

外教社外语测试与教学丛书

语言测试 社会学

杨惠中 桂诗春 编

外教社外语测试与教学丛书

语言测试 社会学

杨惠中 桂诗春 编



图书在版编目(CIP)数据

语言测试社会学 / 杨惠中, 桂诗春编. —上海:

上海外语教育出版社, 2015

(外教社外语测试与教学丛书)

ISBN 978-7-5446-3937-8

I. 语… II. ①杨… ②桂… III. ①语言-测试-

社会语言学 IV. ①H09

中国版本图书馆 CIP 数据核字(2015)第 037752 号

语言测试社会学

出版发行: **上海外语教育出版社**

(上海外国语大学内) 邮编: 200083

电 话: 021-65425300 (总机)

电子邮箱: bookinfo@sflep.com.cn

网 址: <http://www.sflep.com.cn> <http://www.sflep.com>

责任编辑: 蔡一鸣

印 刷: 上海叶大印务发展有限公司

开 本: 787×1092 1/16 印张 21.25 字数 343 千字

版 次: 2015 年 4 月第 1 版 2015 年 4 月第 1 次印刷

印 数: 1 100 册

书 号: ISBN 978-7-5446-3937-8 / H · 1433

定 价: 48.00 元

本版图书如有印装质量问题,可向本社调换

序 言

桂诗春

《语言测试社会学》即将问世。这是从社会性思考的角度来探讨语言测试的理论和实践的一本论文集,作者都是语言测试界富有实际经验的专门人士。他们的共同感受是:测试(Testing and Examination)是一门科学,有其本身的科学体系;但是测试又不能就科学而论科学,它和教育息息相关,所以又称为教育测量学(Educational Measurement)。而且它和许多社会问题,如人才选拔、社会发展、经济建设等,也有密切关系,特别是一些社会的敏感问题,如发展不平衡、社会不公等,也往往会反映到考试和招生录取上面。在国外,考试的社会性(或称为“社会维度”),也是一个越来越受到关注的问题,因为其社会敏感度很高,牵动社会的方方面面。我国是“考试的故乡”,从科举制度的建立和发展,历代统治者无不给以极大的关注。在封建时代,最高的统治者往往就是科举的总主考。因此考试的社会性应该也是考试(特别是外部的公开考试)的一个主要内在属性。离开了它,考试就不复存在。

教育测量是在 20 世纪初的心理测量的基础上发展起来的。在达尔文主义的启发下,人们开始探索和研究人类行为的共同性和个别差异。但在科举时代的因袭和影响下,中国考试并没有与时俱进。在我国的教育学和心理学也都设有这样的课程,要结合我国国情,建立有中国特色的考试体系,还有待努力。也许是因为我国是一个幅员广阔、人口众多、发展不平衡的国家,保证社会公平成为高考和招生录取的核心关注点。建立考试体系亟需从我国实际出发,从科学的角度进行顶层的全面设计。本书的作者们从事测试(特别是语言测试)多年,也力图结合我国实际,在考试社会化方面进行一些探索,没有经验,也有教训,很值得我们参考。

但是一个考试体系的建立需要几方面人员协同努力,最主要的是两个方面:一是考试开发者,另一就是考试的决策者或管理者(他们往往也就是考试用户的代表)。他们都需要懂得教育测量学,但程度可以不一样。前者需要的是较强的专业性,而后者则需要一般的专业知识,做到能统管大局。满足这两部分人的需要,也是本书编写的目的,应该让考试的决策者或管理者了解教育测量的根本原则和考试开发者的所思所为,这是因为:(1)考试的开发和决策管理并无根本的矛盾,只有共同的目标:建立考试的公正和公平;(2)两者都致力于考试招生制度的改革,而且都认识到既不能抱残守缺,也不能全盘照搬,只能结合我国实际革故鼎新;(3)教育测量学是一门科学,它建立了一套体系和原则,但也不是固定不变的,总会有所前进和发展。建立符合我国实际的教育考试体系和原则,必须在科学上作不懈的努力,不能采取实用主义的方法,以为只提出一些权宜措施就能奏效。

顾名思义,这本论文集并非全面介绍和讨论教育测量学的,而只是就语言这个学科在实施教育测量学的理论和原则中所遇到的问题,提供一些实践中的思考。文章的写作时间有所不同,基本上都是在中共中央十八届三中全会和国务院关于深化考试招生的实施意见以前执笔的。我们很高兴地看到中央领导对考试招生工作的重视。国务院的总体评价是“改革开放30多年来,我国考试招生制度不断改进完善,初步形成了相对完整的考试招生体系,为学生成长、国家选才、社会公平作出了历史性贡献,对提高教育质量、提升国民素质、促进社会纵向流动、服务国家现代化建设发挥了不可替代的重要作用。这一制度总体上符合国情,权威性、公平性社会认可,但也存在一些社会反映强烈的问题,主要是唯分数论影响学生全面发展,一考定终身使学生学习负担过重,区域、城乡入学机会存在差距,中小学择校现象较为突出,加分造假、违规招生现象时有发生。”在所列举的社会反映强烈的问题里,我们不难看出考试社会性是一个亟待解决的核心问题。正因为它的解决牵涉到社会的方方面面,所以在具体实施上采取试点先行的慎重举措。

感谢参与本书写作的外中专家和教授,希望他们的努力能够为我国考试招生改革添砖加瓦!

前言

杨惠中

考试制度是社会制度的一个重要组成部分,对保证社会公平和建立和谐社会起到了重要的作用。大规模高风险考试常常会影响考生的未来,具有极强的社会性,考试社会性问题正是当前国际语言测试界的研究重点之一。本书结合我国语言测试实践,对我国社会和教育环境下考试社会性的方方面面进行探讨,希望引起我国教育界和社会各方的讨论,推动语言测试学科的发展,推动教育考试学科的发展,使考试结果能更好地服务于教学和人才培养工作。

本书共收 12 篇论文,分为三个部分。

第一部分包括第 1 至第 3 共三篇文章,从宏观上讨论考试的一般问题,包括考试的起源、考试的社会性、考试的质量保证、考试结果的正确使用等等。

桂诗春的“从社会视角看中国考试的过去和现在”是本书的重头文章。文章从历史发展的角度对科举考试的形成、作用、衰亡进行了深刻而全面的论述。科举的兴起是为了满足各个朝代统治阶级亟需网罗人才来稳固其政权的政治需要。科举把考试作为一种制度,在“取士”方面第一次做到了“三公”(公开、公正、公平),这对现代各国文官制度的形成也产生了重要影响。科举的盛衰和中国的封建制度紧密地联系在一起,科举考试在内容与形式上的僵化,未能及时跟上现代心理测量学的发展,使它最终与封建制度一起走向了衰亡。我们应该持唯物历史观来对待科举,扬其精华,弃其糟粕。科举兴衰的事实充分说明考试是为政治服务的,今天,我们必须与时俱进,实现中国考试的现代化,首先是确立考试的专业

化地位,科学地、公平地、准确地测量出被测量的人的学识、能力和水平。但是高风险考试不但牵涉考试内部的问题,还牵涉到考试外部的问题,因此必须从伦理上考虑考试对考生、教学过程和其他社会环境所产生的种种影响,对考试做更多的社会学思考,这样才能厘清问题的根本,对症下药。桂诗春教授的文章值得从事教育与教育考试研究的人们仔细研读、深入思考。

杨惠中与桂诗春的“语言测试的社会学思考”最初发表于2007年,在国内比较早地提出了从社会学的角度对考试进行思考的问题,文章对语言测试各方利害关系人的职责进行了讨论,所提到的一些观点仍有现实意义。

杨惠中的“有效测试、有效教学、有效使用”一文提出,从系统论的角度看,测试、教学和测试结果的使用构成一个系统,要发挥系统的作用,三者必须协调。实际上三者是不同的过程,虽然相互关系密切,但毕竟不是同一回事,应当区分开来,各司其职,各尽其责。只有各方正确处理好各自的关系,做到有效测试、有效教学、有效使用,才有可能找到正确的解决之道,取得整个系统的最佳效果。

第二部分包括第4至第8共五篇论文,从技术上讨论如何保证考试的专业学术质量。

谢小庆的“教育与心理测试标准”全面讨论了教育测试应当达到的质量标准。事实上,考试提供的是一种评价标准,但考试本身也应当有评价标准。大规模教育考试从设计开发、命题、施考到成绩发放包含诸多环节,每一个环节的失误,都会对考试的结果产生不可弥补的影响,必须为每一个环节制定相应的专业标准,以规范教育考试行为。我国有各种国家工业标准,工业产品达不到国家规定的标准是不准出厂上市的,可是对于人的测量,至今尚无标准,看来我们应当参照国际上的已有经验,结合我国实际,尽快制定我国自己的教育与心理测量标准,以造福千百万学子。

罗冠中的“大规模教育考试的质量保证及设计流程”介绍了我国香港教育界近年来的一项重要改革措施。香港考试及评核局经过十多年的艰苦研究和试验,从2012年起正式实施“香港中学文凭考试+校本评核”的做法,取代实行多年的中学高级程度会考(高考),文凭考试是公开的统一考试,校本评核则由在校任课教师完成,评核的分数以15%的权重计入学

生的香港中学文凭考试成绩。校本评核把教师引入了考试过程,由于任课教师最了解自己的学生,所以由他们进行校本评核是最适当的,校本评核的做法把教和学与评测结合起来,更全面地评价学生的表现。从教师方面来说,直接介入测试,可以影响教师的教学过程并改变他们对教学的认识。的确,香港考试及评核局的经验证明,这样做不但有可能把终结性考试与形成性考试有机地结合起来,把测试结果直接反馈到教学中,而且有可能彻底解决应试教学的老大难问题,改变教育生态。这确实是一项极好的教育与考试改革实验,非常值得我们学习和借鉴。但要取得成功,除了大量地、系统地运用先进的现代教育测量理论与技术,实现方法上、技术上的全面转型外,还有几个先决条件:首先在技术上要制定可靠的各科目能力等级量表;校本评核要进行等值处理,以确保考试的信度与效度,保证社会公正;其次,由于校本评核是以教师为主体进行的,因此教师培训极其重要;再次,必须取得社会各界及各方考试利害关系人的理解与支持;这三个条件缺一不可。香港地区的经验值得我们深入思考和学习。

金艳与范劲松的“语言测试中的道德规范”一文,对国际语言测试道德规范的制定和相关研究进行了回顾,并基于国际语言测试协会(ILTA)制定并颁布实施的《ILTA 道德规范》,探讨我国语言测试工作者所应遵循的职业伦理以及所需承担的专业职责和社会责任,内容包括语言测试工作者的职业伦理范畴、语言测试工作者的专业职责以及语言测试工作者的社会责任等等。鉴于大规模考试的成绩往往被用来作为高风险决策的依据,语言测试工作者除了兢兢业业做好自己的工作外,唯有提高职业道德和专业化的意识,遵循语言测试专业的道德标准进行自律。

杨惠中、朱正才与方绪军的“尽快制定统一的中国语言能力等级量表”一文,深入讨论了我国语言测试中的另一个重要问题。我国是英语教学大国,又是考试大国,各种英语考试数量庞大,但是迄今为止,我国尚没有统一的语言能力等级量表,语言教学和语言测试没有共同的参照框架,甚至连等级的划分也相当混乱,有的语言测试项目分两个级,有的分三个级、四个级、五个级甚至九个级,五花八门,非常混乱。语言能力等级量表就是描述或评价人的语言能力水平的工具,是语言教学和语言测试的基本参照框架。建立我国统一的“语言能力等级量表”有利于组织教学、开发教材、评测教学结果;有利于提高各级各类外语教学的效率;也有利于设计开发各种科学的语言测试,使考试的分数获得统一的可解释性。建立统一的“中国语言能力等级量表”和制定我国“教育与心理测量国家标

准”是我国教育考试改革中的基础工程,这两项基础工程的完成将起到“量同衡、车同轨”的作用,将为建立我国科学的教育考试制度提供强大的技术支持。

Kunnan & Zhang 的“教师如何成为负责任的语言评测者?”一文讨论的是对语言教师普及语言测试知识的问题,文章也用了大量篇幅阐述语言测试专业的道德与行为标准,尤其是公平公正的理念。公平原则近年来受到语言测试界愈来愈多的重视,分为公平和公正两个方面,Kunnan 教授等深入探讨了公平原则的哲学与伦理学基础,并且指出“公平原则先于公正原则,只有有了公平,才谈得到公正。”考试必须促进社会正义,对社会有正面后效。对一项考试来说,只有首先保证测量公平,才能实现教育公正,才能对社会产生正面后效。这些是专业语言测试工作者必须遵循的原则。Kunnan 等认为一般教师也应接受测试理论方面的培训,内容包括理解语言测试开发与设计原理、理解信度与效度基本概念、理解命题原则和试题开发周期,还应当包括统计学基本知识,这样才能使教师更好地进行语言评测,更好地理解语言测试的本质,更好地利用公开考试的数据所提供的反馈信息,更好地开发自己的形成性考试,让语言测试更好地为语言教学服务。

第三部分包括第 9 至第 12 共四篇论文,主要讨论语言测试与语言教学的关系。

亓鲁霞的论文“高风险考试双重功能的冲突”以高考为例,通过广泛抽样采集数据,探讨了高风险考试的促学作用与选拔作用的矛盾。为了扭转中学英语教学重知识、轻能力的倾向,高考英语的设计者把考查的重点放在语言运用上,试图借助考试的影响力,将教学的重心转向培养学生的英语运用能力。但是高考的选拔作用及由此引发的应试教学淡化了考试的促学作用。这里讲的选拔,实际上就是竞争,即为了获得宝贵的高校入学机会进行激烈的竞争,作者认为在这种情况下想利用高风险考试来促进教学改革的美好愿望很容易落空。这是关于大规模考试反拨效应的一项实证研究,规模大、数据详实、分析深入,有说服力。看来解决应试教学的问题还需要做更多社会学层面的深入思考和探索。

Ari Huhta 的“芬兰评测文化——传统与创新的混合体”一文,介绍了芬兰的教育现状,对我们很有启发。芬兰人口约 550 万,实行九年制义务教育。初中毕业后,大约一半学生进入称为文理高级中学的学术型高中,另

一半学生或进入职业学校或直接参加工作。一般来说,进入文理高级中学就是为将来进入大学做准备,但是,据统计,2012年,文理高级中学毕业生中当年就进入大学学习的只有33.4%,而有高达62.5%的学生当年并没有马上进入大学接受高等教育,好像并不存在激烈的大学入学竞争。不但如此,义务教育基本上是就近入学,初中毕业的成绩将决定能不能进和进哪一所文理高级中学,按说这将决定学生未来接受高等教育的机会,因此初中毕业考试理应是高风险考试,但是按芬兰传统,集中统一的考试很少,而初中的毕业成绩主要由教师决定,因为“教师受到家长和教育当局的充分信任。”芬兰唯一的高风险考试是文理高级中学学生参加的大学入学考试,但是按作者的说法“芬兰从来不是一个为了应付……考试而应试教育严重的国家。”由此看来,解决应试教育的问题应从社会经济文化的角度去考察一个国家的教育生态,仅仅想从改革考试着手解决应试教育的问题恐怕未必能够见效。作者在文章中还介绍了芬兰外语教学情况和语言测试现状,总的趋势是不断提高语言测试的专业化程度。

方绪军的“语言测试和教学中试题与习题辨异”指出,即使试题与习题在形式上有相似之处,两者是有根本区别的,这是因为测试过程和教学过程尤其是学习过程是不同的过程,测试关注的是如何准确测得考生的语言能力,而不关心考生的语言能力如何形成。文章从题目的性质、目标与用途、编制过程、题目涉及的内容、形式、题目质量分析与评价的指标等方面讨论和分析语言测试和教学中的试题和习题的区别,懂得这些区别可以帮助教师和学生自觉摒弃以考代教、以考代练、以考代学的做法,因为这严重背离了语言教学要循序渐进培养学生语言能力的根本途径和目标,教师应自觉摒弃应试教学,还课堂教学以本来面貌。

杨治中的“应试教学的社会分析”一文指出,所谓的应试教学就是把宝贵的课堂教学时间用来搞题海战术,安排学生做大量的模拟试题。实际的结果是只有“应试”,没有“教学”,既不能提高学生考试成绩,又毁了教学,贻害青年学生。所谓模拟试题完全是教育领域似是而非的伪劣产品,不可能用来测量考生的知识和能力,而用来“教学”或作考前的操练也只会浪费宝贵的课堂教学时间,是完全无效的。编制模拟试题集的人既不懂测试,也不懂教学,他们不惜通过坑害学生来进行牟利,从语言测试社会学的角度看,这是一个社会问题,不是学术问题,应当由有关部门采取有力措施,加以解决。

语言测试社会学是一个比较新的研究领域。本书所收文章,原文是中文的,都配了英文的详细摘要;原文是英文的,配了中文的详细摘要。英文部分集中放在书的后半部分。文章序号中文部分用阿拉伯数字表示,英文部分用罗马数字表示,一一对应,以便前后相互参照,这样做是希望不懂中文的研究者也能了解我国在这个重要的新兴研究领域里做了些什么。

还有一点需要说明,书名谈的是语言测试,实际上讨论的是教育考试的一般问题。另外,考试与测试两个术语,在使用上不做严格区别,就像英文中的 examination 和 test 也是一样的情况。在中文里,作为研究领域,则用“测试”,如“语言测试”、“测试的效度”等;作为测试项目,则一般用“考试”,如“托福考试”、“大学英语四、六级考试”等。

本书所收 12 篇论文从社会学视角对考试问题进行了深入讨论,鉴于考试的重要性,讨论这些问题及时而必要,希望能引起国内语言测试界与教育界的重视和讨论,但是这些文章未必能涵盖考试社会学的所有方面,论点也未必完全正确,有错误的地方,敬请读者批评指正。

另外,本书从最初构思策划起,我们曾经有机会与国内外多位学者,尤其是 Alderson 教授,进行过广泛的讨论与深入的交流,得到他们许多帮助,我们也听取了他们不少有益的建议,无法一一指出,在这里谨致谢忱。

目 录

序言	桂诗春	I
前言	杨惠中	III
1. 桂诗春	从社会视角看中国考试的过去和现在	1
2. 杨惠中 桂诗春	语言测试的社会学思考	29
3. 杨惠中	有效测试、有效教学、有效使用	40
4. 谢小庆	教育与心理测量标准	79
5. 罗冠中	大规模教育考试的质量保证及设计流程	99
6. 金 艳 范劲松	语言测试中的道德规范——《ILTA 道德规范》 对我国语言测试的启示	115
7. 杨惠中 朱正才 方绪军	尽快制定统一的中国语言能力等级量表	132
8. Antony John Kun- nan & Zhang Limei	教师如何成为负责的语言评测者? (详细摘要)	143
9. 亓鲁霞	高风险考试双重功能的冲突(详细摘要)	150
10. Ari Huhta	芬兰评测文化——传统与创新的混合体(详细摘要)	152
11. 方绪军	语言测试和教学中试题与习题辨异	159
12. 杨治中	应试教学的社会分析	180

I.	<i>Gui Shichun</i>	A Social Perspective of Chinese Exam: Past and Present (Synopsis)	183
II.	<i>Yang Huizhong & Gui Shichun</i>	The Sociology of Language Testing (Synopsis)	190
III.	<i>Yang Huizhong</i>	Valid Testing, Effective Teaching, and Valid Test Use (Synopsis)	193
IV.	<i>Xie Xiaoqing</i>	Standards for Educational and Psychological Testing (Synopsis)	198
V.	<i>Luo Guanzhong</i>	The Design and Quality Assurance of Large-Scale Educational Examinations: Using HKDSE as an Example (Synopsis)	200
VI.	<i>Jin Yan & Fan Jinsong</i>	Code of Ethics in Language Testing: What Can Language Testing in China Learn from the ILTA Code of Ethics? (Synopsis)	203
VII.	<i>Yang Huizhong, Zhu Zhengcai & Fang Xujun</i>	Working towards a Unified National English Language Proficiency Scale (Synopsis)	208
VIII.	<i>Antony John Kun-nan & Zhang Limei</i>	How Can Teachers Be Responsible Language Assessors?	211
IX.	<i>Qi Luxia</i>	Stakeholders' Conflicting Aims Undermine the Wash-back Function of a High-Stakes Test	232
X.	<i>Ari Huhta</i>	Assessment Culture in Finland — a Mixture of Traditions and Innovations	266
XI.	<i>Fang Xujun</i>	The Difference between Items in Language Testing and Exercise Items in Teaching and Learning (Synopsis)	302
XII.	<i>Yang Zhizhong</i>	Test-taking Oriented Teaching — a Social Analysis (Synopsis)	305
	附录一	语言测试常用术语中英文对照表	307
	附录二	专有名词与缩写中英文对照表	321

1. 从社会视角看中国考试的过去和现在^①

桂诗春(广东外语外贸大学)

(一) 前言

“中国是考试的故乡”,这是国内外考试界的共识。中国科举考试从隋末(607年)开始,到晚清光绪31年(1905年)宣布取缔,经历过1,300余年的盛衰,亟须从社会变化、发展和更迭的角度来细加观察,“以史为鉴”,从中获益。近年来,对考试的社会性思考十分活跃,说法很多:有的冠以“批判性”,如 Shohamy (1998; 2001; 2004; 2007) 提出批判性语言测试(Critical Language Testing);有的直接点出其政治问题,如 Alderson (2009) 主编的论文集《语言教育的政治》(*The Politics of Language Education*);有的则把它称为“社会维度”,如 McNamara & Roever (2006);有的从心理和教育测量的角度,把它归为效度的一个问题,如 Messick (1989)。这些思考也引起我们对自以为荣的中国古代一大发明——科举考试作重新认识。

根据 Fulcher (2009) 观察,如果追本溯源,从政治哲学角度看,考试的使用和产生实基于不同的理念。一种理念是 Plato 在《理想国》里所提出的“国家大于个人”,个人主义把国家置于解体的危险。个人“必须把其全部精力专注于其最特别合适的工作,国家与个人的完整和团结才能得以保存。”Plato 明白地指出测试在这种社会里的任务就是选好领袖。和他相对立的观点是 Thucydides 在《伯罗奔尼撒战争史·第二卷》里从 Pericles 在葬礼上的演说所提出的雅典民主原则:“在法律上人人平等”。个人主义可以加强民主,因为每个人都可以按照自己的能力无限制地做出贡献。Fulcher 认为,公元前4世纪政治哲学都是关于国家与个人的基本平衡,我

① 本文初稿写于2013年4月,2015年2月19日完成定稿。

们可以根据一种政治哲学在这两者中倾斜来决定它喜欢哪样的政府和哪样的社会,从而解释(或预测)测试的使用。

由此看来,从考试所依据的政治哲学以及其使用的角度看,考试实际上有两种不同的传统:一种是适用于中央集权制度的传统,Fulcher称之为“集体主义的”;可以中国的科举考试为代表,它蔓延上千年,成为封建主义制度的一根主要支柱。历史学家曾经就中国的封建主义为什么比西方长得多,进行过许多讨论,原因复杂,这也许和科举考试制度不无关系。李约瑟(1990)在他的《中国科学技术史·第一卷》的《序言》里提出过这样的问题:“欧洲在16世纪以后就诞生近代科学,这种科学已被证明是形成近代世界次序的基本因素之一,而中国文明却未能在亚洲产生与此相似的近代科学,其阻碍因素是什么?”郭齐家(1997:221)对这个问题的回应是,“这种原因当然要归于中国的小农业与家庭相结合的自然经济的难以打破;归结于建立在这种经济结构之上的封建专制政治的强固有力,封建典章制度的完备严密;此外,与社会经济相隔离、对科学技术采取粗暴贬斥态度的古代科举考试的长期实行,恐怕也是一个原因。”在封建社会里,科举把教育、考试和做官融为一体,统一于一个共同思想基础:从汉代董仲舒的“罢黜百家,独尊儒法”开始,科举考试内容都离不开四书五经。这种正统文化成为维系封建王朝的重要思想支柱。明代以后,八股文泛滥,更成为禁锢思想、阻碍科技发展的工具。另外一种传统是适用于地方分权制度的传统,Fulcher称之为“个人主义的”,那就是西方在经历过产业革命以后在19世纪建立的心理测量的传统。1863年英国 Francis Galton 受到其表兄弟达尔文的《物种起源》(1859)的影响,开始从遗传学的角度来研究人类行为的共同性和个别差异,1883年发表了著作《人类能力及其发展的探索》;1904年法国人 Binet 开始摸索建立智力测验。1916年,Louis Terman 把它引进到美国。心理测试就开始和教育测试捆绑在一起,由美国三大机构(美国教育研究协会、美国心理学协会、全美教育测量理事会)制定的《教育与心理测量标准》(简称《标准》,后同)从1954年到1999年前后改了六版。美国的大学入学考试 SAT 就是这种理念的体现,SAT 考的是学能(Scholastic Aptitude),而并非考生在中学学习过的知识(但考生需要提供其中学学习成绩作为录取的参考)。SAT 包括三个部分:写作(语法、用法和用词)、数学(数字运作、代数及其函数、几何、统计、概率和数据分析)和批判性阅读(词汇、批判性阅读、句子水平阅读)。

其实,这两种传统从历史上看是不同文明的产物,但从现实上却应该

做到从具体情况出发,因时制宜,不能死守一种传统,就如同有了很多城市并不一定大家都要回到乡村去居住一样。Fulcher 的说法当然不能简单地把考试传统和政治哲学等同起来,它只是受到当时统治思想的影响,因为一个考试有其诸多复杂的外部社会因素和内部心理因素需要考虑,而政治哲学无非是一些政治制度下所滋生的一些理念。在实施的层面上,考试和政治体制并无必然联系,就等于目前全世界有许多国家和地区具有不同政治体制,但它们都必须实行统一度量衡标准一样^①。Spolsky (1995: 356-357) 曾经用 **可行性**(feasibility)、**可用性**(usability) 和 **可靠性**(reliability) 几个因素来比较传统考试和心理测量考试:传统的考试强调前两个因素,对可靠性则不大注重,只要考试“感觉到公正”就可以。其考官主要受过人文科学训练,继承怀疑主义传统和接受多元主义(没有唯一的正确答案),他们由于制度需要或个人偏好,倾向于精英主义,相信考试分数是在人文主义的多元方式下取得的,绝对确定和准确;心理测量则强调技术发展可靠性,即信度(如客观性考试),注意选择受过训练的考官(包括计算机技术的应用),以求达到一致性,并追求人类能力测量绝对准确;但却容易忽略考试效度。这两种考试的对立也不无好处,可以互相补充。但是他认为 **还有一种可称之为“后现代”考试,它认为上述三个因素同样重要**。第一次世界大战以来,心理测量的方法在美国得到很大发展,且很快推广到全球,使我们注意力从更加严重的社会(或反社会)动机和影响中转移开来,但是“树欲静而风不息”,测试从一开始就是“被利用为实施控制和权力的方法——一种选择、驱使和惩罚的方法”。后现代考试从伦理上 **考虑考试在考生身上、在教学过程和其他社会环境中所产生的不良影响**,对考试做更多的社会学思考。Kunnan (2005) 建议要在更广阔背景下对测试作出对社会有利还是有害的估量,这个背景来自过去 20 年来的思考和研究,包括测试在一个社区运作的政治、经济、教育、社会、文化、技术和基础设施、法律和伦理环境。

① 杨学为在《中国考试改革研究》一书第 92 页上说:“有的同志说,统一考试是高度集中的计划体制的体现。恐怕不能这么说。前苏联集中的程度比我们还高,就从来没有任何全国统考,东欧也一样。他们都是学校单独招生考试。相反,很多市场经济比较发达的国家,倒是越来越多地实行了‘统一考试’,比如美国,是资本主义中所谓最自由的一种典型,它的高考,全国是两个机构实施,西部一个,东部一个,特殊的学校自己单独考试,绝大多数学校都使用这两个考试机构的成绩。……所以不能认为统一考试是高度集中的计划体制的体现,现在要改革高度集中的计划体制了,就应当取消统一考试。这种说法,在逻辑上是不能成立的,也没有事实根据。”

(二) 制度化科举——盛衰、遗产和因袭

我们在讨论中国考试制度时,往往只看到它成为制度化科举考试的1,300年,其实早在西周开始就已经有了其面向儿童和成人的教育,13岁开始就要学习“礼”、“乐”、“射”、“御”,20岁成人后进入大学(“辟雍”或“泮宫”),还要学习“书”、“数”,并建立起一套分年考试制度,七年结束,考试及格者为“小成”,九年结束,考试及格者为“大成”。只不过这个九年制的大学在西周并没有真正实行,而只是一种设想。如此说来,中国的考试可以推前到西周,总共有3,000多年,见杨学为等(1992)的《序言》。美国Dubois(1966)也采用了这个说法。

那么科举考试制度建立前的考试又是以什么形式出现的呢?基本上是以荐举为主,考试为辅,这是氏族社会以来残余思想的反映。传说里的氏族首领尧、舜、禹都是按照“天下为公,选贤与能,讲信修睦”(《礼记·礼运》)被推选出来的。西周教育只限于贵族子弟,在王畿内实行“乡里选举”,在王畿外则是“诸侯贡士”。两汉实行察举制、魏晋南北朝改为九品中正制,也都是在举荐的基础上进行考核。隋唐以后,一直到清,都在实行科举,成为中国古代考试制度的一个亮点。科举的盛衰和中国封建制度的盛衰紧密地联系在一起,对科举的记录和讨论已有不少,这里不再赘述。

首先,我们认为**科举是历史的产物,它经历过兴旺发达和衰亡的历史阶段;我们应该持唯物历史观,既不能把它神圣化,也不能把它妖魔化。**在它处在上升期发展起来的一些优秀传统应该发扬,而在它衰亡期的许多糟粕应该扬弃。在新时代里对它的历史局限更应该结合我国的实际,予以认真思考和解决。邓嗣禹(1936:122)在介绍唐代科举时就说过,“唐代科举,得人之盛,非后世所能及;晚年绩弊,亦开后世之先河,要而论之,得失盖参半焉。”

1. 那么从历史角度看,科举比它以前的考试制度有些什么进步呢?在科举以前实行的是魏晋南北朝的九品中正制,这种制度主要所选拔的多为豪门世族子弟,目的在于支撑门阀统治,选官权力由朝廷高官的本籍人掌控,为世家大族霸据中正官开辟道路,出现了所谓“上品无寒门,下品无世族”的现象。邓嗣禹(1936:8)认为,隋唐王朝(一般认为“科举肇基于隋确定于唐”)统治阶级在开始时也试行过九品中正制以笼络各地豪强势力,但是后来他们发现这种选官制度不利于中央集权。