



高考效度研究

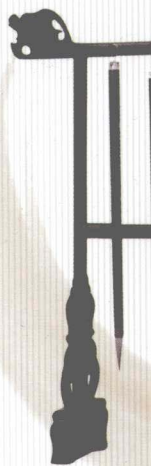
高考改革研究丛书

刘海峰 主编

吴根洲 著



華中師範大學出版社



本书获教育部哲学社会科学研究重大课题攻关项目
“高校招生考试制度改革的理论与实践研究”（03JZD0038）资助

高考效度研究

高考改革研究丛书

刘海峰 主编

吴根洲 著



华东师范大学出版社

新出图证(鄂)字 10 号

图书在版编目(CIP)数据

高考效度研究/吴根洲著. —武汉:华中师范大学出版社,2008.7
(高考改革研究丛书/刘海峰主编)
ISBN 978-7-5622-3733-4

I. 高… II. 吴… III. 高等学校—入学考试—研究—中国
IV. G647.32

中国版本图书馆 CIP 数据核字(2008)第 057802 号

高考效度研究

© 吴根洲 著

责任编辑:肖颖

责任校对:方汉交

封面设计:甘英

编辑室:文字编辑室

电话:027-67863220

出版发行:华中师范大学出版社

社址:湖北省武汉市珞喻路 152 号

电话:027-67867076(发行部) 027-67861321(邮购)

传真:027-67863291

网址:<http://www.ccnapress.com>

电子信箱:hscbs@public.wh.hb.cn

印刷:武汉市新华印刷有限责任公司

督印:章光琼

字数:265 千字

开本:787mm×1092mm 1/16

印张:15.75

版次:2008 年 7 月第 1 版

印次:2008 年 7 月第 1 次印刷

印数:1—3 000

定价:33.00 元

目 录

绪 论	(1)
第一章 高考与效度的理论分析	(21)
第一节 高考辨析	(21)
第二节 效度概论	(35)
第二章 高考效度的定量研究 (上)	(53)
第一节 定量研究说明	(53)
第二节 以大学学业总成绩为效标度量高考的效度	(71)
第三节 以三种学业成绩分别为效标时高考效度的比较	(84)
第三章 高考效度的定量研究 (下)	(95)
第一节 不同群体的高考效度	(95)
第二节 高考各科目成绩与大学学业成绩的相关分析	(110)
第三节 高考英语与高考数学的效度	(120)
第四节 负相关问题与录取标准科学性问题的初探	(131)
第四章 高考效度的解释与提高	(146)
第一节 高考效度的解释	(146)
第二节 高考效度的提高	(163)
第五章 高考效度提高的实践条件	(179)
第一节 直接条件	(179)
第二节 间接条件	(194)
结 语	(213)
附 录	(222)
参考文献	(238)
后 记	(246)

绪 论

恢复高考后的三十年里，高考作为最重要、影响最大的教育考试一直牵动着亿万中国人的神经。高考在人们的一片赞扬声中仅仅徜徉了几个年头，就开始被不断升级的非议所笼罩。对高考的非议主要有两个方面：一是认为高考不能为高等学校选择合适的生源；二是认为高考与应试教育、与学生负担过重密切相关，它扭曲了教育也影响了学生。这两个方面在不同历史时期的影响不同，表达的方式也有所区别。应该说，前一种非议首先出现，在最初曾占据上风，声音主要来自部分教育专家和中学教师，因此表达方式较为理性，一系列的改革也围绕此问题展开；后一种非议随后出现并逐渐占据主流，同时由于声音来源逐渐复杂化，某些表达方式超出了理性的限度，高考改革的动因也越来越多。若不加以严格界定，20世纪80年代主要是基于高考自身的科学性进行讨论与改革，主要成果是高考命题的标准化；20世纪90年代主要是基于如何“减负”、如何实施素质教育来探索改革之路，主要成果是科目调整。

进入21世纪以来，两种非议扭结在一起深深地影响着高考的命运。对后一种非议的回应已经拥有了比较丰硕的成果，而对前一种非议的回应则相对欠缺。在高等教育大众化的背景下，对高考能否为高等学校选择合适的生源这一问题作出科学的回答显得更为迫切，然而现有的答案却似乎不甚有力。三十年来，不断有新的高考改革方案付诸实践，停留在纸面上的方案更是不计其数。然而，方案的形成有多少实证材料的支持，经受过多少专家学者严密的论证呢？若要对高考这一复杂的系统工程进行卓有成效的改革，这些都是必需的基础性工作。本书从实证材料出发，依据测验理论，试图正面回答关于高考能否为高等学校选择合适的生源以及与此密切相关的问题，即关于高考的效度问题。

一、研究什么

随着研究者身份的多元化和相关研究的不断深入，研究对象的复杂化与泛

化往往不可避免, 高考改革作为一个牵涉面极广的系统工程, 更不例外。当下, 以整个高考改革作为研究对象已不是一篇论文、一部专著所能承担的任务。然而, 任何关于高考改革的研究必须首先对高考有一个清晰的界定。随着测验理论不断发展, 人们对效度的认识也在逐步深入, 对本研究中的“效度”与“高考的效度”作一个简单的交代也是非常必要的。

(一) 高考的界定

笼统而言, 在中国大陆有两个高考: 普通高考和成人高考。普通高考形成于 1952 年^①, 在“文革”时期被中断了十一年, 1977 年得到恢复并一直延续至今。成人高考则开始于 1986 年^②。普通高考以普通高中毕业生为主要对象, 成人高考以在职在业的成人为主要对象, 尽管近年来两种考试招生对象的界限有所淡化, 但是两者合一也并非朝夕之事。仅以两者目前在高等教育体系中的地位和社会影响而论, 普通高考远远大于成人高考。一般来说, 人们也常把普通高考简称为“高考”, 而把成人高考简称为“成考”。

从字面上理解, 高考是指为普通高等学校招生而组织的考试。然而, 许多考生、家长、教师甚至一些专家学者在使用这一词语时, 常常包含了考试与录取两个环节, 例如, “高考落榜了”、“高考升学率很高”均有录取的含义在内。仅仅作为一种考试, 高考仍有两种不同的指称, 一种是指统一高考^③, 另一种

① 杨学为:《中国高考史述论》, 湖北人民出版社 2007 年版, 第 9~10 页。

② 刘海峰等:《中国考试发展史》, 华中师范大学出版社 2002 年版, 第 365 页。

③ 在不同的文献中有不同的表述方式, 但均指统一高考。(1) 全国普通高校本专科统一考试。参见刘海峰等:《中国考试发展史》, 华中师范大学出版社 2002 年版, 第 331 页。(2) 全国普通高等学校统一招生考试。参见刘海峰:《高考改革的统独之争》, 载《教育发展研究》2006 年第 11A 期, 第 47 页;《1996 年全国普通高等学校统一招生考试》, <http://www.kedu.net.cn/testpaper/gk/200606/118.htm>, 刻度网络。(3) 全国普通高等学校招生统一考试。参见《全国普通高等学校招生统一考试》, <http://baike.baidu.com/view/983160.htm>, 百度百科;《1992 年全国普通高等学校招生统一考试》, http://www.gdjyw.com/jyfg/16/law_16_1140.htm, 学海荡舟;《2006 年全国普通高等学校招生统一考试(广东卷)物理试卷》, <http://www.szzsks.com/pxfd/gkzy/200609/4036.htm>, 苏州招生考试网。(4) 普通高等学校招生全国统一考试。参见教育部:《2007 年度普通高等学校招生工作规定》, http://www.jyb.com.cn/zs/gxzs/ptgxzs/zsxc/t20070306_68019.htm;《普通高等学校招生全国统一考试大纲》, <http://www.eol.cn/html/g/2007dagang.shtml>, 中国教育在线。

是指包括统一高考在内的所有普通高等学校招生考试^①。因此,无论是理论探讨还是实践评说,普通高考有三种含义:等同于普通高校招生,包括招生考试与招生录取两个部分;等同于普通高校招生考试,包括为普通高校招生而举办的招生统一考试、招生联合考试、招生自主考试等各种招生考试;等同于普通高校招生统一考试。三种含义的外延依次缩小。

其实,高考是一个特指名词^②,高考就是指统一高考,它是与高校招生举办的单独考试、几所高校招生举办的联合考试相对而言的,只是统一的含义不是一成不变的。高考形成的初期,时任高等教育部部长的马叙伦指出,所谓统一,是指统一的计划,统一的组织领导,统一的报考,统一的录取调配^③。显然,这些统一是针对整个招生而言的。如果仅把高考定位在招生考试的范围,那么高考的统一性主要是指全国范围内统一报考条件、统一考试日期、统一考试科目、统一命题、统一阅卷标准,大行政区或省区范围内统一组织考试与阅卷^④。

不断的改革使得统一高考的含义逐渐发生了变化。自1985年至1999年,主要有上海市的高考单独命题、从广东省开始推行的高考命题标准化试验、“三南”方案以及随后推行的“3+2”方案等高考科目变革。1999年以后的改革主要有当年开始的“3+X”科目探索、2002年广西的本专分考、2004年分省命

① 认为高考是“高等学校招生考试”的简称。参见颜莉冰:《高考制度评价研究》,厦门大学硕士学位论文,2006年,第76页。时下许多论文把统一高考、高校自主招生考试,甚至保送生考试都纳入高考这一概念之内,有的论文在题目上就明确了观点。如孟照彬:《大学自主招生:高考的最终走向》,载《湖北招生考试》2004年第6期下半月刊,第4页;李文鑫:《高校自主招生应是高考改革的方向》,载《湖北招生考试》2006年第12期下半月刊,第10页;王美林:《自主招生:高考改革的新亮点》,载《湖北招生考试》2004年第12期下半月刊,第14页。中国人民大学的顾海兵教授还专门撰文指出,许多文章作者往往没有区分高考与统一高考,进而指出“高考”不是问题,“统一高考”是个问题。参见顾海兵:《高考与统一高考之辩——兼与孙东东教授商榷》,载《湖北招生考试》2005年第2期下半月刊,第9页。

② 刘海峰:《高考改革的统独之争》,载《教育发展研究》2006年第11A期,第47页。

③ 杨学为:《高考文献》(上),高等教育出版社2003年版,第22~23页。

④ 1958年因受极“左”影响实行各省单独招考,1977年因时间仓促实行分省命题考试,这两个年份除外。

题范围的大幅扩大。然而经过多次改革的高考仍保持了全国范围内考试大纲、报考条件、考试时间的统一,省区范围内组织考试与阅卷的统一,即统一的考试大纲和统一的操作程序。

招生考试与招生录取在理论上是彼此独立的两个环节,在实践中两者的关系也会随环境的变化而变化,统一高考则是整个普通高等学校招生考试的主体与改革的基准点。本研究中的高考取最严格的含义,仅指统一高考,不包括高校招生录取与高校招生其他考试,即高考是指“普通高等学校招生全国统一考试”^①的简称。

(二) 高考的效度

考试的科学性主要用信度、效度、难度、区分度来评价^②。效度是用来衡量考试有效性或准确性的指标,即考试能够测出它欲测特质的程度。难度、区分度、信度的大小也对考试的效度有着直接的影响。因此,效度是评价考试科学性最主要的指标。

传统上,人们将效度分为内容效度、结构效度、效标关联效度三种主要类型。现代教育和心理测量学已经将效度看做是以各种不同类型证据为基础的统一概念^③。发展中的效度概念不再谈及效度的种类,而是指效度凭证的各个方面^④。美国教育研究会、美国心理学协会和全美教育测量学会 1999 年联合出版的《教育与心理测试标准》依据凭证来源把效度凭证分为基于测验内容的凭证、基于解题过程的凭证、基于内部结构的凭证、基于和其他变量关系的凭证、基于测试后果的凭证五个方面^⑤。调查不同方面的效度凭证追求的是相同的基本

① 以《2007 年度普通高等学校招生工作规定》为准。

② 谷振宇:《基于模糊关联规则的高考数据分析》,清华大学工程硕士学位论文,2004 年,第 13 页。

③ 雷新勇:《大规模教育考试:命题与评价》,华东师范大学出版社 2006 年版,第 282 页。

④ [美]美国教育研究会、美国心理学协会、全美教育测量学会著,燕妮琴、谢小庆译:《教育与心理测试标准》,沈阳出版社 2003 年版,第 8 页。

⑤ [美]美国教育研究会、美国心理学协会、全美教育测量学会著,燕妮琴、谢小庆译:《教育与心理测试标准》,沈阳出版社 2003 年版,第 16~24 页。高考作为标准化的考试,在工具层面上等同于测验,因此,在本书的行文过程中,若非特别注明,考试与测验所指是一致的,具体论述见第一章的第一节。

目标：理解分数的意义和正确使用分数。

调查效度凭证，为正确诠释测验分数积累实证，提供充分科学依据的过程被称为效度化或效度检验。一个做得好的效度论证会将各种凭证融为一体而使现有的凭证和理论能支持对指定用途的测验分数的预期诠释^①。在保证质量的前提下，所掌握的效度凭证越多越有利于对测验分数的科学解释与使用。然而，在效度检验实践中，效度凭证能否获取常常受到诸多局限条件的限制，并且不同效度凭证的价值也不是等同的。有些效度凭证在某些情况下特别关键，而其他凭证就不那么重要^②。因此，在一项研究中，总是倾向于以一个相对重要的凭证来源为主。

没有任何一种凭证天生就比别的凭证更可取，而是凭证对预期测验用途的质量和相关性决定具体凭证的价值^③。高考作为一种测验，它的主要用途是为普通高等学校选择合适的生源。对高考分数的解释与使用指向录取决策的作出：为不同高校、不同专业录取合适的生源。度量“是否合适”最为恰当的标准是高考分数与大学学业成绩具有的一致性程度，因此，高考分数与大学学业成绩之间的相关程度便成为表达高考效度最重要的凭证。

尽管“内容效度”、“决策效度”、“效标关联效度”、“测量效度”等术语不再指不同的效度类型，然而在一定的语言背景中，考虑效度术语的不同用法仍然存在某些价值^④。在本书中，大学学业成绩与高考分数的关系作为高考效度的凭证，两者的关系用这两个变量之间的相关系数来表达。此时，高考的效度也可以称之为效标关联效度，这一相关系数被称为效度系数；这一相关系数也可以表达高考分数对大学学业成绩预测的准确程度，因此又可以称为预测效度。

① [美] 美国教育研究会、美国心理学协会、全美教育测量学会著，燕妮琴、谢小庆译：《教育与心理测试标准》，沈阳出版社 2003 年版，第 25 页。

② [美] 美国教育研究会、美国心理学协会、全美教育测量学会著，燕妮琴、谢小庆译：《教育与心理测试标准》，沈阳出版社 2003 年版，第 13 页。

③ [美] 美国教育研究会、美国心理学协会、全美教育测量学会著，燕妮琴、谢小庆译：《教育与心理测试标准》，沈阳出版社 2003 年版，第 25 页。

④ [美] 凯温·R·墨菲、查尔斯·O·大卫夏弗著，张娜、杨艳苏、徐爱华译：《心理测验》，上海社会科学院出版社 2006 年版，第 138~139 页。

二、已有研究

对高考效度问题的探讨依赖于对测验理论的认可与掌握。“文革”结束以后,在恢复高考的同时,测验理论也重新拥有了研究的空间。20世纪70年代末80年代初,一些介绍测验理论的论文成为高考效度问题研究的先声。紧接着,应用测量理论探讨考试效度包括高考效度问题的文章陆续发表,由于作者多为教育统计测量方面的专家,因此这一时期的研究成果具有相当的分量。自20世纪90年代以来,研究高考效度问题的论文几乎未曾间断,然而由于作者知识背景与职业的多元化,对高考效度问题的探讨从不同的角度展开,其研究成果的质量也开始变得参差不齐。世纪之交,一批更为系统、更为深入的测验理论专著在中国大陆不断地涌现,为高考效度问题的深入研究提供了重要学术支撑。高考研究的整体进展也拓宽了高考效度问题的研究空间,高考的效度已不再是一个单一的问题,而是以它为中心形成一个包括一系列相关问题的问题簇,因此,这些对高考效度问题来说较为间接的研究成果同样值得重视。

(一) 关于高考效度的文献

若以研究的着眼点为标准,研究高考效度的文献可以分为着眼于评价高考选才有效性和着眼于探讨高考成绩对大学学业成绩的影响两种类型,前者如北京师范大学心理系高考研究组的《改革高考,更加准确有效地选拔人才》^①,后者如尚鹤睿、邱颀曾的《医学生大学英语四级考试成绩影响因素分析》^②。前一种类型的文献数量远少于后一种类型的文献数量,说明多数研究者关注高考成绩对大学学业成绩的影响,目的是为大学教学服务,作者也多为各专业的大学教师,少数研究者关注高考自身的科学性,目的是为高考改革服务,作者多为心理与教育测量专业人士。这两类文献都在一定意义上表达了人们对高考效度状况的描述乃至判断,若按价值判断的结果又可以分为肯定高考的效度状况与否定高考的效度状况两种类型。刘春锋等学者在《医学生高考成绩与入学后在校期间学习成绩相关性研究》一文中认为高考总成绩与在校期间总成绩之间存在

^① 北京师范大学心理系高考研究组:《改革高考,更加准确有效地选拔人才》,载《教育研究》1985年第6期,第54~59页。

^② 尚鹤睿、邱颀曾:《医学生大学英语四级考试成绩影响因素分析》,载《广州医学院学报》1995年第6期,第65~66、62页。

明显的正相关关系^①，而王方根、王闽在《医学生入学成绩与毕业总成绩相关性研究》一文中认为两者的相关性很低^②。甚至同是数学专业，高晶、陆仲伟《高考成绩与大学学习成绩的回归分析》一文的研究结果是肯定的^③，而柴俊、陆竞、俞曼《高考数学分数高，大学数学学习成绩一定好吗》一文的研究结果是否定的^④；同为英语专业，许之所、张丽芳《高考英语试卷预测效度实证研究》一文的研究结果是肯定的^⑤，而周国辉《英语高考成绩预测效度》一文的研究结果是否定的^⑥。因此，在涉及效度状况描述时，我们有必要对高考的具体年份、高校类型、学科专业、数据的选取标准与处理口径、样本量、计算公式等因素加以考虑；在涉及高考效度的价值判断时，我们必须重视判断所依据的是什么标准，例如刘德峰、郭雅在《入学成绩与在校学习成绩的相关性及其影响因素》一文中以0.001水平作为判断标准^⑦，显然不太符合高考效度研究的实际。

文献所涉及的高考年份自1977年（如吴丽卿教授的《高等学校学生入学年龄问题初探》^⑧）持续到现在几乎不曾间断；涉及的高校类型既有中山大学、山东大学等研究型院校，如谢宗江、冼利青的《“3+X”与医学生在校成绩相关分析》和山东大学教务处的《电子计算机应用于大学生成绩分析》，也有天津财经学院、国际关系学院等应用型院校，如罗永泰、李小妹的《高考入学成绩对

① 刘春锋等：《医学生高考成绩与入学后在校期间学习成绩相关性研究》，载《洛阳医学专学报》1994年第4期，第235页。

② 王方根、王闽：《医学生入学成绩与毕业总成绩相关性研究》，载《江苏高教》1995年专辑，第61页。

③ 高晶、陆仲伟：《高考成绩与大学学习成绩的回归分析》，载《丹东师专学报》1994年第1期，第64页。

④ 柴俊、陆竞、俞曼：《高考数学分数高，大学数学学习成绩一定好吗》，载《数学教学》2003年第8期，封二。

⑤ 许之所、张丽芳：《高考英语试卷预测效度实证研究》，载《武汉理工大学学报》（社会科学版）2004年第2期，第249页。

⑥ 周国辉：《英语高考成绩预测效度》，载《淄博师专学报》1996年第2期，第95页。

⑦ 刘德峰、郭雅：《入学成绩与在校学习成绩的相关性及其影响因素》，载《高教探索》1994年第1期，第64页。

⑧ 吴丽卿：《高等学校学生入学年龄问题初探》，载《教育研究》1981年第12期，第23~27、72页。

后续课程影响的统计分析》和叶景山的《高考成绩、人格因素与英语四级成绩的关系》，以及杭州医学高等专科学校、丽水师范专科学校等实用型院校，如顾秀娟、邹立人、许亮文的《大专生高考成绩与在校学习成绩关系的分析》和卢巧焕的《对我校物理系学生的人校成绩与在校学习成绩的统计分析》^①；覆盖了绝大部分的学科门类，如医学、理学、文学、农学等等^②，其中关于医学类的成果最为丰富。这些研究大部分采用了皮尔逊积差相关法^③，还有一部分采用了斯皮尔曼等级相关法、多元回归和逐步回归法、典型相关分析法等各种研究方法^④。此外，雷新勇博士在《2004年高考（上海卷）地理科考试评价》一文中，从内容方面的证据和内部相关关系的证据入手探讨了高考的效度^⑤。另有一些研究虽不是直接研究高考效度，但宽泛地说，实质上也是在探讨高考为高

① 谢宗江、冼利青：《“3+X”与医学生在校成绩相关分析》，载《医学教育探索》2003年第4期，第46~48、55页。山东大学教务处：《电子计算机应用于大学生成绩分析》，载《教育研究》1983年第2期，第77~80页。罗永泰、李小妹：《高考入学成绩对后续课程影响的统计分析》，载《数理统计与管理》1996年第2期，第14~16页。叶景山：《高考成绩、人格因素与英语四级成绩的关系》，载《国际关系学院学报》2002年第2期，第40~43页。顾秀娟、邹立人、许亮文：《大专生高考成绩与在校学习成绩关系的分析》，载《中国高等医学教育》1996年第2期，第37~38页。卢巧焕：《对我校物理系学生的人校成绩与在校学习成绩的统计分析》，载《丽水师范专科学校学报》2004年第2期，第119~121、128页。

② 张靖等：《对医学生在校学习成绩变迁因素的初步探讨》，载《医学教育》1994年第6期，第17~19页。宋健：《高考成绩对大学主干课程成绩的影响》，载《集宁师专学报》1999年第4期，第30~32页。陈葆、陈艳辉：《城乡大学生英语基础与四级成绩的相关研究》，载《中国成人教育》2005年第6期，第101~102页。马骥等：《农科院校学生入学成绩、基本情况对大学成绩影响的实证分析》，载《高等农业教育》2003年第11期，第50~52页。

③ 钱钟、吴祖俭：《高考成绩与发展潜力的相关性研究》，载《江苏高教》2002年第3期，第39~41页。

④ 别雪君、李祖超：《大学成绩与高考成绩相关关系的统计分析》，载《建材高教理论与实践》1997年第4期，第49~52页。高晶、陆仲伟：《高考成绩与大学学习成绩的回归分析》，载《丹东师专学报》1994年第1期，第60~61、64页。何杭佳：《大学成绩与高考成绩的相关分析》，载《玉林师范学院学报》（自然科学版）2002年第3期，第133~135页。

⑤ 雷新勇：《2004年高考（上海卷）地理科考试评价》，载《考试研究》2005年第1期，第56~70页。

等学校选择合适生源有效性的问题,如宁静等学者在《高考成绩与大学成绩的相关性研究》一文中采用聚类分析的方法,得出高考成绩达到一定水平的同学在入学后的努力程度和学习态度是决定他们大学成绩的主要因素的结果^①。这些研究成果对我们全面、深刻地认识高考效度不无裨益。

近年来,也有数篇硕士论文把高考的效度问题纳入研究的视野。宋玉宏的论文《从高考成绩对后续课程的影响看高招录取工作的合理性》选取首都师范大学数学系1995~1998级学生数据,分析了高考总分、各科成绩与大学第一年总成绩的关系^②。肖婕的论文《实际数据的统计分析与建模》对中国科学技术大学1999级学生的高考成绩和大学一年级统考的数学及英语成绩进行了聚类分析和相关性研究^③。林秀莲的论文《我国研究型高校本科生学习过程规律研究——以厦门大学为个案》选取厦门大学1996~2004级除艺术类外来自福建、四川、浙江、湖北四个省份的四年制本科生为样本,分年度分科类进行相关分析,考察了高考成绩与大学学习成绩的相关系数^④。

综上所述,对高考效度研究的成果不可说数量太少,也不可说缺乏高质量的研究成果。然而,这些研究成果采用的研究方法不尽相同,却没有阐明依据;在效标选择上大多采用第一学年的学业成绩,也有研究采用大学学业总成绩、大学主干课程成绩、第一学期的学业成绩、几门主干课程成绩等,不一而足,同样缺乏一个明确的理由。有的研究选择一所或几所高校的几个专业同一年份入学的学生作为样本,有的研究选择一所高校的一个专业连续几年入学的学生作为样本,也有的研究选择几所高校的同一专业同一年份入学的学生作为样本,大部分研究是选择一所学校的一个专业同一年份入学的学生作为样本。撇开研究本身的科学性程度不论,由于研究口径的不统一与研究结论的多样性,普通民众、使用高考成绩的高校与从事相关研究的学者难以从这些研究中获取一个相对完整、相对确定的关于高考效度状况的描述与判断,决策者也难以依此而

① 宁静等:《高考成绩与大学成绩的相关性研究》,载《高等理科教育》2001年第3期,第46页。

② 宋玉宏:《从高考成绩对后续课程的影响看高招录取工作的合理性》,首都师范大学硕士学位论文,1999年。

③ 肖婕:《实际数据的统计分析与建模》,中国科学技术大学硕士学位论文,2001年。

④ 林秀莲:《我国研究型高校本科生学习过程规律研究——以厦门大学为个案》,厦门大学硕士学位论文,2006年。

确定高考改革的方向并选择合适的改革策略。

(二) 关于效度理论的文献

测验理论中关于效度的研究主要分为论文和专著两类,这种划分不仅代表了成果表达方式的不同,而且代表着研究成果处于不同的发展阶段。

测验理论中关于效度的研究论文集中发表在《教育研究》杂志上,时间多在 20 世纪 70 年代末到 80 年代中期。张述祖教授的《论教育测量的重要性和教育测量中的一些基本概念》(1979)^① 和邱焕元教授的《对教育考试八项指标的综合评判》(1986)^② 两篇论文较早对一些基本概念包括效度进行了简要的介绍。李伟明教授在《谈谈测验考试的效度》(1981)^③ 一文中,结合一些具体的实例对考试效度作出了有益的探讨。邹有华教授、曾桂兴教授的《高考命题标准化初探》(1983)^④ 和刘育民教授、高凌飙教授、张敏强教授的《从广东省高考标准化试验看高考改革的方向》(1988)^⑤ 两篇论文从命题的角度阐述了高考的效度问题。在高考效度的研究过程中,高考总分和大学学业成绩的计算方法是非常重要的,席宁华教授的《用标准分计算高考总分的运算和研究》(1984)^⑥ 和叶佩华教授、曾桂兴教授的《正确计算多科考试成绩的方法》(1983)^⑦ 两篇论文就此问题分别作出了富有成效的研究。信度是影响效度的重要因素,而评分者信度则是高考实践中较易对效度产生重要影响的一个信度类型,郑日昌教授、张厚粲教授的课题组对此问题进行了非常深入而严谨的研究,

① 张述祖:《论教育测量的重要性和教育测量中的一些基本概念》,载《教育研究》1979 年第 5 期,第 69~70 页。

② 邱焕元:《对教育考试八项指标的综合评判》,载《教育研究》1986 年第 12 期,第 55 页。

③ 李伟明:《谈谈测验考试的效度》,载《教育研究》1981 年第 10 期,第 88~91 页。

④ 邹有华、曾桂兴:《高考命题标准化初探》,载《教育研究》1983 年第 6 期,第 58~61 页。

⑤ 刘育民、高凌飙、张敏强:《从广东省高考标准化试验看高考改革的方向》,载《教育研究》1988 年第 7 期,第 37~40 页。

⑥ 席宁华:《用标准分计算高考总分的运算与研究》,载《教育研究》1984 年第 4 期,第 38~44 页。

⑦ 叶佩华、曾桂兴:《正确计算多科考试成绩的方法》,载《教育研究》1983 年第 3 期,第 55~58 页。

并形成论文《对高考评分客观性的调查分析》^①发表在《教育研究》(1985)上。虽然这一时期的研究成果数量不多,但是覆盖了高考效度研究的主要方面,更为重要的是研究质量很高。

效度是测验理论的重要概念。比较成熟的经典测验理论形成于20世纪50年代,稍后兴起的概化理论、项目反应理论因受计算机技术的限制直到20世纪七八十年代才较为成熟。在这相当长的时期内,我们接受的基本仍是传统意义上的效度概念,而在世界范围内,对效度的理解已经有了极大的进展。

20世纪90年代尤其是21世纪以来,随着一部部测验理论专著和译著的完成,国内理论界对效度的理解也有了明显的进展。以对本专题的影响而言择其要者予以简单罗列:叶佩华教授主译的《心理测验分数的统计理论》较为全面地介绍了回归、预测、复相关、偏相关等有关度量效度的术语^②;王孝玲教授的《教育测量》(修订版)对理论效度与操作效度作了有益的区分,并较为全面地介绍了度量效标关联效度的方法^③;张敏强教授的《教育测量学》对中间效标和最终效标的区分对度量高考效度的效标选择具有重要的启发意义^④;燕妮琴、谢小庆教授翻译,美国教育研究会、美国心理学协会、全美教育测量学会合著的1999年版《教育与心理测试标准》介绍了效度的凭证类型以及选择合适效度凭证类型的依据^⑤,而雷新勇博士的《大规模教育考试:命题与评价》较为全面地介绍了不同类型效度凭证的检验方法^⑥;张娜、杨艳苏、徐爱华翻译的《心理测验》对效度系数的解释大大拓展了人们对高考效度的理解^⑦,张厚

① 郑日昌、张厚粲:《对高考评分客观性的调查分析》,载《教育研究》1985年第2期,第26~31页。

② [美]罗德、诺维克著,叶佩华主译:《心理测验分数的统计理论》,福建教育出版社1992年版,第297~321页。

③ 王孝玲:《教育测量》(修订版),华东师范大学出版社2005年版,第65~76、93~94页。

④ 张敏强:《教育测量学》,人民教育出版社1998年版,第122页。

⑤ [美]美国教育研究会、美国心理学协会、全美教育测量学会著,燕妮琴、谢小庆译:《教育与心理测试标准》,沈阳出版社2003年版,第12~37页。

⑥ 雷新勇:《大规模教育考试:命题与评价》,华东师范大学出版社2006年版,第282~304页。

⑦ [美]凯温·R·墨菲、查尔斯·O·大卫夏弗著,张娜、杨艳苏、徐爱华译:《心理测验》,上海社会科学院出版社2006年版,第168~177页。

梁教授等翻译的《心理测量与评估》对增益效度的探讨和对泰勒-鲁塞尔表的介绍则进一步加深了对高考效度的理解^①；漆书青教授的《现代测量理论在考试中的应用》详细探讨了信度和范围限制对效度系数的影响，使人们认识到实有效度系数是小于测量所得效度系数的，而且这个差值是不能忽略不计的^②。

（三）与高考效度问题密切相关的研究

高考效度的研究在向内部拓展的同时也深受外部因素的影响。前者包括高考效度研究的方法、技术路线、结果的解释、提高的途径等，后者主要包括对公平、成本、与基础教育的关系等外部制约因素的分析。

在高考效度研究中，如前所述，王孝玲教授、雷新勇博士的著作均探讨了研究高考效度的一些基本方法。这些研究成果对初入此领域的研究者提供的是理论上的指导，即具体的教育测量学的指导，若要进行具体深入的研究，还需要掌握必要的统计学知识和相应的统计软件。对于前者，张厚粲教授、徐建平教授撰写的《现代心理与教育统计学》（2004）和凌云教授撰写的《考试统计学》（2002）是非常必要的参考文献；对于后者，米子川教授的《统计软件方法》（2002）、米红教授的《实用现代统计分析方法及 SPSS 应用》（2004）、胡咏梅教授的《教育统计学与 SPSS 软件应用》（2002）则较为有益。关于数据处理的技术路线，席宁华教授、叶佩华教授提出了切实可行的研究途径。关于效度系数的解释，前述《心理测验》一书有较为详尽的研究。研究高考的效标关联效度，效标的选择是核心。北京师范大学心理系高考研究组指出，人们一般采用大学一年级的学习成绩，大多数国家也是如此。然而由于我国大学一年级开设的多属公共课和基础课，不能很好地显示出专业的特点，因此第一年的学习成绩不足以作为选拔人才准确有效的效标。若改用高年级的学习成绩或大学各年级学习的平均成绩作效标，可能会使分析更有意义^③。卫子光等学者提出了效标选择的四个原则：（1）效标必须准确地体现高考的总目标；（2）效标必须是稳定、可靠的；（3）效标必须是客观的，避免偏见；（4）在保证有效性的

① [美] Lewis R. Aiken 著，张厚粲、黎坚译：《心理测量与评估》，北京师范大学出版社 2006 年版，第 100～107 页。

② 漆书青：《现代测量理论在考试中的应用》，华中师范大学出版社 2003 年版，第 85、423 页。

③ 北京师范大学心理系高考研究组：《改革高考，更加准确有效地选拔人才》，载《教育研究》1985 年第 6 期，第 57 页。

前提下,效标必须尽可能简单、可操作^①。尽管这些原则的提出具有明显的积极意义,但是作者在运用这些原则进行实例分析过程中却也有所出入,当然,这也充分反映了在具体的高考效度研究中效标选择的复杂性。至于提高高考效度的途径,大多数文献只是从理论上作一些简单的介绍,也有少量的文献从命题科学化^②、分卷考试^③等角度进行了相应的探讨。

诚然,许多有识之士逐渐认识到保证高考具有一定效度的重要性,但是若要真的落到实处,则至少应迈过公平、成本、对基础教育的影响这三关。笔者认为,迈过这三关的首要基础是理论的清晰化,一些学者对此问题作出了具有重要价值的研究。先看公平,刘海峰教授在谈到科举考试的公平问题时指出,“追求细节上、形式上的公平有可能偏离考试取才的根本意图”,“在舞弊不公与僵化刻板之间,考生宁愿公平竞争而接受死板的考试”,“这就容易驱使考试制度走向违反设科取才本意的死胡同”^④。刘虹教授认为,科举制度因为过度“防奸”、过度追求公平,最终导致了终极目的的迷失,给自身造成了戕害。公平是科举制度存在的坚厚的基石,然而对公平的过分热衷却又导致了科举制度最终的灭亡,这不能不引起人们的反思^⑤。两相对比,高考若要保持活力,较好地履行其职能,则必须把对公平的追求保持在一定的限度范围之内。再看成本,刘海峰教授认为,高考一方面要提高考试的信度、效度和区分度,准确地测验出应试者的实际水平,将优秀者选拔出来供高等学校挑选;另一方面要使考试本身做到高效、经济,能够使考试简便易行,省时、省事、省力。他还进一步指出,当经济条件还不够好时,高考改革要从两方面同时考虑;当经济大为发展,条件大为改善之后,则应更多地满足第一方面^⑥。可以说,刘海峰教授明确点出了不同局限条件

① 卫子光等:《化学高考预测效度初探》,载《化学教育》1995年第11期,第4页。

② 邹有华、曾桂兴:《高考命题标准化初探》,载《教育研究》1983年第6期,第60页。

③ 杨学为:《关于高考能力的问题》,载《教育研究》2005年第3期,第34页。

④ 刘海峰:《高考改革中的公平与效率问题》,载《教育研究》2002年第12期,第81页。

⑤ 刘虹:《科举作弊与科举制度功能的异化》,载《科举文化与科举学学术研讨会论文集》,中国科举博物馆2006年,第107页。

⑥ 刘海峰:《高考改革中的公平与效率问题》,载《教育研究》2002年第12期,第83页。