



中国书籍文库

China Books Library

汇集优秀原创学术论著

推动科研成果转化交流

口语测试 理论与实践

——口语测试理论及国内外主要口语测试种类简介

KouYu CeShi LiLun Yu ShiJian

毛忠明 黄自然 等编著



中国书籍出版社

China Book Press

口语测试 理论与实践

——口语测试理论及国内外主要口语测试种类简介

KouYu CeShi LiLun Yu ShiJian

毛忠明 黄自然 等编著



中国书籍出版社
China Book Press

图书在版编目(CIP)数据

口语测试:理论与实践:口语测试理论及国内外主要口语测试种类简介/毛忠明等编著. —北京:中国书籍出版社, 2012. 9

ISBN 978-7-5068-3193-2

I. ①口… II. ①毛… III. ①口语—测试—研究

IV. ①H09

中国版本图书馆 CIP 数据核字(2012)第 217289 号

责任编辑/ 许艳辉

责任印制/ 孙马飞 张智勇

封面设计/ 中联学林

出版发行/ 中国书籍出版社

地 址: 北京市丰台区三路居路 97 号(邮编:100073)

电 话: (010)52257143(总编室) (010)52257153(发行部)

电子邮箱: chinabp@vip.sina.com

经 销/ 全国新华书店

印 刷/ 北京天正元印务有限公司

开 本/ 710 毫米×1000 毫米 1/16

印 张/ 12.5

字 数/ 225 千字

版 次/ 2013 年 1 月第 1 版 2013 年 1 月第 1 次印刷

书 号/ ISBN 978-7-5068-3193-2

定 价/ 38.00 元

版权所有 翻印必究

0 序言

本书是上海市语言文字工作委员会立项资助，上海市语言文字水平测试中心科研规划重点项目——“语言文字水平测试研究课题”的一项成果。2011年初，为适应越来越多外籍人士来沪求学、工作、经商、旅游、居住的需要，上海市语委立项设立了“实用汉语能力测试”研究课题，旨在为外籍在沪人士构建一个学习汉语、检测汉语水平的公共服务平台。本书几位作者承担了该课题的子课题项目——“口语测试理论及国内外主要口语测试种类简介”，负责梳理语言测试理论的发展脉络，研究口语考试的设计、施测、评估的原则和要点，评析国内外几种主要的口语考试项目，为该课题的教材编写及测试的设计和提供理论支撑和可资借鉴的案例分析。

在上海市语委领导和上海市语测中心同仁的指导下，课题组成员不懈努力，协同合作，终于使课题研究报告一段落。几个月来，课题组成员广阅博览，条分缕析，梳理了国内外学者有关语言测试的理论，尤其是学界关注较少的口语测试理论，分析了国内外几种较有影响的口语测试模式，形成了本课题报告。本课题报告分为6个部分：1. 语言测试理论研究综述，由张锡九、黄友执笔；2. 语言测试的类型、内容和方法，由黄自然、张锡九执笔；3. 作为第二语言的口语测试，由黄自然执笔；4. 国外主要语言测试中的口语考试，由杨显华、黄友、刘凯红、黄蓓、常志斌执笔；5. 国内主要语言测试中的口语考试，由黄自然执笔；6. 关于汉语口语考试设计、测试、评估的建议，由课题组成员共同承

担。毛忠明负责本课题的总体设计及本书的统稿。由于时间仓促，资料有限，本课题研究疏漏在所难免，敬请读者谅解和斧正。希冀本课题的研究及本书的出版能为改进和提高“汉语口语水平考试”，或为同类型的标准化口语考试的设计和施测提供有益的借鉴。

毛忠明

2012. 07. 08

于上海大学

目 录

CONTENTS

1 语言测试理论研究综述	1
1.1 语言学的发展阶段及相应的语言测试模式	/ 2
1.1.1 科学前语言测试模式	/ 2
1.1.2 心理测量—结构主义语言测试模式	/ 3
1.1.3 心理语言学—社会语言学测试模式	/ 4
1.1.4 交际语言测试模式	/ 4
1.2 国内外现有语言能力标准	/ 8
1.2.1 ALTE 框架 2011	/ 9
1.2.2 CEF 2001	/ 11
1.2.3 ILR(FSI) 量表 2011	/ 14
1.2.4 ACTFL 水平指南 2012	/ 17
1.2.5 “21 世纪外语学习标准”及“全美中小学中文学习目标” 1999	/ 18
1.2.6 加拿大语言等级标准 (CLB) 2000	/ 23
1.2.7 澳大利亚 ISLPR 量表 1999	/ 29
1.2.8 国际汉语能力标准	/ 31
1.2.9 汉语口语水平等级标准及测试大纲(HKC2010)	/ 35
1.2.10 分析和结论	/ 36

2 语言测试的类型、内容和方法	39
2.1 语言测试的作用	/ 39
2.2 语言测试的类型	/ 40
2.2.1 按测试目的划分	/ 40
2.2.2 按分数解释的参照标准划分	/ 43
2.2.3 按评分方法划分	/ 44
2.2.4 按测试内容或认知态度划分	/ 44
2.2.5 按测试所用手段划分	/ 45
2.2.6 其他分类方法	/ 48
2.2.7 结论	/ 48
2.3 语言测试的项目和内容	/ 50
2.4 试卷的设计	/ 51
2.4.1 卷面的构成	/ 51
2.4.2 试题的类别	/ 52
2.5 语言测试的开发	/ 53
2.5.1 设计阶段	/ 54
2.5.2 操作阶段	/ 55
2.5.3 施测阶段	/ 57
2.6 语言测试的质量分析	/ 58
2.6.1 信度	/ 58
2.6.2 效度	/ 61
2.6.3 区分度	/ 63
2.6.4 难度	/ 63
3 作为第二语言的口语测试	65
3.1 口语水平的定义	/ 65
3.2 口语水平的维度	/ 65
3.3 口语测试的设计原则	/ 67

3.4	口语测试的基本方式	/ 69
3.5	口语测试的评分	/ 71
3.5.1	口语测试的评分方法	/ 71
3.5.2	评分员	/ 76
3.6	口语测试的反拨作用	/ 78
4	国外主要语言测试中的口语测试	80
4.1	英语作为第二语言测试中的口语测试	/ 81
4.1.1	口语能力面试(OPI – Oral Proficiency Interview)	/ 81
4.1.2	托福考试(TOEFL – Test of English as a Foreign Language)	/ 85
4.1.3	雅思考试(IELTS – International English Language Testing System)	/ 89
4.1.4	剑桥主体考试(MSE – Main Suite Examinations)	/ 91
4.1.5	剑桥少儿英语考试(CYL – Cambridge Young Learners' English Tests)	/ 97
4.1.6	剑桥商务英语证书考试(BEC – Business English Certificate)	/ 99
4.1.7	英语口语测试证书(TSE – Test of Spoken English)	/ 103
4.1.8	爱尔兰交互英语口语测试(OTIE – Oral Test of Interactive English)	/ 106
4.1.9	博思考试(BULATS – Business Language Testing Service)	/ 108
4.2	非英语语言测试中的口语测试	/ 110
4.2.1	德语考试	/ 111
4.2.2	俄语考试:对外俄语等级考试(ТРКИ)	/ 119
4.2.3	西班牙语 DELE 考试	/ 121

4.2.4	法语考试	/ 122
4.2.5	日语考试	/ 127
5	国内影响较大的口语测试	131
5.1	国内自主组织的包含口语测试的外语考试	/ 131
5.1.1	全国外语水平考试(WSK)	/ 132
5.1.2	全国出国培训备选人员外语考试(BFT)	/ 134
5.1.3	外语专业四、八级考试	/ 135
5.1.4	上海市通用小语种应用能力考试	/ 138
5.1.5	全国翻译专业资格(水平)考试	/ 139
5.1.6	小语种导游资格考试	/ 141
5.2	国内汉语作为第二语言测试中的口语考试	/ 142
5.2.1	汉语水平考试(HSK)	/ 143
5.2.2	新汉语水平考试(HSK)	/ 147
5.2.3	少数民族汉语水平考试(MHK)	/ 152
5.2.4	华语文能力测验(TOP-Hanyu, 台湾)	/ 156
5.2.5	汉语口语水平测试(HKC)	/ 160
5.2.6	实用汉语水平认定考试(C. TEST)	/ 162
5.2.7	其他汉语口语考试	/ 163
6	评论和思考	165
附录 1:	HSK(高等)口语水平评价标准	/ 173
附录 2:	HSK(高等)口语水平五级评分标准	/ 174
附录 3:	新 HSK 高级试题样卷——口语	/ 175
附录 4:	新 HSK 初级试题样卷——口语	/ 177
图表索引	179	
参考文献	181	

1 语言测试理论研究综述

语言测试是一门跨学科的综合性科学，其理论涉及语言学（理论语言学、应用语言学、心理语言学和社会语言学）、教育学、数理统计学、心理测量学和计算机科学等学科领域。口语能力是语言交际能力的重要方面，口语测试是语言交际能力测试的重要手段，是语言测试的一个重要组成部分，受到各国语言测试专家和学者越来越多的重视。目前，国外影响力较大的英语测试项目，如雅思（IELTS）、托福（TOEFL）、剑桥商务英语证书考试（BEC）等都有口语测试。新托福考试也通过互联网考试方法全面测试考生包括口语在内的英语应用能力。近年来，我国口语测试也出现日益繁荣的景象：大学英语口试（CET）、英语专业四级和八级（TEM）口语考试，公共英语等级考试（PETS）中的口语部分，研究生入学考试中的英语口语面试，还有近几年刚刚兴起的高考和中考中的英语口语部分等等（徐海铭，2006）。因此对口语的测试研究成了第二语言教学的一大重要课题。

在汉语教学中，与听力、阅读和语法水平的单项考试相比，汉语口语水平的单项考试起步较晚，以前仅仅局限在 HSK（高等）考试中。近两年来，面向初、中等的口语考试刚刚在新 HSK 考试中试行。专项的汉语口语水平测试（HKC）虽已推出，但还在研发和试验阶段，尚未推广开来。至于汉语口语成绩测试，虽然已经进行多年，但也还有不少迫切需要解决的理论问题和实际问题。而在汉语口语的学能测试方面，迄今几乎是空白。综观我国对外汉语教学界历年来发表的专著和论文，

其内容涉及语言测试的不多，探讨口语测试及其具体做法的更为少见。

如何进行大型的对外汉语口语水平测试？如何才能做到统一化和科学化？如何才能有效地测试出学生真正的口语水平及其运用汉语完成交际的能力？如何提高测试的科学性？这些问题也是国内外学者在当前研究第二语言口语水平测试时遇到的重点和难点问题。为使对外汉语口语测试更加科学可靠、更加正确可信、更加客观公平、更具操作性，我们应该深入了解语言测试理论的基础及其发展过程，口语测试的题型、施测方法及评分方法，以及口语测试的信度、效度与反拨作用等各个方面。

1.1 语言学的发展阶段及相应的语言测试模式

语言测试体系随着语言教学和对语言能力研究的发展而发展。一般认为，语言测试经历了四个发展阶段：前科学时期、心理测量—结构主义语言学时期、心理语言学—社会语言学时期和交际语言测试时期。这四个阶段分别对应四种语言测试模式，即前科学测试模式、心理测量—结构主义模式、心理语言学—社会语言学模式和交际语言测试模式。（熊明丽，2004）

1.1.1 科学前语言测试模式

科学前时期指 19 世纪 40 年代至 20 世纪 40 年代这一漫长的时期。在这一时期，语法翻译法在外语教学中占主导地位。与此相适应的第一代测试模式，称为科学前语言测试模式（The pre-scientific language testing），即所谓的传统测试法。其主要特点：重视语法规则、词形变化和词语的用法，通常测试语法、词汇、阅读和翻译。重视书面语，一般只有笔试，没有听力和口试。常见的题型有填空、语法分析、句子改

写、翻译、书面问答、缩写以及命题作文等。这些题型对语言能力的测试大多是间接的、不完全的。不能直接地、完全地反映语言交际活动和输出性的语言能力。传统测试法受机械语言学的影响，苛求语言形式的准确，注重语言知识，但忽视了语言应用，也忽视了口语测试。总之，传统测试法缺乏明确的指导思想和科学的分析方法，处于教师自发、分散、随意的组织状态中，忽视对测试成绩的分析与评估，测试的效度和信度都不理想。因此这一时期在测试学上被人称为科学前时期。（刘润清，1991）

1.1.2 心理测量—结构主义语言测试模式

心理测量—结构主义语言学时期（psychometric - structuralist era）始于20世纪40年代，迅速发展于五六十年代，至70年代达到鼎盛时期。与之相适应的语言测试模式称为心理测量—结构主义测试模式。它以结构主义的语言观和行为主义的语言学习观为理论基础，又吸收了心理测量学的科学方法，形成了一种科学、客观、可靠的测试体系。结构主义语言学派认为，学习语言就是学习语言成分以及掌握各语言成分之间相互关系的规则。此外，受当时行为主义心理学的影响，结构主义语言观还认为语言学习是一个机械的刺激—反应—巩固的过程。因此，重视口头语言的句型结构，在分析语言结构基础上进行大量的脱离语言环境的句型操练，成为当时外语教学的主要内容，主张听说领先，读写跟上。测试也往往采用分离式试题，间接地考查考生的语言水平，注重围绕语言这个由语音、词汇、语法构成的符号系统，逐项测试学生掌握操作这套系统的各项技能。口试以测试说的技能以及与说有关的分项技能为宗旨，强调口语形式的准确性与口语技能的熟练性。考试题型常以朗读、复述、补全对话、看图说话、口头报告、问答等形式出现，注重语音语调、语法、词汇、流利程度以及内容等五个方面。它强调测试的科学性、准确性和可靠性，多采用客观题，评分快捷、方便、客观、公

正, 测试信度高。但是分离式测试的效度却很有问题, 因为分离意味着把语言成分分离开来, 单独加以测量, 但当语言成分被分离开时, 语言的重要本质属性便不复存在了。(熊明丽, 2004)

1.1.3 心理语言学—社会语言学测试模式

心理语言学—社会语言学时期 (psycholinguistic – sociolinguistic era) 始于 20 世纪 70 年代。在这一时期, 语言不仅看做是一种可以分解的体系, 更被看做是一种动态的具有创造性的功能体系。(杨惠中, 2000) 乔姆斯基 (Chomsky) 在 1965 年提出了语言能力 (linguistic competence) 和语言行为 (linguistic performance) 的概念。(刘润清, 1991) 语言测试界开始重视所谓总体综合法 (global integrative approach) 的研究。主张测试在一定的语境中进行, 不刻意区分各单项语言成分、技能和能力, 而是强调多项的综合评估。综合测试法主张通过一次测试全面地评价学生的总体语言水平, 学生必须综合地运用各种语言知识和技能。因此在测试中采用比较多的是综合题, 如完形填空、综合改错、听写、口试、作文等。这样的试题能多方位地检查考生的语言能力, 在同一试题中考查语音、词汇、语法、语篇、拼写等多种能力。(熊明丽, 2004) 值得指出的是, 综合测试法顾名思义是综合测试考生运用各种语言技能的能力, 然而该种测试方法中的很多题型, 如完形填空、改错、多项选择等, 在现实生活的语境中是很少运用的。因此, 综合测试法更准确地说只是相对综合地测试考生潜在的语言运用能力。

1.1.4 交际语言测试模式

交际语言测试兴起于 20 世纪 70 年代。1972 年社会语言学家海姆斯 (Hymes) 首次提出交际能力 (communicative competence) 概念。(Hymes, 1972) 他认为语言使用不但要能够按照语法规则产生出符合语

法的句子,而且还应具有在不同语境中合理使用这些句子的能力,也就是说,语言使用涉及社会文化因素。他认为交际能力应包含四个社会文化因素:(1)可能性,即所说的话是否符合语法规则,是否在形式上可行;语言是交际工具,交际的成功不仅受到语言规则本身的制约,而且受到社会文化规则的约束。(2)可行性,即所要表达的语言行为是否能付诸实施及可实施的程度。(3)合适性,即话语在语境中是否得体及得体的程度。(4)有效性,即言语行为是否已付诸实施及已实施的程度。(文秋芳,1999)

20世纪80年代卡内尔(Canale)和斯魏恩(Swain)等人补充和发展了海姆斯的交际能力概念。他们认为交际能力至少包括四个方面的知识和技能:(1)语言能力,指正确理解和表达所说的话语的字面含义所需的语言知识;(2)社会语言能力,指在不同语境中恰当理解和表达言语的能力;(3)语篇能力,指根据不同体裁,将语言形式和语言意义结合起来,组成统一的口语体或书面体篇章的能力;(4)策略能力,指在交际过程中的应变能力。(樊长荣,2004)

20世纪90年代初,巴克曼(Bachman)发展了卡内尔、斯魏恩等人的研究成果,提出了新的交际语言能力理论模式。巴克曼的交际法语言测试理论对全球的语言教学和语言测试领域产生了很大影响,被认为是语言测试史上的里程碑。(刘润清,1991)他“首先认为语言习得和语言教学的研究和发展与语言测试之间存在内在的联系。两者之间互相影响、互相促进。”(陈菁,2002:51)他对语言能力的认识更为全面、更加深刻。“他还认为,学习者的语言能力应该有更广泛的涵义,除了对语言系统知识的掌握外,还必须包括对句子之外语言使用语境的掌握。语言的交际不只是简单的信息传递,而是发生在情景(situation)、语言使用者(language user)和语篇(discourse)之间的动态交互(dynamic interaction)。语言测试应包括语法、语篇、社会语言能力和策略能力在内的交际语言能力,同时它还应该以真实(authenticity)为准则,使被测者能同时处理清晰的语言信息(explicit linguistic informa-

tion) 和隐含的言语意义或功能意义 (implicit illocutionary or functional meaning)。” (陈菁, 2002: 51) 他对交际语言能力进行了十分详细和严格的定义, 提出了由以下三大部分组成的交际能力模式: (1) 组织篇章能力 (organizational competence, 如词汇、词法、句法、衔接、组织等) 和语用能力 (pragmatic competence, 如言语知识和社会语言知识) 组成的语言能力 (language competence); (2) 策略能力 (Strategic Competence); (3) 心理—生理语言机制 (psychophysiological mechanisms)。(Bachman, 1990) 组织篇章能力包含生成或辨认语法正确的话语能力、理解话语中主题内容的能力和把话语组织成口语的或书面的语篇的能力。语用能力包含在具体的语境中语言的表达能力和解释在语段中符合社会习惯的语用行为。它们包括语义能力、功能能力和社会语言能力。策略能力是指在具体情景下运用语言知识进行交际的心理能力 (mental capacity), 它是语言使用中的一个重要部分, 是一种综合运用已掌握知识解决问题的能力。语言使用者的策略能力使其能将语言能力与知识结构 (knowledge structure) 和交际情景 (context of situation) 的特征联系起来, 例如, 当语言使用者的语言能力出现障碍时, 他的策略能力会起到某种补偿作用, 以保证交际的顺利进行。他认为, 策略能力实际上是人们使用语言进行交际时的心理认知过程, 由四组元认知策略组成。它们是评估策略 (assessment strategies)、确定目标策略 (goal - setting strategies)、制定策略 (planning strategies) 和执行计划策略 (execution strategies)。不同的人在不同的时期对语言知识和元认知策略的掌握程度不一样, 即使同一个人在完成某一特定测试任务时, 对某些策略的使用也不一样。(Bachman, 同上) 巴克曼指的心理生理机制是语言交际时的一种神经和生理过程。如在接收语言的过程中使用的是视听技能, 而在产生语言的过程中使用的是神经肌肉技能。“巴克曼提出交际语言测试理论的重要意义在于, 把语言的使用看做一个动态的过程, 重视语言在具体环境、社会文化背景下的使用, 突出语言能力不仅指对语言知识的掌握, 更重要的是运用所掌握的知识进行有效的交际。这就

意味着测试时要把涉及交际能力的各种因素结合成一个整体加以测试。”(陈菁, 2002: 51~52)

20世纪90年代末,李筱菊进一步提出了新的交际行为能力模式范畴,即语言能力、语篇能力和语用能力。她把三者形象地比喻为一种化合物,而不是一种混合物。她认为,交际式语言测试与别的语言测试方法最突出的不同点在于它不是仅仅测试语言知识,而是把测试的重点放在应用语言的能力上,即要着重了解受试者用目标语进行交际的能力。她强调,外语测试不能分别测试交际能力的各个组成部分,而应当把交际能力的各个组成部分结合成一个整体来加以测试。典型的交际式语言测试的题项是要求受测者借助所学的语言来完成具体的、现实生活中存在的或极有可能存在的某一任务。(李筱菊, 1997)

文秋芳在分析和比较卡内尔—斯魏恩模式、巴克曼模式和李筱菊模式的基础上提出了跨文化交际能力的模式。跨文化交际能力由交际能力和跨文化能力组成。交际能力包括语言能力、语用能力和策略能力。跨文化能力即指对文化差异的敏感性、宽容性及处理文化差异的灵活性。语言能力又可细分为语法能力和语篇能力。(文秋芳, 1999)“语篇能力指将话语组成互相衔接、连贯的完整的语篇的知识以及运用这些知识的能力。语篇最重要的特点是衔接与连贯。口语交际语篇能力的一个重要标志就是要掌握一些约定俗成的会话规则,如引起注意、提起话题、发展话题、维持交谈及如何应对等。语用能力包括实施语言功能的能力,即理解和表达语言功能的能力和社会文化语言能力。也就是说,口语交际要根据交际场合和交际对象的身份,在风格、措辞、口气上作相应的调整。交际策略能力指交际者运用语言或非语言手段为克服由语言能力不足引起的交际困难的能力。”(李传芳, 2003: 64)

文秋芳总结交际测试的特点为功能性、真实性、交互性、情景性、得体性和综合性。(文秋芳, 同上)交际语言测试与其他语言测试模式最突出的不同有以下几点:(1)强调语言交际功能这一本质。交际语言测试虽也测试语言知识,但更侧重对语言行为的测试。考生不仅要掌

握好语言知识，而且还要具备正确理解各种交际情景并能在不同的语境中恰当地运用所掌握的语言知识的能力。(2) 力求真实。真实性是交际语言测试中的重要因素。交际语言测试所提供的情景要尽可能地与语言运用的客观相一致，以最大限度地反映受试者的语言运用能力。测试中的任务是现实生活中实际存在的，不是虚构的。(3) 强调听说读写四种技能在语篇层次上的综合运用能力。(4) 交互性。受试者是“局内人”，参与交际的整个过程。(5) 测试方法更注重效度而不是信度。这种测试虽难以提供客观可靠的量化数据，但测试效度很高，可以更加直接地测量考生运用语言知识的能力。(6) 评估方法为标准参照。评估标准是交际的有效性，根据所制定的评分细则由考官主观地综合评判考生是否能有效、得体地使用语言完成所规定的交际任务。(熊明丽，2004)

1.2 国内外现有语言能力标准

一个标准的语言测试都应该有其相应的能力标准。语言能力标准是指导标准化语言测试的纲领性文件，标准化语言测试的目标、内容、等级划分与成绩评定都以语言能力标准为参考指南和依据。目前国内外的语言能力标准主要有欧洲的 ALTE 框架 2011、CEF2001、美国的 ILR (FIS) 量表 2011、ACTFL 水平指南 2012 和“21 世纪外语学习标准”、“全美中小学中文学习目标”1999，加拿大语言等级标准 (CLB) 2000、澳大利亚 ISLPR 量表 1999、中国的“国际汉语能力标准”2007 和“汉语口语水平等级标准及测试大纲”(HKC2010) 等，下面分而述之。