DOI: 10.3785/j.issn.1008-942X.2013.04.161

语言测试研究的新视角：批判语言测试

何莲珍　吕洲洋
（浙江大学 外国语言文化与国际交流学院，浙江 杭州 310058）

[摘　要] 长期以来，语言测试工作者研究的核心是如何提高测试的信度与效度，以客观、精确地测量考生的语言能力。语言测试尤其是大规模高风险语言测试带来的社会影响，很难由改良测试本身来解决。批判语言测试（Critical Language Testing）的提出把语言测试置于更广阔的社会政治语境中，对测试后效（consequences）的反思不再局限于反拨效应（washback），而是重新审视测试中渗透的权力与控制，测试所具有的社会价值、政治功能等。

[关键词] 语言测试使用；批判语言测试；社会政治语境；考生语言能力；替代性评估方式

A New Perspective of Language Testing Research: Critical Language Testing

He Lianzhen　Lü Zhouyang
(*School of International Studies, Zhejiang University, Hangzhou* 310058, *China*)

Abstract: For a long time, the main focus of research conducted by language testers and language testing researchers has been on improving reliability and ensuring validity of tests in order to assess test-takers' language proficiency objectively and accurately. With their widespread use, large-scale and high-stakes language tests have exerted a great influence on test-takers' lives as well as the lives of other stakeholders, which has drawn more attention of the public to fairness and social consequences of language tests. Language testers are fully aware that improving tests can only solve the problems concerning the internal quality of tests, without providing effective solutions to those embedded in social and political contexts, such as minimizing misuse or abuse of tests, enhancing fairness of test administration and use, etc. Therefore, the field of language testing has shifted its focus to the issues of test use and impact, and efforts have been made in recent years to look at the social dimensions of language testing and language education. The research of social dimensions of language testing has been nourished by the development of validity theory. The first section of the article reviews the development of validity theory of language testing and the background of critical language testing theory. Elana Shohamy advanced the theory of Critical Language Testing (CLT) in the plenary talk at the annual conference of the

[收稿日期] 2013-04-16　　[本刊网址·在线杂志] http://www.journals.zju.edu.cn/soc
[在线优先出版日期] 2013-07-16
[基金项目] 国家社会科学基金项目(10BYY092)

[作者简介] 1. 何莲珍，女，浙江大学外国语言文化与国际交流学院教授，博士生导师，主要从事语言测试、外语教学及语料库语言学等方面的研究；2. 吕洲洋，女，浙江大学外国语言文化与国际交流学院博士研究生，主要从事语言测试研究。

American Association for Applied Linguistics in 1997. She pointed out that "The act of language testing is not neutral. Rather, it is a product and agent of cultural, social, political, educational and ideological agendas that shape the lives of individual participants, teachers and learners." The theory of critical language testing illuminates a broader social and political context of language tests. It reconsiders the consequences of tests which are not confined to washback, in terms of effects on instruction, learning and educational advancement. Instead, it reexamines the exercise of consensual power and control of tests over individuals and educational systems. The second section of the article introduces the main idea of CLT in five aspects in detail, including emergence of power of language tests, transformation of language tests from measuring tools to powerful tools, fairness of language tests, the alternative assessment of language tests and social responsibilities of language testers. The third section suggests the implications of CLT for language testing practice and foreign language education in China. First, efforts should be made to balance Chinese learning and foreign language learning while attaching importance to the latter. Students' awareness of learning Chinese should be enhanced in order to preserve, promote and develop Chinese culture. Second, the critical perspective should be taken to examine the use of tests in wider social context and to guard against the misuse and abuse of tests. Third, attempts should be made to develop alternative assessment aiming at reducing the risk of some large-scale tests as the only evidence for decision-making. As one of the various perspectives reflecting language testing, CLT broadens the research field from psychometrics and statistics to sociology and signifies a paradigm shift in language testing research. Rather than limiting its concerns to technical sophistication and efficiency, CLT shifts its focus to social discourse and social dimensions of language testing. Reflections on how to apply CLT to language testing research and practice and further develop the theory will prove a worthwhile endeavor.

Key words: language testing use; Critical Language Testing; sociopolitical context; test-takers' language proficiency; alternative assessment

语言测试作为一种社会实践具有悠久的历史。长期以来，语言测试专家将研究焦点集中于语言测试的技术话语，沿袭传统心理计量学方法，致力于研究如何采用并开发新的测量手段以提高语言测试的信度和效度。近年来，效度理论的发展及效度与公平性关系的讨论体现了语言测试界对语言测试使用及其社会影响的关注，对语言测试及语言教育的社会性的思考与探索。受后现代主义批判理论的影响，批判语言测试视角着重审视了语言测试中渗透的隐性权力，考量了语言测试在社会语境中扮演的控制者、监管者、支配者的角色，探讨了语言测试社会功能的演变对测试利益相关者(stakeholders)产生的影响。本文将在回顾批判语言测试这一概念提出的背景的基础上，着重介绍批判语言测试的研究领域，阐释其主要观点及核心思想，并讨论批判语言测试对我国语言测试实践及外语教育的启示。

一、批判语言测试的研究背景

近年来，语言测试工作者的研究重点逐步从提高语言测试信度和进行效度验证方面转向对语言测试社会影响的关注。语言测试社会维度的研究，即语言测试在社会中的使用情况以及产生的

社会影响和后果,是在效度框架的发展过程中逐步丰富的。Cronbach 和 Meehl 提出用构念效度替代效标关联效度,强调效度验证的对象是基于受试所得分数所做出的推论和解释,而不是分数本身[1]283。Cronbach 在效度验证的论著中并未强调分数解释中的社会因素,这与他在有关项目评价的研究中对政治立场与价值观影响的考量形成鲜明对比①。后期受项目评价研究的影响,Cronbach 开始意识到评估的社会语境的重要性,试图将语言测试的社会维度理论化[2]4。Messick 在 1989 年提出了具有突破意义的整体效度观,把效度概念的涵盖范围由分数意义扩展到相关性和使用、价值意义及社会影响,拉开了语言测试社会维度的研究序幕[3]20。Messick 的效度理论涉及社会维度的两个层面:测试分数的社会、文化意义以及测试使用对社会造成的影响。其中两个关键概念:构念代表性不足(construct-underrepresentation)和构念无关差异(construct-irrelevant variance),体现了他在效度理论中包含的对公平性问题的思考。受 Messick 的影响,从 20 世纪 90 年代起,语言测试研究者开始关注语言测试的社会话语,强调语言测试效度验证中社会用途及社会后果的重要性。Kane 基于论证的效度验证模型与 Messick 的整体效度观一脉相承,认为基于分数所做的实际决定受社会价值和政策的制约,应考虑可能产生的社会后果[4]38-40。Bachman 和 Palmer 建立了测试有用性框架,该框架包括测试的信度、构念效度、真实性、交互性、影响和可操作性[5]。Bachman 和 Palmer 把"影响"纳入到该框架中,认为语言测试的使用必须考虑其对个体、教育体系和社会可能产生的影响。Bachman 于 2005 年首次提出"评估使用论据"(Assessment Use Argument,简称 AUA)框架[6],并在他与 Palmer 2010 年的新著中进行了系统的论述,将测试有用性框架发展为对测试结果解释和使用的合理性进行检验的理论,并进一步强调语言测试使用要考虑测试后果、道德和公平性等社会问题[7]。

McNamara 和 Roever 指出,效度理论的发展在不同程度上包含了对语言测试社会维度多角度的思考,但缺乏对测试所给予的更广泛的社会政治环境的关注及系统的理论支撑,并对测试承担的社会角色及受到的政治、政策影响表示默许和接受,而没有采取批判视角审视测试背后无形的权力之手[8]40-41。

受后现代主义批判理论研究范式的影响,对测试的批判性思考始于 20 世纪 70 年代末 80 年代初。福柯在《规训与惩罚:监狱的诞生》一书中着力论述了测试作为象征权力和控制的工具,把日常生活的常态转化为权力的行使,认为"检查(考试)(examination)把层级检视的技术与规范化裁决的技术结合起来,把权力的仪式、试验的形式、力量的部署、真理的确立都融为一体"[9]208。Pennycook 提出批判应用语言学(Critical Applied Linguistics)的概念,将其定义为应用语言学范畴内的实践与社会、政治、文化思考的结合,主张语言和话语应与广泛的社会语境相结合,应具有政治和意识形态特征[10-11]。Pennycook 总结了批判手段的三大特征②:(1)对特定领域兴趣浓厚,如性别、社会阶层、性、种族、民族、文化、身份、政治、意识形态和话语,并拒绝对这类问题给予标准化回应(normative response);(2)包含转换教学法(transformative pedagogy);(3)对批判理论持自我反思立场(self-reflexive),即审视"压制"、"不平等"、"意识形态"等一类术语所揭示的对世界的特定认识方式。受 Kramsch[12]、Pennycook[13]和 Moss[14]影响,Shohamy 在 1997 年美国应用语言学协会(American Association for Applied Linguistics)年会上做题为"Critical Language Testing and Beyond"报告时明确提出了批判语言测试(Critical Language Testing,简称 CLT)的概念,指出测试行为并不是中立的,事实上,测试是文化、社会、政治、教育及意识形态的产物和代表,它们对每一位参与者、教师和学生的生活都有影响[15]。此后,Shohamy 在《测试的权力》一书中对测试权力的生

① 转引自 T. McNamara & C. Roever, *Language Testing: The Social Dimension*, Oxford: Blackwell, 2006, p. 11。

② 转引自 B. K. Lynch, "Rethinking Assessment from a Critical Perspective," *Language Testing*, Vol. 18, No. 4(2001), pp. 351 - 372。

成和运作机制，测试在社会和政治语境中引入的目的、使用情况及所产生的后效和社会影响进行了充分的阐释，提出了批判语言测试的 15 项原则，开拓了语言测试的新领域[16]。受批判理论对事物之间权力关系及意识形态观点的影响，Shohamy 把测试视为植根于社会、政治语境中的权力工具，批判性地思考测试引入的意图、用途、后果及影响等，提倡监控测试的权力，揭示测试的误用，减少测试的负面影响，并保护测试利益相关者的权利。其他学者也对语言测试进行了批判性研究。Lynch 探讨了测试用途的道德性问题，提出了衡量一项测试是否具有公平性的评判标准[17]。《客观语言测试》一书标志着语言测试的政治功能开始重新受到关注。Spolsky 追溯了语言测试的心理计量学传统的发展，并在书的后半部分以托福考试为例，说明美国教育考试服务中心（Educational Testing Services，简称 ETS）如何利用心理计量的精确性对测试利益相关者施加权力影响、实施控制[18]。

二、批判语言测试的主要研究问题

（一）语言测试的权力机制

首先，测试的权力产生于测试活动的本质，即对某种构念进行量化的过程。测试的固有特征，如测试语言科学化、测试形式书面化、测试结果分数化等，营造了自身客观、公平的表象，也使测试被推崇为将先赋性社会（ascribed society）改造为自致性社会（achieved society），为大众提供平等机会的有力措施。正如 Broadfoot 指出的，“正是这种客观性和合理性使语言测试享有了合法性并使之难以被辩驳”[19]86。其次，测试所获得的社会权重强化了测试的权威地位。对语言能力测量的精确性愈高，意味着测试的可信度愈高，由此获得的社会权重也愈大[20]369。测试对所有利益相关者产生的影响及风险（如入学、求职、移民、入籍等）使测试在社会中树立了权威地位，并使测试使用者把测试当作控制、支配、监管的工具。再者，测试维护了社会秩序，因而受到受试者的拥护和支持。在社会中，拥有权力、起主导作用的强势一方与受制于权力的弱势一方之间形成了不成文的合同[21]。后者出于对测试科学性、客观性的信任，广泛认可测试作为选拔工具的公平性，并主动将权力交予前者，以确保自己获得与他人公平竞争的权利。前者则利用后者对测试的信任，将测试作为维持社会秩序、维护和巩固自身社会地位的工具。双方的关系出现一种颇具深意的悖论：被测试控制的一方却成为测试强有力的拥趸。如家长强烈反对废除测试，因为测试被视为衡量教育是否有效的良好标尺，在他们质疑学校和教师的教学质量时，只有测试是唯一令他们信服的衡量标准。

（二）语言测试从测量手段到权力工具的转变

批判语言测试否定了语言测试的中立立场，检视了语言测试在教育、政治、文化等各个领域的应用中渗透的“规训性的权力”（disciplinary power）[9]，批判性地指出语言测试已从测评语言能力、为教学提供反馈、促进语言能力发展的手段演变为实现多种政治目的，如规范社会行为、表征社会身份、加剧社会意识形态转变的权力工具。

在教育领域，测试成为当权者和政策制定者改革教学内容与实践、规范与监管教师和学生行为、建构并定义知识以及对学生实施量化、分类、规范化和惩戒的工具。福柯认为：“检查（考试）试图在监视的目光中寻求规范化裁决、定性和分类。通过检查，个人成为权力对象，成为整个规训化的产物，或者说是权力产生了现实的人。”[9]208 Madaus 认为测试体现了政策制定者、测试开发者及使用者的价值，并通过测试自身的“科学”手段使价值永恒化[22]。Hanson 指出测试自身的特征决

定其成为大众普遍接受且无可争议的社会化机制,定义和催生了个人,并使他们处于权力缺失与被监督的地位[23]。Shohamy则指出测试作为规范性的工具体现在学校推行全校范围的考试以促进教师教学,或教师采用考试和课堂测验以激励学生学习[16,24]。Shohamy明确阐释了测试内容成为实际课程内容的过程。若教育系统的官员认为某个领域的知识应该教授而事先并未涉及,为确保高效、快速地推进教学改革,并在短期内得以见效,引入与该内容相关的测试具有事半功倍的效果。由于教师需要承担学生没有取得好成绩带来的风险,并承受校方、家长及学生本人对好成绩的期望而产生的压力,即使教师深知改变教学方式及教学内容可能对学生的学习体系造成不利影响,也会出于无奈而改变教学内容,将测试所包含的知识作为实际的课程内容,并制订相应的课程计划,采用相应的练习资料等。而考生迫于考试对毕业、升学、评奖等带来的影响,追求分数最大化,以取得高分作为学习的最终目的。在这个过程中,考试实际上变成了控制测试利益相关者行为的有效工具。

在政治领域,语言测试被当作实现多种政治目的的工具,如筛选或剔除外来群体、重构社会身份、按政府机构意图控制移民者水平及数量、削弱外来文化入侵以促进民族融合等。在这一领域,测试的构念源于社会语言学,即测试表现是某个群组身份的体现。社会语言学的研究建立了语言与社会身份的联系。语言具有族群属性标记功能,是民族政治诉求中的重要内容。把语言能力作为公民资格和国籍的基本要求,反映了当本民族意识形态遭遇外来冲击时,语言作为国家和民族显性认同标志的重要性,在民族主义运动中起着对内唤醒民族意识、提高民族向心力,对外与其他民族相区隔的重要作用[25]5。McNamara和Roever详细论述了语言测试基于该构念的应用,如战争时期敌方军队身份的鉴别、难民国籍的确定、公民资格的授予、移民数量的控制等[8]。在《语言和一体化中的话语:对欧洲语言测试体制的批判性思考》一书中,多位作者从不同角度指出入籍考试将语言作为壁垒对他者予以排斥:van Avermaet认为语言测试被当作排他和控制移民的机制[26]37;Blackledge强调语言意识形态承担着把关的角色,创造、维持和强化人与人之间的界限[26]85;McNamara则进一步指出,语言测试不再用以测量实践性的交际技能,而是被用来掩盖测试的深层次动机,即利用测试强化语言文化霸权[26]157。

此外,测试正逐渐演变为意识形态传播的媒介与定义不同语言文化等级的工具。作为产生事实上的隐性语言政策的机制,语言测试定义了不同语言的优劣、等级与地位,压制语言及文化的多样性,抵御他者文化的进入。Shohamy以希伯来语和阿拉伯语在以色列的地位为例,指出在大多数民族国家,语言测试成为当权者维护语言忠诚度、增强民众的爱国主义与集体认同感、巩固自身地位与实施统治的有力工具[27-28]。官方文件在表面上倡导民主多元化,宣扬对多元文化的包容性,而事实上利用测试在实践中的权力,使占主导地位的民族语言及官方语言在高风险考试中具有优势,以确保其在该国语言等级中的权威地位,而其他少数民族语言则被边缘化,甚至濒临消亡。

(三)语言测试的公平性

受社会正义理论影响,测试公平性的定义有很多,测试界对效度与公平性之间的关系也有不同的看法,归纳起来,主要有以下三种观点。第一种观点以美国测试实践联合委员会(Joint Committee on Testing Practice,简称JCTP)制定的《教育测试实践公平性准则》(The Code of Fair Testing Practices in Education)(JCTP 2004)及美国教育考试服务中心(ETS)制定的《ETS公平性和质量标准》(ETS Standards for Quality and Fairness)为代表(ETS 2002),认为公平性是相对独立的测试特征,与效度之间没有清晰、连贯的关系。第二种观点把公平性视为重要的测试特征,将效度包含在内,隐含效度是实现测试公平性的必要条件。该观点在Kunnan的相关论著中更为鲜明。Kunnan把公平性定义为包含三个层次的概念:效度、途径及公平,在测试的心理计量学维度

上附加了公平性的社会、道德、法律及哲学层面[29]。第三种观点则把公平性视为效度的重要组成部分，将公平性定义为群组间的可比效度[30-31]。

McNamara 和 Roever 赞同将语言测试的效度包含在公平性框架内，并将关注点集中于公平性的社会、政治层面，强调基于心理计量学研究测试偏颇时，应融入受社会、政治因素驱动的价值判断[8]。这种公平性研究不再局限于心理计量学的范畴，而是承认测试开发设计过程中受到政治因素的制约，有助于从社会学角度提高测试的质量。Shohamy 认为长期以来测试工作者通过使用更合理的评分手段以提高测试的精确性，但测试内容并没有从根本上发生改变，仍然由当权者掌控。测试者与受试者之间不对等的权力关系影响了测试的效度。对来自不同国别、文化背景、民族传统、教育体系，且没有平等拥有所测知识的学习渠道和学习环境的考生，对他们测试相同的内容，并基于测试结果做出高风险的重大决定，具有不公平性[16,32]。Lynch 指出公平性应包含对测试利益相关者权力关系的思考，即是否考虑了所有利益相关者的立场，是否实现了福柯提出的与规范性道德准则相关的个人行为的最大化①，也即是否使权力关系处于流动的（mobile）、可逆的（reversible）、互惠的（reciprocal）状态[33]365。

批判语言测试特别指出了测试的误用与产生的社会后效对测试公平性的影响，拓宽了测试公平性的含义。Shohamy 指出测试所宣称的目的与实际目的之间的差异导致了测试的不道德、不公平。她以三门语言考试，即某国四五年级学生全国性阅读水平测试、以色列七到九年级学生阿拉伯语言水平测试及某国高考英语口语水平测试为例批判性地指出：引入语言测试的实际目的是将测试作为媒介，实现政治目的，控制教学内容与实践，而不是官方所声称的为了衡量学生语言能力，且不给予学生关于考试的任何反馈，并不能真正帮助学生提高语言水平。就这个意义而言，测试是不道德且不公平的[16,24]。

（四）语言测试的替代性评估

批判语言测试理论承认测试能测量到的领域是有限的，测试不应作为评估知识水平的唯一手段，测试者和评估者需要通过多种途径来收集测试无法测量到的被试能力的证据，多元化的评估方式能为考生能力提供效度更高的解释。Shohamy 否定了权力模型（power model）测试，认为在这种模型中，知识和价值观体现的是权力中心者的意志，其他测试利益相关者如教师、考生家长、考生本人等的想法并未得到关注。在此基础上，Shohamy 进一步提出了民主评估模型（democratic assessment model），即权力不是从权力中心转移到地方，而是两者相互合作，通过测试之外的多种方式收集考生行为的相关信息以进行评估，包括观察、访谈、个人陈述、参与课堂活动、建立学生个人档案等。Shohamy 还提出了民主评估应该遵循的五项原则：（1）应把批判语言测试的观点应用到测试实践中，监控测试作为权力工具的用途，质疑它的前提并检视它的后果；（2）测试以与被试者相互合作的方式实施；（3）测试开发、实施过程中的相关者应对测试本身及其用途承担责任；（4）应在测试设计过程中把不同群体的知识包括进来；（5）应保护被试者的权利[32]376。Lynch 指出，替代性评估的提出象征着研究范式的转变，是一种区别于传统测试的文化。替代性评估意味着采用非后实证主义视角，用非定量的手段收集和分析数据。事实上，采用定性手段不仅是方法论上的转变，也是知识论和本体论上的转变。语言不再是一个独立存在的客观实体，被动等待人们的探索和衡量，而是存在于人们对它的使用、研究和解释中[33]361。

---

① 参见 M. Foucault，″The Subject and Power，″ in H. L. Dreyfus & P. Rabinow（eds.），*Michel Foucault：Beyond Structuralism and Hermeneutics*，Chicago：University of Chicago Press，1983，pp. 208－226。

(五) 语言测试工作者的社会责任

由于语言测试可能造成的巨大影响,如入学、求职、移民、入籍等,它对被试及其他利益相关者而言意味着生活的"妨碍者"(intruder)。因此,语言测试工作者面临着在职业道德规范和个人道德意识之间寻求平衡点的困境。对语言测试工作者承担的社会责任的探讨,是语言测试伦理问题研究范畴的一个方面,不同学者所持立场各异。

Davies赞成对语言测试工作者提出伦理要求,但呼吁这些要求不是无限量的,认为测试开发者很难考虑到测试可能带来的所有社会后果。建议通过成立行业协会,制定道德规范,以资格证书作为职业门槛,创造伦理环境,以培养测试工作者的职业道德意识[34]335-337。Hamp-Lyons则认为测试工作者承担的责任范围更广,应对所有可能产生的社会后果负责[35]302。Shohamy把测试工作者定义为所有参与行动和决策、实施测试的人员,包括政策制定者、命题者、统计学家、测试出资方、测试审校者、考生家长等,把被试及其他利益相关者也视为社会责任的承担者,认为他们应与测试工作者共同分担责任[16]145。

在我国,从社会学角度分析语言测试伦理道德及测试工作者社会责任的论著较少。2007年,杨惠中、桂诗春两位教授合作撰写的《语言测试的社会学思考》一文论述了语言测试工作者的责任,认为一项好的考试要收到好的社会效果,需要考试各方的协同努力,包括语言测试工作者、政府教育主管部门、教师、学生、家长、社会用户等,呼吁为保证考试的正面社会效益,各方应共同做出努力[20]。

三、批判语言测试理论对我国语言测试实践及外语教育的启示

首先,在重视英语教育的同时应平衡母语学习和外语学习,提高母语的学习意识,保护和弘扬中国文化。随着经济全球化的发展,英语在世界范围内已成为使用范围最广、使用人口最多、使用领域最宽、使用频率最高的一种交际工具[36]233。英语作为一种国际通用语言,在国际交往中扮演着日益重要的角色。加强外语教育对促进我国社会经济的发展及全球经济一体化具有重要的作用。《国家中长期教育改革和发展规划纲要(2010—2020)》明确提出了"培养各种外语人才"的重要任务,体现了我国在全球化背景下对具备国际化素养的人才的迫切需求,而较高的外语能力是其中不可或缺的条件。语言是文化的重要载体,个人或社会整体接受和掌握一种语言的过程,意味着其世界观、价值观与意识形态潜移默化地受到该语言背后的文化的影响。外来文化的普及必将在不同程度上对本国文化造成冲击。如何正确处理母语与外语的关系、本国文化与外来文化的关系,在重视外语教育尤其是英语教育的同时增强母语学习意识,保护和弘扬本国优秀的传统文化,是我国制定外语教育相关政策时应思考的重要问题。近年来,各种英语水平考试,如大学英语四/六级、托福、雅思、托业、剑桥商务英语等,在学生毕业、求职、出国留学中的重要性与日俱增。许多家长从基础教育阶段开始就对孩子的外语教育给予高度关注,在孩子的中文尚处于起步阶段时,英语教育已经同步甚至抢先一步开始了。对英语的关注反映的是一种社会需求,而强调英语在经济、政治、学术及日常生活中的重要性及门槛作用,在一定程度上巩固了英语的强势地位,对汉语的主导地位产生了冲击。学生的母语意识淡化,并在跨文化交际中频频出现"中国文化失语症"[37]。面对这样的情况,有必要全面开展英语能力需求调查,制定更切合实际的英语教育政策。同时,在英语教育的过程中,应该更多地融入中国文化元素,把文化传承贯穿于英语教学全过程各环节,引导学生吸收借鉴国外优秀文化成果,树立文化自觉和文化自信,提高跨文化交际能力,增强对中华文化和社会主义文化的认同。

其次，从批判角度思考测试结果在社会领域内的使用，避免误用及滥用。语言测试从测评语言能力、为教学提供反馈、促进语言能力提高的初衷逐渐演变为权力渗透和实施监管的工具，已经超越了测试专业领域的管辖范围，对测试分数的解释和使用与政策制定及测试使用者高度相关。而某项测试的社会权重越大，对利益相关者的影响也越大。因此，要从批判视角思考测试结果的使用，避免误用及滥用，并需要测试相关者如政府部门、学校领导和教师、家长和学生等的共同努力。如在我国已实施二十多年的大学英语四/六级考试的初衷是对教学大纲所要求的我国大学生英语能力进行客观、公正、准确的评价，为推动我国大学英语教学服务，为我国各级用人部门选拔人才、合理使用人才的目的服务[38]21。但近年来频现对大学英语四/六级考试结果误用的现象：如将四/六级考试成绩与学位授予、毕业证书发放挂钩；一些地方和高校对四/六级考试成绩统计排名，并以此作为经费划拨的依据；还有一些用人单位将四/六级考试成绩作为招聘的准入条件等。考试被赋予的社会权重迫使一些英语基础薄弱的考生为获得证书而采用题海战术或临场猜题等各种应试技巧，并没有真正提高自身的英语能力，违背了四/六级考试开发设计者的意图。测试还引发了一系列负面的反拨效应，如应试教学导致教学计划无法完成，教学内容狭隘化，学生学习的兴趣点受考点影响，所测知识与文化呈现制度化、标准化、同质化等。以大学英语四/六级考试为主的大规模高风险考试的社会公信力已获得公众的广泛认可，而如何降低其社会权重，避免将考试作为某些选拔的唯一标准，并对考试用途进行充分论证，是回归测试结果合理使用的根本出发点，也是相关政策制定及完善的方向。

最后，应尝试采用多元评估方式，降低以单一的大规模考试作为决策依据的风险。现阶段我国教育仍倾向于“一考定终生”，呈现出了不公平性和失误零容忍性。对部分平时表现优良但在考试中因不可抗力如身体不适、交通堵塞等意外情况发挥失常的学生有失公允。以一次采样区分考生能力无法全面、客观地评测考生在不同阶段的学习进展及语言能力，有一定的局限性。此外，一些大规模考试如高考、研究生入学考试、大学英语四/六级考试的命题者因考试巨大的社会影响力及风险性而承载了额外的心理压力。替代性评估方式的提出有助于降低一次考试作为唯一衡量标准的风险，通过学生自测、观察、访谈、个人陈述、课堂活动参与、建立学生个人档案等方式，把形成性评估与终结性评估结合起来，全面真实地反映考生的语言能力。考虑到替代性评估可能出现的问题，如由于我国考生人数庞大，形成性评估与终结性评估的结合在实施过程中会耗费大量人力、财力和时间，需先在个别城市和学校进行试点，评估其可行性及可操作性，进而探究是否适合在全国范围内推广。

四、结　语

作为审视语言测试的多种视角之一，批判语言测试理论的提出，敦促语言测试界将研究重点从语言测试的心理计量学、统计学范畴向社会学范畴延伸，象征着传统语言测试研究范式向批判理论范式的转变。批判语言测试理论不再单纯研究测试本身及测试的技术话语，而是转向对测试的社会话语的关注，为语言测试的社会维度研究提出了批判性视野，对测试中渗透的权力关系深入挖掘剖析，倡导规避大规模测试带来的高风险，减少测试结果的误用及滥用，对公平、合理地利用已具备较高信度和效度的测试，回归测试设计开发者的初衷具有重要意义。尽管从本文的上述归纳中可以看出这方面的研究已取得一些进展，但总体而言仍然十分有限。首先，语言测试的主要研究问题决定了其作为应用语言学较为独立的一个分支，与其他分支以及其他人文社会科学的关系相对疏远；其次，由于缺乏系统翔实的理论与操作性较强的框架支撑，将语言测试视为一种社会、政治实践的研究难以开展；最后，语言测试研究者对社会学、批判理论等了解有限，导致批判视角仍处于非主

流地位。虽然困难重重,但语言测试的跨学科研究具有积极意义,进一步思考如何借鉴和发展批判语言测试理论,并将其应用于语言测试研究与实践中,具有深远的意义。

[参 考 文 献]

[1] L. J. Cronbach & P. E. Meehl, "Construct Validity in Psychological Tests," *Psychological Bulletin*, Vol. 52, No. 4(1955), pp. 281 - 302.

[2] L. J. Cronbach, "Five Perspectives on the Validity Argument," in H. Wainer & H. I. Braun(eds.), *Test Validity*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1988, pp. 3 - 18.

[3] S. Messick, "Validity," in R. L. Linn(ed.), *Educational Measurement*, New York: Macmillan, 1989, pp. 13 - 104.

[4] M. T. Kane, "Validating High-Stakes Testing Programs,"*Educational Measurement: Issues and Practices*, Vol. 21, No. 1(2002), pp. 31 - 41.

[5] L. F. Bachman & A. S. Palmer, *Language Testing in Practice*, Oxford: Oxford University Press, 1996.

[6] L. F. Bachman, "Building and Supporting a Case for Test Use," *Language Assessment Quarterly*, Vol. 2, No. 1(2005), pp. 1 - 34.

[7] L. F. Bachman & A. S. Palmer, *Language Assessment in Practice*, Oxford: Oxford University Press, 2010.

[8] T. McNamara & C. Roever, *Language Testing: The Social Dimension*, Oxford: Blackwell, 2006.

[9] [法]米歇尔·福柯:《规训与惩罚:监狱的诞生》,刘北成、杨远婴译,北京:北京三联书店,1999年。[M. Foucault, *Discipline and Punishment: The Birth of the Prison*, trans. by Liu Beicheng & Yang Yuanying, Beijing: Beijing Joint Publishing Company, 1999.]

[10] A. Pennycook, *Critical Applied Linguistics: A Critical Introduction*, Mahwah, NJ: Lawrence Erlbaum Associates, 2001.

[11] A. Pennycook, "Introduction: Critical Approaches to TESOL," *TESOL Quarterly*, Vol. 33, No. 3(1999), pp. 329 - 348.

[12] C. Kramsch, *Context and Culture in Language Teaching*, Oxford: Oxford University Press, 1993.

[13] A. Pennycook, *The Cultural Politics of English as an International Language*, New York: Longman, 1994.

[14] P. Moss, "Enlarging the Dialogue in Educational Measurement: Voices from Interpretive Research Traditions," *Educational Researcher*, Vol. 25, No. 1(1996), pp. 20 - 28.

[15] E. Shohamy, "Critical Language Testing and Beyond," *Studies in Educational Evaluation*, Vol. 24, No. 4 (1998), pp. 331 - 345.

[16] E. Shohamy, *The Power of Tests: A Critical Perspective on the Uses of Language Tests*, London: Longman, 2001.

[17] B. K. Lynch, "In Search of the Ethical Tests," *Language Testing*, Vol. 14, No. 3(1997), pp. 315 - 327.

[18] B. Spolsky, *Measured Words: The Development of Objective Language Testing*, Oxford: Oxford University Press, 1995.

[19] P. Broadfoot, *Education, Assessment and Society: A Sociological Analysis*, Buckingham: Open University Press, 1996.

[20] 杨惠中、桂诗春:《语言测试的社会学思考》,《现代外语》2007年第4期,第368 - 374页。[Yang Huizhong & Gui Shichun, "Social Dimensions of Language Testing," *Modern Foreign Languages*, No. 4(2007), pp. 368 - 374.]

[21] P. Bourdieu, *Language and Symbolic Power*, Cambridge, MA: Harvard University Press, 1991.

[22] G. Madaus, *Testing as a Social Technology: The Inaugural Annual Boise Lecture on Education and Public Policy*, Boston, MA: Boston College, 1990.

[23] F. A. Hanson, *Testing Testing: Social Consequences of the Examined Life*, Berkeley, CA: University of California Press, 1993.

[24] E. Shohamy, "Testing Methods, Testing Consequences: Are They Ethical? Are They Fair?" *Language Testing*, Vol. 14, No. 3(1997), pp. 340 - 349.

[25] 陈平：《语言民族主义：欧洲与中国》，《外语教学与研究》2008 年第 1 期，第 4 - 13 页。[Chen Ping, "Linguistic Nationalism: Europe and China," *Foreign Language Teaching and Research*, No. 1(2008), pp. 4 - 13.]

[26] G. Hogan-Brun, C. Mar-Molinero & P. Stevenson(eds.), *Discourses on Language and Integration: Critical Perspectives on Language Testing Regimes in Europe*, Amsterdam: John Benjamins, 2009.

[27] E. Shohamy, "Language Tests as Language Policy Tools," *Assessment in Education*, Vol. 14, No. 1(2007), pp. 117 - 130.

[28] E. Shohamy & T. Kanza, "Language and Citizenship in Israel," *Language Assessment Quarterly*, Vol. 6, No. 1(2009), pp. 83 - 88.

[29] A. J. Kunnan, "Fairness and Justice for All," in *Fairness and Validation in Language Assessment*, Cambridge: Cambridge University Press, 2000, pp. 1 - 14.

[30] W. W. Willingham & N. Cole, *Gender and Fair Assessment*, Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

[31] W. W. Willingham, "A Systemic View of Test Fairness," in S. Messick (ed.), *Assessment in Higher Education: Issues in Access, Quality, Student Development, and Public Policy*, Mahwah, NJ: Lawrence Erlbaum Associates, 1999, pp. 213 - 242.

[32] E. Shohamy, "Democratic Assessment as an Alternative," *Language Testing*, Vol. 18, No. 4(2001), pp. 373 - 391.

[33] B. K. Lynch, "Rethinking Assessment from a Critical Perspective," *Language Testing*, Vol. 18, No. 4 (2001), pp. 351 - 372.

[34] A. Davies, "Demands of Being Professional in Language Testing," *Language Testing*, Vol. 14, No. 3(1997), pp. 328 - 339.

[35] L. Hamp-Lyons, "Washback, Impact and Validity: Ethical Concerns," *Language Testing*, Vol. 14, No. 3 (1997), pp. 295 - 303.

[36] 金光美：《英美霸权与英语全球化》，《甘肃社会科学》2009 年第 2 期，第 230 - 233 页。[Jin Guangmei, "British and American Hegemony and Globalization of English," *Gansu Social Sciences*, No. 2(2009), pp. 230 - 233.]

[37] 从丛：《"中国文化失语"：我国英语教学的缺陷》，《光明日报》2000 年 10 月 19 日。[Cong Cong, "Chinese Culture Aphasia: The Defect of English Teaching in China," *Guangming Daily*, 2000 - 10 - 19.]

[38] 杨惠中：《大学英语四、六级考试十五年回顾》，《外国语》2003 年第 3 期，第 21 - 29 页。[Yang Huizhong, "15 Years of the CET and Its Impact on Teaching," *Journal of Foreign Languages*, No. 3(2003), pp. 21 - 29.]