查显示，在恢复高考30年以后的今天，社会上已经产生了强烈的高考改革的呼声。

我在《为什么要进行高考改革》一文中表达了自己的主要观点：高考需要改革的主要原因是“缺乏效率”，而不是“不够公平”。延续了1300年的科举考试在100多年前被取消，并不是由于科举“不公平”，而是由于科举“没效率”，是由于科举考试脱离实际，引导人死读书，读死书，是由于科举不是以“德才”取人，而是以“背书”取人。30年前恢复了高考，恢复高考的原因主要是为了“提高教育效率”，而不是为了“维护社会公平”。考试作为一种人员评价手段，它主要是提高效率的工具，而不是维护公平的工具。今天的国家政策明确规定不允许在小学和初中招生中采用考试，就是因为采用了考试就损害了每个儿童与生俱来的平等的受教育权利，就是因为采用了考试就损害了每个儿童都应平等分享优质教育资源的权利。我们今天进行高考改革，主要是为了提高教育的效率，主要是为了改变青少年的普遍厌学状况，主要是为了保护青少年的好奇心和青少年的创新能力，主要是为了改变“读

^{*} 原刊于《中国教师》2007年第10期。

^① 见《中国青年报》2007年2月5日。

一本书、教一本书、背一本书、考一本书”的局面，主要是为了改变我们跟在空中客车、波音、英特尔、微软、朗讯、松下、索尼、夏普后面喝一点残汤、吃一点剩饭的局面。

大学毕业生本来应是我们民族最宝贵的人力资源，不幸的是，一些“宝贵资源”经过高等教育却被转化成“人口负担”。数以百万计的大学生“毕业即待业”的严峻局面，已经将教育的“低效率”问题凸显出来。根据《中国青年报》2007年8月21日的报道，在8月12日举行的中国有色金属工业发展与人力资源开发高层论坛上，有色金属工业人才中心总经理丁学全指出：“一边是企业招不到人，一边是学生找不到合适的企业。”这种“企业为找人发愁、毕业生为找工作发愁”的现象在今天的各行各业都很普遍。这种现象，突出反映了今日中国教育的“缺乏效率”。这之中，既有能力发展方面的问题，也有人格发展方面的问题。

高考改革主要是为了提高教育的效率。为了提高效率，需要通过改革高考来保护青少年的好奇心和青少年的创新能力。保护好奇心和发展创新能力对于提高教育效率是非常重要的，但并不是最重要的，改革高考，还有更重要的原因，就是为了保证学生健康人格的发展。

2007年8月17日，由来自内蒙古呼伦贝尔草原深处的37名6~13岁的少数民族儿童组成的“五彩传说”儿童合唱团在北京展览馆剧场演出，感动了到场的观众。余秋雨先生写出了《追回天籟》一文来表达自己所受到的感动。他在这篇文章中谈到自己在多次担任青年歌手大奖赛评委时的感受：“歌手都很年轻，绝大多数受过严格的专业训练，拥有大专学历。但是，一旦让他们谈谈自己，谈谈父母，谈谈家乡，谈谈音乐，立即出现一种惊人的景象。多数人都不假思索，随口吐出，用词华丽，充满了成语、形容词和排比，却又都严重雷同。他们谁也没有意识到，他们说得多虚假和空洞。不管你怎么追问，他们还给你的，是加倍的虚假和空洞。”

余秋雨先生对青年歌手们的批评，折射出今天在高考指挥棒驱使下的“应试教育”中缺乏“爱心教育”的严重问题，这是比“应试教育”挫伤儿童好奇心和摧残创造力更严重的问题。这使我们联想到夏丏尊在他为《爱的教育》一书写的“译者序言”中对当时的教育的批评：“单从外形的制度方法上，走马灯似的更变迎合，而于教育的生命，从未有人培养顾及。好像掘池，有人说四方形好，有人说圆形好，朝三暮四地改个不休，而对于池所以为池的要素的水，反无人注意。教育上的水是什么？就是情，就是爱。教育没有了情爱，就成了无水的池。任你四方形也罢，圆形也罢，总逃不了一个空虚。”

笔者历来认为，教育的第一职能是发展学生的人格，其核心是品德。其次才是发展学生的能力，再次才是传授学生以知识。在20世纪80年代，笔者曾经批评当时的教育“恰恰是本末倒置了，注重传授知识而忽视发展能力，更不重视发展学生的人格。在这种教育状况下，不仅学生知识面窄，实际分析问题和解决问题能力差，独立思考能力差，而且缺乏社会责任感，缺乏理想和抱负，审美能力低，兴趣狭窄。”^①近20年过去了，令人遗憾甚至沮丧的是，这种局面不仅没有丝毫扭转，而且愈演愈烈。这种忽视学生人格发展的局面，与20年来高考改革的脚步迟缓有很大关系。

有色金属工业人才中心丁学全总经理在分析“企业找不到人、毕业生找不到工作”的原因时指出了两方面的原因，一是由于“行业已经发展到21世纪的高科技了，一些大学还在用20世纪的教材，讲授20世纪的知识”，一是由于“有色金属企业大多在深山里，学生们都不愿去”。前者是由于学校教育在能力培养方面的低效率，后者则是由于学校教育在人格发展方面的低效率。今天，在我国许多欠发达地区对大学毕业生有较高的需求。但是，相当数量的大学毕业生宁愿留在家中“啃老”或“混在城市”，也不愿

① 谢小庆：《心理测量学讲义》，华中师范大学出版社1988年版，第6页。