

# 招生考试研究

Admission & Examination Research

2013

3

总第19辑

上海市教育考试院 主办



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

# 招生考试研究

上海市教育考试院 主办

(2013-3)

上海交通大学出版社

# 招生考试研究

(上海市教育考试院主办)

## 编 委 会

总 顾 问：王荣华

顾 问：(以姓氏笔画为序)

沈晓明 林蕙青 戴家干

主任委员：李瑞阳

委 员：(按姓氏笔画为序)

王 钢 王斯德 王厥轩 刘海峰 张民选

张华华 杨学为 陈 勇 胡启迪 谢小庆

蔡达峰 雷新勇 漆书青

主 编：李瑞阳

执行编辑：褚慧玲 李立峰 王 彬

英文编辑：觉 先

通讯地址：上海市钦州南路 500 号 《招生考试研究》

邮 编：200235

电 话：021-64946609 021-64946068

传 真：021-64946679

电子邮箱：zkyj@shmeea.edu.cn xdlf2000@163.com

## 图书在版编目(CIP)数据

招生考试研究. 2013 年. 第 3 期 /上海市教育考试院

院主编. —上海:上海交通大学出版社, 2013

ISBN 978-7-313-10547-9

I. ①招… II. ①上… III. ①招生—考试—中国—

文集 IV. ①G473.2-53

中国版本图书馆 CIP 数据核字 (2013) 第 259279 号

## 招生考试研究(2013-3)

主 编: 上海市教育考试院

出版发行: 上海交通大学出版社

邮政编码: 200030

出 版 人: 韩建民

印 制: 上海交大印务有限公司

开 本: 787mm×1092mm 1/16

字 数: 157 千字

版 次: 2013 年 12 月第 1 版

书 号: ISBN 978-7-313-10547-9 AG

定 价: 25.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 6.75

印 次: 2013 年 12 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系

联系电话: 021-54742979

# 目 录

## Contents

### • 特别关注 •

教育测量:从数学模型到法学模型

Education measurement: from math model to legal model

谢小庆 (1)

### • 口述历史 •

艰难的探索:上海高考单独命题改革的推出

何顺华 (9)

上海春季高考改革的推行与实施

胡启迪 (16)

### • 理论探讨 •

我国大学招生考试制度改革的几个视角

Perspectives on the reform of China's college entrance  
examination and admission system

张继明 (22)

我国高考制度改革的价值取向分析

An analysis of the value orientation in China's  
college entrance examination reform

王 彬 (30)

对高校博士研究生招生工作的思考

Reflections on the recruitment of doctoral  
students of higher education

李京肽 (40)

考试公平论:法和经济学的视角

On fairness in examinations from the perspectives  
of laws and economics

徐光木 (47)

### • 比较研究 •

香港高校招生考试制度改革的价值取向

Value orientation of the college entrance examination

system in Hong Kong

程 军 刘清华 (56)

• 实践探索 •

上海市本科院校体育特长生培养的调查研究

A study on the cultivation of sports students  
in the colleges of Shanghai

车荣华 刘 隽 马 杰 (63)

上海市高等教育自学考试专业合作现状分析及建议

On the professional cooperation in the self-study examinations  
in Shanghai's higher education——An analysis

of the status quo and suggestions for improvement 范玲玲 孟 萍 曹 虹 (77)

• 史海钩沉 •

清代解元进士之量化分析

A quantitative analysis of Jie Yuan who passed the Metropolitan  
Examination and became Jin Shi in Qing Dynasty

吴根洲 (84)

科举文化学的现象学诠释

The phenomenon interpretation of imperial examinations culture

冯用军 (90)

## 教育测量：从数学模型到法学模型

谢小庆

**[摘要]** 美国教育协会和美国国家教育测量学会共同组织编写的《教育测量》在业内被称为是“测量领域的《圣经》”。2006年出版的《教育测量(第4版)》将图尔敏的论证模型作为效度研究的基本范式。这不仅是效度研究范式的转变,更标志着教育测量研究从数学模型向法学模型的转变。本文讨论了这种教育测量研究范式的转变。

**[关键词]** 测验;考试;教育测量;图尔敏

效度研究是教育和心理测量研究领域中最重要的问题。美国教育协会(American Council on Education)和美国国家教育测量学会(National Council on Measurement in Education)共同组织编写的《教育测量》(*Educational Measurement*)在业内被称为“教育测量领域的《圣经》”。在2006年出版的《教育测量(第4版)》中,将图尔敏的论证模型作为效度研究的基本范式<sup>①</sup>。在新的效度研究范式中,“理据(warrant)”成为核心概念,效度研究被视为一种通过构造理据系统、理据链条和理据网络而对效度进行的“论证(argument)”,效度研究被视为一种对测验分数做出普乐好(plausible)解释的过程<sup>②</sup>。

作为一门学科,教育测量学已经走过了百余年的历史。在教育测量学的发展历史中基本的研究模型是数学模型,是借助数学工具改进教育评价的质量,从而提高教育评价的有效性、可靠性和公平性。百年间,教育测量研究的数学模型取得了很大的成绩,研究成果被广泛地应用于考试实践,既促进了教育的公平,也提高了教育的效率。每当数学家和计算技术专家发明了一种新的数学工具或计算技术,教育测量学家往往会在第一时间就拿来应用。有时候,教育测量学家甚至会自己发明一些新的数学工具用于自己的研究。例如,今天被广泛应用于自然科学和社会科学各个领域的因素分析(factor analysis)方法,就是由斯皮尔曼

---

**作者简介:**谢小庆,男,博士,研究员,北京语言大学教育测量研究所名誉所长。

<sup>①</sup> Brennan R L, ed. *Educational Measurement* (4<sup>th</sup> edition) [M]. Washington, DC: American Council on Education/Praeger, 2006:17-64.

<sup>②</sup> 谢小庆. 测验效度概念的新进展[J]. 考试研究, 2013(3):56-64.

谢小庆. 效度:从分数的合理解释到可接受解释[J]. 中国考试, 2013(7):3-8.

(Spearman)、瑟斯顿(Thurstone)、卡特尔(Cattell)等人在教育测量的研究中发展起来的一种数学工具。

Mislvey 是今天国际教育测量领域中最有影响的学者之一。他曾说:“如果我们将支配着今天教育测量活动的测验理论概括成将 20 世纪的统计学应用于 19 世纪的心理学,那么,这不能算是一种夸张”<sup>①</sup>。Mislvey 已经看到,在描述人的复杂的心理过程时,仅仅靠数学模型是不够的。实际上,人的心理过程的复杂性,远远超出最先进的数学模型的处理能力。教育测量,不仅需要先进的数学模型和计算技术,还需要先进的心理学模型。

## 一、数学模型的局限性

在数学模型被应用于教育评价之前,教育评价主要依靠基于个人经验之上的定性方法。在许多情况下,我们都可以根据经验做出正确的判断,并做出正确的决策。但是,由于不同的研究者各自的经验、观点、倾向、偏好不同,有时候对同一问题会得到非常不同的结论。例如,客观性选择题能否考察出学生的写作能力?英语 4,6 级考试成绩应否作为学位授予的必要条件?物理系招生是否需要语文最低分数要求?古代汉语研究生招生是否需要外语最低分数要求?……对于许多问题,不同教师的看法相去甚远,甚至截然不同。孰是孰非?仅凭各自的经验只会争论不休。为了支持自己的观点,双方都可以举出大量的个案,然而个案是不足为据的。这些时候,就需要借助定量方法,需要借助统计方法。这时,定量分析可以帮助我们各执一词的争论中摆脱出来。因此,在几乎整个 20 世纪,数学模型成为教育测量研究的基本研究范式。

在教育心理测量的研究过程中,人们逐渐认识到数学模型的局限性。这种局限性,突出地表现在以下几个方面:

### 1. 概率的逻辑基础

今天应用于教育测量的数学方法大部分基于概率理论之上。然而,概率理论能否应用于教育测量研究,尚是一个颇值怀疑和非常棘手的问题。何谓概率?通俗讲即“重复试验中事件发生的可能性”。对于概率概念,“重复试验”是一个非常重要的前提。例如,只有多次重复抛掷硬币,才可能得到正面朝上的概率。倘若是不可重复的试验,倘若每次抛掷时硬币的重量、质地、成分、形状等会发生变化,就无所谓概率。教育研究的对象是人,每个人不仅具有不同的遗传特点,而且经历、需要、欲望、情感、能力水平等各异。对不同的人进行的观察能否被视作重复试验呢?能否被视作与将一枚硬币多次抛掷相似的重复试验呢?这个问题的答案至少不是不言而喻的。

退一步,即使我们将对不同的人的观察视为重复试验,这种基于概率基础之上的统计规

---

<sup>①</sup> Mislvey, R. J., Foundations of a new test theory, in Test for a new generation of tests [M]. New Jersey: Lawrence Erlbaum Associates Publishes, 1993:19.

律性能否成为关于有个性的人的教育决策的依据呢？这仍然是一个需要讨论的问题。即使根据我们的研究知道具有某一组神经生理心理特点的人中 90% 难以完成某一水准的学业，我们能否根据这一研究结果而预言一个具有这些神经生理心理特点的人不能完成学业呢？这里，人的能动性是一个不容忽视的因素。不用说预言一个具有能动性、选择性的人，即使是从大量抛硬币中得到的统计规律，对于预测下一次抛掷硬币的结果也是毫无意义的。

归根结底，概率方法是基于归纳逻辑之上的。然而，正如恩格斯早在 100 多年前所指出的：“按照归纳派的意见，归纳法是不会错误的方法，但事实上它是很不中用的，甚至它的似乎最可靠的结果，每天都被新的发现所推翻<sup>①</sup>”。将归纳法用于研究死的、被动的物理现象尚存在着“不中用”的一面，对于研究能动的、有选择性的人的心理现象，则具有更大的局限性。

## 2. 显著性检验

今天教育测量研究中经常运用的一种定量分析方法是显著性检验，包括正态检验、t 检验、卡方检验、F 检验等。通常，只有基于一定的定性分析之上，在一定的问題情境之中，显著性检验才有意义。

当我们用组间比较的方法<sup>②</sup>对一项测验的效度进行论证的时候，我们可能犯两种错误，一种是“接受无效测验”，一种是“拒绝有效测验”。在统计学中将这两种错误分别称为“第一类错误（或  $\alpha$  错误）”和“第二类错误（或  $\beta$  错误）”。犯某一类错误的可能性的减少必然以犯另一类错误的可能性的提高为代价。差异显著性标准愈严格，就愈不容易犯“接受无效测验”错误，同时，就愈容易犯“拒绝有效测验”的错误。显著性检验标准的设定（ $P=0.001, 0.01, 0.05, 0.1$  或  $0.2$ ），需要根据测验的实际应用情况来确定。对于用于飞行员选拔的测验，我们可能会设定较严格的标准；对于用于高中招生的测验，我们可能会设定较宽松的标准。设定怎样的标准，需要基于先于显著性检验的定性分析之上。

## 3. 相关分析

今天教育测量研究中最常用的定量分析方法是相关分析，包括回归分析、因素分析等。这些定量分析方法可以揭示出事物之间的相关关系。然而，相关并不等于因果。学校早上八点上课，商店早上九点开门，两者相关很高，但并不存在因果关系。人类的许多误解都是源于错误地对事物之间的相关关系作出了因果的解释。“重物下落较快”这一错误看法就是由于人们对“质量大”和“下落快”之间的相关关系错误地作出了因果解释。“心脏是思维的器官”这一错误看法就是由于人们对心跳与思维之间的相关关系错误地作出了因果的解释。

教育测量总是力图揭示考生的心理属性与教育成果之间的因果联系，从而为教育决策提供依据。相关是因果关系的必要条件，但不是充分条件。能否对相关关系作出因果的解释？仅仅靠定量分析是不够的。

<sup>①</sup> 恩格斯. 自然辩证法[M]. 北京：人民出版社，1972：206.

<sup>②</sup> 谢小庆. 心理测量学讲义[M]. 武汉：华中师范大学出版社，1988：156.

## 二、图尔敏的工作逻辑学和论证模型

斯特芬·图尔敏(Stephen Toulmin, 1922~2009)是一个在世界范围有广泛影响的科学哲学家和逻辑学家。他出生于英国,1948年在剑桥大学因逻辑学研究获得博士学位。他在剑桥大学所受到的数学、逻辑学和物理学方面的良好训练为他日后的科学哲学研究打下了坚实的基础。在剑桥学习期间,他直接得到当时在剑桥任教的罗素(Bertrand Russell)和维特根施坦(Ludwig Wittgenstein)的指导。他既受到罗素和早期维特根施坦的分析哲学的影响,看重理性和严格的形式逻辑在认识世界中的作用,也受到晚期维特根施坦的影响,认识到理性和形式逻辑的局限性。他1958年出版了《论证的使用》(*The Uses of Argument*),从“概率”概念入手,揭示了理性和形式逻辑在面对复杂的科学、社会问题时存在的局限性。他发现,仅仅借助于数学模型和形式逻辑,很难在现实生活中形成有效的论证。对于一个理论、一个观点、一个命题的论证,不是一个可能立即得到答案的实验室研究,不是一场可以决出胜负的球赛。一个新理论、新观点被接受,一个旧理论、旧观点被放弃,往往是一个漫长的过程,往往是一个旷日持久的论证过程。就像科学理论不可能被“证实”一样。持有某种观点的人完全将自己的论辩对手说服的情况很少,持有某种观点的人将所有的论辩对手说服的情况很少。

图尔敏对以形式逻辑为主体的传统逻辑学进行了反思,对始于亚里士多德的以“三段论”为代表的逻辑学体系进行了反思,对罗素和怀特海(Alfred Whitehead)所进行的逻辑学数学化的努力进行了反思。他认为,逻辑学的出发点不应是符合逻辑的理论,而应是符合逻辑的实践;逻辑学不应局限于研究理想的逻辑,更应该研究工作的逻辑(*working logic*),更应该研究日常生活实践中的逻辑。他指出,那种数学化的、跨时间的、跨学科领域的逻辑远远不能满足实际生活中论证和决策的需要。他认为,逻辑学中不仅需要包含形式逻辑,还需要包含非形式逻辑;不仅需要包含数学模型或几何学模型,还需要包含法学模型。

在正视理性和形式逻辑局限性的基础上,为了进行更有效的论证,图尔敏提出了一个同时基于形式逻辑和非形式逻辑基础之上的论证模型。在这个论证模型中,包含资料(*datum*, D)、必要条件(*backing*, B)、理据(*warrant*, W)、限定(*qualifier*, Q)、反驳(*rebuttal*, R)和结论(*claim*, C)等6个基本要素。论证的基本过程是:资料(D)和必要条件(B)共同构成了理据(W),在接受了反驳(R)之后,经过限定(Q),使结论(C)有条件地得以成立<sup>①</sup>。

2006年出版的权威文献《教育测量(第4版)》将图尔敏的论证模型作为效度研究的基本模式。

---

① 谢小庆. 测验效度概念的新进展[J]. 考试研究, 2013(3): 56-64.

谢小庆. 效度: 从分数的合理解释到可接受解释[J]. 中国考试, 2013(7): 3-8.

### 三、数学模型与法学模型的区别

将图尔敏的论证模型作为效度研究的基本模式，标志着教育测量范式从数学模型向法学模型的转变。

2009 年圣诞节，美国德克萨斯州 30 岁的吉尔伯特(Ezekiel Gilbert)因 23 岁的应招女弗拉戈(Ive Frago)收取了 150 美元后拒绝提供性服务，将她枪杀。2013 年 6 月 6 日，法院终审宣判吉尔伯特无罪。在这起案件中，指控被告有罪的检察官是有理由的：被告不应为 150 元就夺取一个人的生命；为被告辩护的律师也是有道理：私有财产不可侵犯，被告的行为是为了保卫自己的私有财产。最终，陪审团基于“保护私有财产”的考虑支持了律师。在这里，既没有正确的(right)判决，也没有合理的(rational)判决，仅仅有普乐好(plausible)的判决。

2012 年 2 月 26 日，28 岁的协警齐默尔曼(George Zimmerman)巡逻时射杀 17 岁黑人少年马丁(Trayvon Martin)。2013 年 7 月 13 日，法院终审宣判齐默尔曼无罪。在这起案件中，指控被告有罪的检察官是有理由的：马丁并没有携带武器，被告使用武力过当，剥夺了一个并无大错的年轻人的生命；为被告辩护的律师也是有道理：警察是高危行业，需要得到社会的高度保护。最终，陪审团基于“保护警察安全”的考虑支持了律师。在这里，既没有正确的(right)判决，也没有合理的(rational)判决，仅仅有普乐好(plausible)的判决。

从这两个案例可以看出一些“法学模型”的特点：首先，需要以事实为依据，判决不能基于虚假或虚构的事实之上；其次，必须符合形式逻辑，判决不能与形式逻辑相冲突，必须是合理的；第三，在符合前两项的基础之上，基于不同的价值取向，可能存在多种可能的选项，这些选项，基于不同的前提假设和价值倾向之上，它们之间的区别在于是否属于普乐好的一项。

法学模型将教育测量研究视为围绕考试的效度、信度、公平性、及格标准设定等问题进行的辩论，类似于法庭上控辩双方的辩论。一项考试的支持者努力为考试进行辩护，努力拓展考试的应用领域。相反，该考试的反对者则提出种种质疑，谨慎地界定考试的应用范围，避免考试分数的误用。

数学模型与法学模型的区别主要表现在：

#### 1. 时间依赖性

数学模型独立于时间，不存在时间维度。 $1+1=2$ ， $2^3=8$ ，世代永远如此。法学模型则不同，有时会表现出与时俱进的特点。在美国，不久以前，买卖黑奴是合法的，今天已经属于非法。不久以前，黑人乘坐白人专用的公共汽车是违法的，今天已经成为合法的。1984 年 6 月 30 日邓小平接见日本客人时说：“所谓小康，从国民生产总值来说，就是年人均达到 800 美元……国民生产总值可以达到一万亿美元<sup>①</sup>”。2012 年，中国人均国民生产总值已经

① 邓小平，邓小平文选(第三卷)[M]. 北京：人民出版社，1993：64.

超过6 000美元,总量已经超过8万亿美元。“小康社会”是否已经实现?对此,需要进一步的论证。1995年,根据教育部668号文件,外国留学生达到汉语水平考试(HSK)3级才可以进入理工科学习专业,HSK达到6级才可以进入文史哲和中医学科学习专业。由于来华留学生人数的快速增长,今天,北京大学等一流中国大学实际采用的留学生入系学习专业的汉语标准,已经远远高于1995年的教育部标准。

## 2. 领域依赖性

数学模型具有跨各个研究领域的一致性。 $1+1=2$ ,  $2^3=8$ , 对于各个研究领域是一致的,对于物理学是这样,对于生物学也是这样。法学模型则不同,在不同的研究领域中可能表现出不同的特点。在对于盗窃罪的认定上,盗窃一般民用物资与盗窃军用危险品是不同的。在塑料拖鞋生产车间,合格率标准可以是95%;在载人航天器关键元件的生产车间,合格率则要达到99.99%以上。在作文评分时,误差控制范围可以是总分的10%;在采用光电读卡器对选择题试卷进行扫描时,误差控制范围可以是 $10^{-5}$ ;在作弊甄别时,误差范围则要求控制在 $10^{-17}$ 以下。

## 3. 情境依赖性

数学模型具有跨各种情境的一致性,在各种不同的情境中,数学模型计算得到的结果是一致的。法学模型则不同,即使在同一研究领域中,对于不同的情境,也可能作出不同的选择。在强奸罪的认定上,成年受害人与未成年受害人是不同的。医师、护士、律师、会计师等职业资格考试的合格标准,关系到患者、当事人和顾客的利益。如果报考者达不到必要的能力和知识要求,患者、当事人和顾客的利益就会受到损害,安全就得不到保障。按照数学模型,“合格标准”应该是全国统一的。但是,今天在东部沿海,许多人手握资格证书却找不到工作岗位;在西片偏远地区,严重缺乏这些方面的专业人才。在职业资格的合格标准设定上,往往需要根据实际情况对合格标准进行调整。

## 4. 答案的唯一性

借助数学模型,一般可以得到唯一正确的答案。法学模型则不同,依据事实并符合逻辑的答案,往往不是唯一的。无论是对枪杀应召女的吉尔伯特的无罪判决,还是对枪杀17岁黑人青年的齐默尔曼的无罪判决,都不是唯一的“正确答案”,都仅仅是多种合理选项之一。

2013年7月18日,作为某项职业资格考试专家委员会成员,笔者与专家委员会委员们共同确定了该项考试2013年的及格分数线。考试主持单位从4个方面对及格线设定进行了研究。第一是安哥夫方法,第二是借助作为“外锚”的共同题实现的试卷分数等值,第三是以几十所业内骨干学校的近10万名考生作为样本的等百分位等值,第四是2013年该项专业人员的需求分析。4个方面的研究结果互相验证,结果高度一致。综合4个方面的研究结果,专家委员会最终面临两个候选方案,两个方案仅仅相差1分。虽然仅仅相差1分,却关系到12 000余名考生能否取得职业资格。

事实上,相差1分的两个方案都是符合事实的,也都是合理的,“高1分方案”有利于保护服务对象的利益,“低1分方案”则有利于保护求职者的利益。最后,专家委员会经过无记

名投票,选择了低分方案。

低分方案胜出的主要原因是“西部因素”。专家委员会对受到影响的 12 000 余名考生的构成进行了分析,发现其中相当大的比例来自新疆、西藏、青海、甘肃等西部省份。这些省区,严重缺乏此类专业人员,人员缺口巨大。因此,多数专家投票支持了低分方案。从此例也可以看出,与不受情境因素影响的数学模型相比,法学模型往往需要考虑情境因素。

#### 四、教育测量范式转变对学习革命的影响

2013 年 4 月 26、27 两日,在美国国家教育测量学会(National Council on Measurement in Education,简称 NCME)年会期间,NCME 安排资深教育测量专家们进行了长达 16 小时的专业培训,集中介绍教育测量领域的一些最新进展。这次培训课程向亚洲、非洲和南美洲的一些国家免费直播。与旧金山课堂现场的学员们一道,北京语言大学教育测量所的研究生们通过网络收看了这些培训课程。借助专用的网络教学软件,北京的学员不仅可以在同一个屏幕上同步看到讲课教师和 PPT 课件,而且随时可以向讲课教师提问。

计算机技术和网络的迅速发展,正在改变着传统的学习方式。一轮新的学习革命正在向我们走来。

教育测量研究从数学模型向法学模型的转变,将改变传统的教育评价方式,将对新的学习革命产生重大的影响。参照法学模型来审视认识的发展,不难发现,对于一个理论、一个观点、一个命题的论证,就像对于吉尔伯特和齐默尔曼的案件审理一样,并不存在唯一正确的答案。辛亥革命已经过去了百年。今天,辛亥革命对于中国现代化进程的影响仍然是激烈争论的话题。“五四”已经过去了近百年。今天,“五四”对于中华民族文化建设的正面和负面的影响仍然是学术界激烈争论的话题。“罗斯福新政”已经过去了近 80 年。今天,对其得失成败仍然存在巨大争议,仍然是经济学家们和政治学家们激烈争论的话题。

今天,中国学校中广泛流行的是形成于 20 世纪 50 年代的学习方法,是深受前苏联影响的学习方法。学校中广泛流行的是形成于 20 世纪以前的“真理—谬误”的简单思维方式。这种思维方式把学习过程理解为一个学生学习和掌握“科学真理”的过程,理解为一个老师向学生传授“科学真理”的过程。事实上,在今天的学校中讲授的许多标有“科学真理”标签的东西都是非常可疑的。这种学习方式,大大地摧残了学习者的好奇心,大大地打击了学习者的怀疑精神,大大地压抑了学习者的创造性。

教育评价方式的改变,将改变这种陈旧的学习方式。教育将不再是简单地向学生灌输特定的结论,而是倡导研究性的学习,而是发展学生的审辩式思维能力(critical thinking competency)。在学校中,学习者的好奇心将受到小心翼翼的呵护,学习者的怀疑精神将受到鼓励。教育者将努力保护和激发学习者的创造力,倡导研究性的学习,倡导审辩式论证。这样,学习将成为一个探索和发现的过程,而不仅仅是一个记忆和拷贝的过程。这样,学习

将成为一个快乐和享受的过程,而不再是一个枯燥和痛苦的过程。

## 五、结 语

在 2006 年出版的《教育测量(第 4 版)》中,将图尔敏论证的模型作为效度研究的基本范式。这不仅是效度研究范式的转变,这标志着教育测量研究从数学模型向法学模型的转变。这种转变,将改变传统的教育评价模式,将对新一轮的学习革命产生重要的推动作用。

---

### Education measurement: from math model to legal model

Xie Xiaoqing

[Abstract] A new validation paradigm based on Toulmin argument model was proposed in *Educational Measurement* (4th edition), jointly published by American Council on Education and American Educational Research Association in 2006. This new validation model marks a transformation from math model to legal model in the field of education measurement research, which is the theme under discussion in the present paper.

[Keywords] test, examination; education measurement; Toulmin

## 艰难的探索：上海高考单独命题改革的推出

口述：何顺华

采访：李立峰、王 彬

整理：李立峰

**口述前记：**何顺华，男，上海人，1932年2月出生，1953年6月加入中国共产党。1986年前任上海市人民政府教育卫生办公室教育处副处长，后调任上海市教育考试中心办公室主任，市教育招生考试中心正处级副主任，副研究员。1995—1996年获得上海市机关事业单位“老有所为”精英奖，研究成果集中在上海的教育招生考试研究方面，主要成果有《略论上海高中会考和高考制度的改革》、《完善考试制度，深化教育改革》等。

**采访信息：**在系统了解上海高校招生考试制度的基础上，经过搜集整理资料、拟定访谈提纲，我们于2013年5月6日，对上海市教育招生考试中心副主任何顺华先生进行了采访，采访地点在本市何顺华先生的寓所。采访结束后，经过多次的文字整理和查阅档案资料后，再请何顺华先生对此文进行审阅校对，最后形成定稿。

### 一、落实办学自主权引发了高考改革

上海高考单独命题的改革，要回溯到对高校招生自主权改革的呼吁。恢复高考之后，对高校办学自主权改革的呼声越来越高。1979年12月，《人民日报》刊登了上海复旦大学的苏步青、同济大学的李国豪、华东师大的刘佛年和交通大学党委书记邓旭初等四个大学领导的文章，呼吁给高校一点自主权。几个校长一致认为，在现行的管理体制下，校长的手脚被束缚住了，很难办好学校，甚至校长在“每年招生的时候，连招收一个学生的权利都没有”；教育行政部门在管理高校时不能只用行政手段。这些主张和观点在当时引起了很大的社会反响。

时任上海市副市长（分管教育）的杨凯，看到了这四位校长的文章，就请市政府教育卫生办公室（简称教卫办）的相关同志一起研究如何解决这一问题。当时奚心雄同志分管教育，

我当时在教卫办的教育处。我记得当时杨凯指示教卫办的同志,要用报纸来回答这四个校领导的问题。经过教卫办的反复讨论,初步的考虑是:高等学校招生的改革要与中学的教育教学改革联系起来一起考虑。当时大家批评最多的是学生负担重,因为高考是分科考试,要考7门,科目比较多,致使中学生负担较重。中学普遍反映高三这一年等于是考试年了。所以我们的设想是建立一个用来检测学生学业水平的会考制度,用会考把关,学生成绩达到一定的要求以后,高考考试的科目能减少一些。因此,我们就提出上海要单独高考命题、市内统一进行会考,在会考基础上,减少高考科目。改革的目的是减轻学生负担,提高中学教学质量。

围绕这一改革设想,上海市教卫办请四大高校的负责人苏步青、刘佛年、李国豪和邓旭初召开座谈会,听取他们对改革草案的意见与建议。几位校长也赞成用会考来减轻学生负担的指导思想,他们认为,会考达到一定的要求以后,学生不仅完成了高中的基本学业,而且拿到了参加高考的资格。在此基础上,高考的科目可以减少一些,这样一来,就可以给学生减轻学业和心理压力。校长们还提出,在高校招生时给高校一定的自主权,比如学校招100个学生,在投档时可以投120个,给高校有自由选择的空间。不仅一本院校可以这样做,而且二本院校也可以采取120%投档的方式。

除了高校领导的座谈会之外,我们还开了一些中学校长的座谈会,参加座谈会的有上海中学的唐盛昌和时任育才中学名誉校长的段力佩。对会考以及会考基础上减少高考科目的设想,中学普遍表示欢迎,他们认为这样可以有效地减轻学生的学业负担,解放一部分学生,使学生不再背负高考的沉重压力。

## 二、争取高考单独命题的过程

1985年,上海拿到了高考的单独命题权,这是上海高考改革标志性的一个事件,因为有了单独命题权,上海才开始实行会考基础上的高考科目改革等一系列的招生考试制度改革。1985年秋季,我从教卫办的教育处调到了教育考试中心。其实,我们拿到单独命题权的时候,实际上已经在谈会考的事情了。会考要先行试点,上海是全国第一家开始会考试点的省市。随后,浙江也开始进行会考的试点。上海会考改革的特色在于将“会考改革和高考改革配套在一起”,这句话是很完整的一句话,在会考基础上减少高考,如果光减少高考科目,中学生就偏科了,但是有了会考的基础就不存在偏科的问题了。所以这两个问题是连在一起的、完整的。

当时面临的困难是原来的高考命题权不在地方,而是统一集中在中央,是全国统一命题。所以在当时上海要出台会考基础上减少高考的改革方案,面临的重大问题就是高考的单独命题权。对国家教委来说,会考你地方可以自己命题,你要考就去考,而且会考本来就应该是中学出题目,现在是集中起来统一出题目,没有问题;但问题是高考的命题权在国家教委手里。从国家教委的态度来看,一开始不同意单独命题,因为害怕这个事情会起连锁

反应，上海单独命题了，那北京呢？天津呢？事实上，国家教委在1985年批准了上海单独命题之后，一直到2002年，各省要求自主命题的呼声渐高，自主命题才开始大面积的推行。

当时的情况是，单独高考命题权是同基础教育的改革和会考改革结合起来的，正因为实行了会考制度，就可以减少高考的科目了，所以国家教委就批准了上海的高考单独命题。会考制度最先在上海和浙江搞试点，到1991年在全国范围推行会考制度。上海在向国家教委提出要求单独命题时，就已经考虑到会考的问题了。但问题的关键并不是建立会考制度，因为单单会考也不存在单独命题的问题。当时，上海的改革面临着一个非常重要的问题，也就是全国区域差异的问题。对上海来说，中国幅员这么辽阔，全国一张考卷，当然有问题。另外一个问题就是教材，西部边远地区、中部沿海地区和其他发达地区都用这个教材，就有问题了。而中学的教育改革一定是教学内容的改革，上海教学内容的改革要根据上海的情况来改。上海的一期课改大概是1987年正式推出的，课改的推出直接影响到中学的教学，作为对中学教学的评价，高考制度必然要面对课改的问题，因此上海必须要单独命题。从目前来看，上海其实是蛮特殊的，因为我们的教材是可以自己编的，课程标准也是自己制定的，其源头是1985年那个时候留下来的。我们去申请高考单独命题，就是为了推进一期课改，然后会考是在有了高考命题权的基础之上再来推进的。在上海市高招办上报国家教委的请示报告中就明确指出，“我们的目标是：从一九八五年本市单独命题着手，经过两三年的努力，过渡到高中毕业生全市组织会考，减少高考科目；再逐步过渡到一般高校招生不再考试，即凭考生中学会考成绩和志愿进行录取……因此，上海要求一九八五年单独命题考试，是整个高校招生改革的一个组成部分，也是改革中一个试验，为今后进一步改革摸索经验<sup>①</sup>。”所以说，我们当时已经有了一套完整的会考方案，这个方案是将课程改革、会考和高考单独命题结合在一起考虑的。

我记得，申请高考单独命题的过程是相当艰苦的，时任教卫办副主任的奚心雄代表上海去国家教委汇报工作，前后一共去了五次。从国家教委的情况来看，也有非常支持上海高考改革的，如时任国家教委学生司副司长的杨学为就非常支持上海的改革，他一直主张考试制度要进行改革。为更好地观察和指导上海统一会考基础上的高考改革，他有段时间常驻在上海，住在华东师大招待所，而且还参加每次教卫办组织的座谈会。但也有人理解上海的高考改革。我们与国家教委前后讨论了很多次，经过多次的讨论之后，国家教委逐步开始接受上海的改革试点。从国家教委来讲，搞统一会考，它没有什么大的意见。最关键的问题是高考要单独命题，上海要把命题权拉到上海来，如果上海拿到的话，接下来北京就来了。

为什么向国家教委反复汇报了好几次？一个是他对我们的高考科目减少有想法。减少科目的方案就是语、数、外是必考的，不管文科还是理科。具体方案是理科考语、数、外加上一门物理或化学或生物；文科考语、数、外再加上一门政治或地理或历史。这样高考的科目

<sup>①</sup> 《关于一九八五年上海市全日制普通高校招生试行改革的请示报告》沪高招(84)第10号，载上海市教育考试院编，上海市普通高中会考和普通高校招生考试制度改革1985—1995[Z]，1996：9-14。