

《中国英语学习者水平测评——考试 构念、后果和困境》评介^①

王隽 上海交通大学

DOI:10.13564/j.cnki.issn.1672-9382.2017.02.015

1 引言

我国是考试大国，考试在我国具有悠久的历史。自改革开放以来，由于国家对英语教育的重视，英语教学与测试发展迅速。英语考试已成为学生学习和生活中不可或缺的一部分。我国的英语考试往往规模大、风险高，而且对教学的反拨作用明显。正如语言测试学者 Shohamy (2001) 所说，考试的影响力渗入教育、政治、经济等各个层面。我国目前拥有全世界最庞大的英语考试考生群体，例如2015年报考大学英语四、六级考试的人数约为1 800万，是全世界规模最大的单科考试 (Jin, 2014)。此外，我国目前也是全世界最大的留学生输出国之一，中国学生在托福、雅思等国际性英语考试参试人群中的比重也是最大。鉴于考生群体的数量庞大及考试的社会影响广泛，对中国学生英语水平的测试近年来已成为热点话题之一。虽然已有一些学术著作问世 (如 Cheng & Curtis, 2010; Jin & Cortazzi, 2011)，但是总体而言，针对中国英语学习者群体^②的测评研究仍相对缺乏。

在此背景下，由英国布里斯托大学的 Guoxing Yu 博士和上海交通大学的 Yan Jin 博士主编、国际著名出版机构帕尔格雷夫·麦克米伦出版公司出版的《中国英语学习者水平测评——考试构念、后果和困境》Assessing Chinese Learners of English: Language Constructs, Consequences and Conundrums 一书问世。该书汇聚国内外众多知名学者的观点和见解，通过丰富多样的实证案例展示了中国英语学习者独特的群体特点；同时该书对中

国的英语考试环境进行深入剖析，引发了对当代语言测试理论研究和考试实践的社会性思考。

2 内容简介

本书共包括14个章节。在第一章中编者介绍了该书的编撰背景、目的以及全书的脉络和结构。从第二章开始按主题编排可分为5个部分，以下简要介绍各部分的主要内容。

第一部分即第二章，主题为香港地区校本考试 (School-Based Assessment, 简称SBA) 的设计与实施。SBA由香港考试局开发，是兼具形成性测试和终结性测试特点、旨在以考促学的新型评测模式。作者Hamp-Lyons是香港SBA的核心开发者之一，她在本章中详细讲述了该考试的设计理念和考试实施过程中遇到的挑战，并在此基础上提出了未来的解决方案。

第二部分包括4个章节 (第三至六章)，主题为中国学生在英语口语考试中的表现。第三章为香港SBA口语考试效度研究的一部分。Lam发现，不同学校在小组交互任务中给予学生的准备时间不同，有些学生甚至可以借此在考前编好对话并在考试中将其背诵出来。作者指出，考前准备时间的差异严重影响了考试任务的真实性，从而影响到SBA的测试效度。第四章中，Jin & Zhang探讨了大学英语四、六级口语考试的施考模式对口头交际策略使用的影响。结果显示，在面试和机考两种考试模式下学生所使用的口头交际策略基本相同，是否采用机考对交际策略的使用没有造成显著影响。该研究为大学英语四、六级口语机考提供了一定的效度证据。第五章和第六章的主题是中国

作者简介：王隽，上海交通大学外国语学院在读博士。研究方向：语言测试。E-mail: jwang1989@sjtu.edu.cn。

学生在英语口语考试中词汇运用的特点。第五章中, Xu运用基于语料库的对比研究观察到中国学生在面对面的口语考试中会过度使用连词和感叹词, 而对于能够促进交际的词或词块使用率则很低。第六章中 Wang & Chen指出, 在英语专业四级口语考试的复述任务中, 学生在口语惯用语上的使用非常有限, 即使是原文中出现过的惯用语也很少被使用。以上两项研究表明中国英语教学中针对口语词汇的教学和学习相对缺乏, 在一定程度上阻碍了学生英语口语能力的提高。

在第三部分(第七章)中, 针对中国学生英语词汇习得困难的问题, Tang & Treffers-Daller通过实证研究检验了Laufer & Hulstijn (2001) 所提出的“投入量假设”(Involvement Load Hypothesis)。研究发现无论是在即时还是延迟的情况下, 投入量高的阅读任务能够让学生更有效地记住单词。该研究结果对通过控制阅读任务的投入量来促进词汇习得、评测词汇学习进程等提供了重要参考。

第四部分(第八至十二章)重点聚焦考试利益相关群体对考试的想法。在第八章中, Xi et al.报道了学生对用于考前练习的托福网上口语自动评分系统(SpeechRaterSM)的评价和态度。调查结果显示, 学生对该系统总体上持正面态度, 但多数学生认为传统的人工评分比机器评分更加精确; 高风险考试最好同时结合人工、机器两种评分方法, 单独的机器评分则更适用于低风险考试。第九章报告了香港高校英语专业本科生对基于项目的小组评估(project-based group assessment, 简称PBGA)的态度。Qian发现学生普遍认为由于团队分工不均等干扰因素的影响, PBGA并不能真实、公平地体现他们的语言水平, 因此PBGA在课堂教学中应谨慎使用, 其成绩也不适合用于做高风险的决策。第十章同样关注考试改革, Chen & May报道了学生对在大学英语课程中添加形成性评估这一教育政策的看法和意见。通过采用个案研究方法, 作者揭示了各类社会因素(如区域经济发展程度)

对教育政策在语言课堂中的实施所产生的影响。第十一章调查了中国大学生的英语学习动机。Zheng发现, 学生的英语学习动机受到了来自社会、课堂、个人等诸多方面因素的影响。第十二章的研究对象为教师群体。Zhang发现, 母语为汉语和英语的语言教师对于在教学中使用何种英语语言变体的态度存在差异。英语为本族语的教师更能接受英语的多样化变体, 而中国教师则更偏向于使用标准英语。由此, Zhang建议考试机构在聘用评分员时, 应适当考虑他们的语言背景以进一步确保评分信度。

本书的最后两章(第十三、十四章)主要探讨考试对教学的影响。在第十三章中, Lin采用民族志式个案分析的方法调查了在我国台湾两所社会背景不同的大学中“全民英检(General English Proficiency Test, 简称GEPT)”对英语教学影响的异同。该研究验证了考试的风险对考试影响的助推作用, 即考试风险度越高, 考试影响就越深远。最后一章中, Gu & Saville回顾了过去20年剑桥英语系列考试在中国的发展情况, 并分析了社会因素(如考试文化、学习方法等)对各年龄段学生英语学习的影响。基于分析结果, 他们认为国际语言考试的设计应充分体现当地的环境特征, 这样才能对当地的英语教育产生更为积极的影响。

3 简评

本书紧扣“中国英语学习者水平测评”这一主题, 融理论和实践为一体, 深入剖析了当代语言测试领域的前沿和趋势。本书具有以下三个鲜明特色。

第一, 本书通过丰富多样的实证研究案例, 从多个角度探讨了英语测评的相关问题。本书所收录的研究涉及多个中国英语学习者群体和多种英语测评。这些中国英语学习者群体来自不同地域, 包括大陆、香港和台湾地区, 也包括不同学习背景的英语学习者, 如非英语专业和英语专业的大学生、中等职业学校学生、中小學生和社会考生等。除了探讨不同的

英语学习者以外, 书中也涉及多种英语测评方式, 包括大规模、高风险的英语考试(如大学英语四、六级考试、英语专业四级考试)、英语课堂中的形成性评估(如基于项目的小组评估)、计算机辅助考试(如托福网上模拟考试、大学英语四、六级口语机考)、国际英语考试(如剑桥英语系列考试)等。本书收录的研究将以上各类考试放置于特定的社会教育环境中, 结合英语学习者的群体特点, 探讨了英语考试效度、学生在考试上的表现、考试利益相关群体对于考试的想法、考试对教学的影响等众多语言测试研究中的关键性问题。值得一提的是, 本书侧重于展现对考试过程中考生和其他考试利益相关群体的关注, 体现了语言测试的社会学属性(参见如McNamara & Roever, 2006)。语言测试本质上是一种社会活动, 围绕语言测试的研究不应囿于对考试成绩的统计分析, 而需更多关注考试与社会、文化和教育环境之间的互动关系, 并将考试后果、文化价值等社会因素融入考试效度理论探讨和实证研究中去。

第二, 本书准确、深入地把握了英语考试在当代中国的发展现状。我国是名副其实的“考试大国”, 考试历史源远流长。在我国目前的教育环境中, 高风险、大规模考试占据语言测评的主流地位, 如本书中出现的大学英语四、六级考试(第四、五、十一章)、英语专业四级考试(第六章)、托福(第八章)等。对以上这些考试的关注, 是对中国主流英语考试现象的深刻剖析。

尽管大规模、标准化考试依然是关注的热点, 但是本书除此以外还重点关注了正在兴起的新型评测模式。随着测试理论与实践的不断发展, 当前“以考促学”(Assessment for Learning, 简称AFL, 参见如Black & Wiliam, 1998)的理念影响深远。在这个理念的推动下, 以学习为目的的新型评测模式逐渐成为当代语言教育改革的新主角。同大规模、标准化考试相比, 新型评测模式常常以学校、课堂为研究对象, 更加注重学习的过程和结果, 降低了对高信

度、高区分度和高分的追求。例如,本书第二、三章中的香港SBA考试改革,将考试融入教学大纲,使考试成为课堂教学的一部分,突出对“学”的重视;第九章中出现的PBGA,要求小组合作完成项目,强调项目完成过程中小组成员间的合作、交流和互动;第十章提及的形成性评估重视学生在学习过程中的参与程度,并要求通过提高参与度达到预设的学习目标。

与此同时,本书也讨论了形成性评估在语言课堂中实施所遇到的各种挑战。例如Hamp-Lyons在第二章中所述,在SBA实行之初,校长、家长和老师们焦虑情绪很高以至于社会上的负面情绪高涨,后经多方努力才得以缓解。此外,在SBA实行过程中,考试开发者们发现实际情况同期望相差甚远,绝大部分老师对于政策的理解和实施仍需要大量的引导。这些案例说明,新兴的考试理念往往同传统的标准化考试理念和考试高风险权重存在矛盾,如何平衡这个矛盾是当代语言测试工作者们无法回避的一个问题。

因此,本书通过呈现高风险、大规模考试的社会主流地位和新型评测模式的逐渐兴起,展示了当代中国极富特点的英语考试发展现状。

第三,本书收录多种定性研究方法,展示了领域内部的前沿研究趋势。随着时代的发展,现代语言测试领域的研究已不再是定量统计方法一枝独秀,越来越多的研究采用定性研究方法。本书收录多种定性研究方法,根据话题可分为三类:(1)运用话语分析方法分析学生的口语考试表现;(2)采取反思法和启发式访谈调查考试利益相关群体的看法;(3)应用个案分析和民族志式研究方法调查考试影响。具体而言,首先,话语分析方法对于深入调查学生在考试中的表现非常适用。例如,本书第三、四章中的研究均使用该研究方法剖析学生在口语考试交互任务中的话语特点,并以此为基础探究了测试方法对学生考场表现的影响。其次,在收集考试利益相关群体的意见和看法时,通过反思法、访谈等定性数

据收集方法可以获取更为深刻的反馈。如本书第三章中,在对学生的考试表现进行会话分析的基础上,研究配合运用反思法从学生和考官角度调查了其他影响考试任务真实性的因素;第十一章中Zheng运用启发式访谈的方法深度调查了学生英语学习和参加英语考试的动机。最后,在分析考试的社会影响时,定性研究视角有助于深入揭示考试同社会各个层面相互交织的复杂性和动态性。如本书第十章中Chen & May运用个案分析方法调查了社会文化因素是如何影响学生在英语课堂活动中的参与度以及他们对英语考试和英语测评改革的态度;第十三章中Lin运用民族志式个案分析的方法,深入、全面地剖析了GEPT在台湾社会中的反拨效应。总之,本书收录的研究中运用了多种定性研究方法,是当代语言测试领域前沿研究趋势的体现。

综上所述,本书对现代语言测试领域中的理论探索和实证研究均有重要贡献。首先,本书通过丰富的实证研究,从多个角度展示了中国英语学习者的群体特点,为研究者提供了可学习、借鉴的宝贵经验。其次,本书结合中国社会的实际情况,实时、准确地把握了中国英语考试的发展现状,促进了世界对中国的英语考试和中国英语学习者群体的了解。最后,本书通过展现考试影响现象的复杂实质,引发了考试的社会学角度思考,有助于语言测试的利益相关群体更深刻、更理性地看待考试以及考试改革。尽管如此,本书依旧存在一些美中不足的方面。首先,本书主要关注的是口语测试,对其他语言技能关注度不高。其次,本书的研究对象主要集中在大学阶段的英语学习者群体,对中、低级别的学习者关注较少。最后,全书虽然各部分重点明确,但章节间的衔接不够明显。但正如Weir在本书前言中所说,本书的意义不仅仅在于加深了世界对中国英语测试与学习者的了解,同时对语言测试领域的知识积累也做出了重要贡献。当前我国外语测试与教学正处于重要的改革时期,本书对于改革的内容与方向都极具启迪作用,是一部值得语言测评的研

究人员、语言测评及相关专业方向的研究生、英语教育工作者等仔细阅读的佳作。□

注释

- ① 衷心感谢上海交通大学金艳教授和复旦大学范劲松博士的悉心指导。
- ② 本文中的“中国英语学习者群体”指母语为汉语的英语学习者群体。

参考文献

- [1] Black, P. J. & D. Wiliam. Assessment and classroom learning [J]. *Assessment in Education*, 1998, 5 (1): 7-74.
- [2] Cheng, L. & A. Curtis. *English Language Assessment and the Chinese Learner* [C]. NY & London: Routledge, 2010.
- [3] Jin, L. & M. Cortazzi. *Researching Chinese Learners: Skills, Perceptions and Intercultural Adaptations* [C]. Basingstoke: Palgrave Macmillan, 2011.
- [4] Jin, Y. The limits of language tests and language testing: Challenges and opportunities facing the College English Test [A]. In D. Coniam (ed). *English Language Education and Assessment* [C]. Singapore: Springer, 2014.
- [5] Laufer, B. & J. H. Hulstijn. Incidental vocabulary acquisition in a second language: The construct of task induced involvement [J]. *Applied Linguistics*, 2001, 22(1): 1-26.
- [6] McNamara, T. & C. Roever. *Language Testing: The Social Dimension* [M]. Oxford: Blackwell Publishing, 2006.
- [7] Shohamy, E. *The Power of Tests: A Critical Perspective on the Uses of Language Tests* [M]. London: Longman & Pearson Education, 2001.
- [7] Yu, Guoxing & Yan Jin. Assessing Chinese Learners of English: Language Constructs, Consequence and Conundrums[G]. UK: Palgrave Macmillan, 2016.