

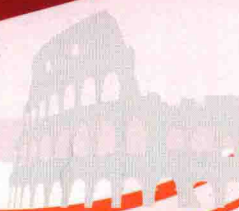


汉语国际教育核心课程系列教材

丛书主编 阮桂君 副主编 张洁 欧阳晓芳

教学测试与评估

张洁 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

汉语国际教育核心课程系列教材

丛书主编 阮桂君 副主编 张洁 欧阳晓芳

教学测试与评估

张洁 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

教学测试与评估/张洁编著. —武汉:武汉大学出版社,2017.7

汉语国际教育核心课程系列教材/阮桂君主编

ISBN 978-7-307-19211-9

I. 教… II. 张… III. 汉语—对外汉语教学—教学评估 IV. H195.3

中国版本图书馆 CIP 数据核字(2017)第 082430 号

责任编辑:白绍华

责任校对:李孟潇

版式设计:马 佳

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北恒泰印务有限公司

开本:787×1092 1/16 印张:19 字数:449千字 插页:1

版次:2017年7月第1版 2017年7月第1次印刷

ISBN 978-7-307-19211-9 定价:39.80元

版权所有,不得翻印;凡购买我社的图书,如有质量问题,请与当地图书销售部门联系调换。

总 序

2007年，几个朋友在一次会议上碰面，说希望能够把自己正在授课的内容编成教材，大家都觉得是好事，因为在对外汉语本科教学阶段，那个时候，所用的教材，基本上都是借用汉语言文学专业的，没有自己独有的教材。不过言易而行难，这句话一说，就过去了十年。我们从初涉这个领域，到埋头苦干于这个领域，也整整十年。

从对外汉语的学科名称更名为“汉语国际教育”开始，意味着中国的汉语教学正式开始大规模地从本土走向世界，国际视阈、全球思维、跨文化能力将成为国际汉语教师最基本的修养。随着孔子学院的发展，汉语志愿者外派人数与日俱增，而师资，尤其是优秀的国际汉语师资，依然缺乏。而师资的缺乏又跟培养单位的条件以及是否有合体合身的教材有着莫大的关系。

对外汉语更名为汉语国际教育后，我们认为至少在本科、硕士培养阶段，需要有以下观念的转变：

(1)从“对外”到“国际”的转变。对外汉语，大部分是针对中国境内的汉语大环境下的第二语言教学而言的，而国际汉语教学则涵盖中国境内和境外的第二语言教学，尤其是境外的第二语言教学，需要我们探索。

(2)在理念上增加“国际化”的同时，本科、硕士阶段的教材需要进行全面的调整，需要及时地在教材中增加国际化的元素，创设各种国际汉语教学中的现实场景，并始终贯穿国际汉语教师的职业精神、职业梦想。

基于此，我们邀请了中国人民大学、武汉大学、华中师范大学、首都师范大学、中南民族大学、广东外语外贸大学、湖北第二师范学院等高校战斗在汉语国际教育第一线的教师们共同来完成这套丛书的编写工作。丛书的总体指导思想是：

- (1)编写具有鲜明专业色彩的汉语国际教育专享教材；
- (2)改变以往只重视理论阐述，忽略实践应用的教材模式，把理论知识与案例结合起来；
- (3)将任课教师多年来对自己所负责课程的教学经验以及心得体会融入课程，彰显各自特色。

我们希望，学生在学习本教材后，提高三个能力：

- (1)对该学科的解释能力；
- (2)客观系统的观察分析能力；
- (3)自主学习勇于创新的能力。

我们在编写的过程中，尽量做到在现有相关材料的基础上，客观、稳妥、积极地吸收

最新的成果，通过不同章节多角度开拓学生的学科视野，既让学生掌握本学科的传统基础，又引导学生注重事实，看到不同的观点和理论，进行开放性思考，培养创新意识和素质。

我们也从体例上做了一些探索，每本教材，每个章节下的每一个小节均由 5 个部分构成：

(1) 案例导入。通过一个具体的案例导入正文基础知识内容，类似于教案中导入部分，以事实为引导，带学生进入学习内容；

(2) 基础知识。同于现在使用的一般教材除练习以外的内容；

(3) 思考与练习。由两部分构成，一个是对基础知识的问答，属于识记类型的。另外一个练习，主要是根据这一节内容，由 2~3 个操作性比较强的练习题构成，多为社会实践操作之类的活动。

(4) 案例与分析。学习完基础知识后，对导入部分的案例进行剖析，分析原因，寻求对策；

(5) 延伸阅读。这部分列出几篇与章节相关的重要文献，供学生进一步阅读使用。

令人感到高兴的是，这几年专门针对汉语国际教育这个学科本身的专享教材越来越多，大家的参与，使得这个专业正经历大浪淘沙的发展阶段，最终走向真正的成熟。我们的丛书愿意作为其中的一颗沙子，为汉语国际教育事业的发展，尽一份绵薄之力。

阮桂君

2017 年 3 月 19 日于法国巴黎东亚语言研究所

目 录

第一章 语言教学、测试与评估	1
第一节 测试的理论与分类	1
案例导入	1
基础知识	1
一、语言测试理论	1
二、测试类型	6
(一)按测试的目的不同	6
(二)按分数解释的不同	9
(三)按测试形式的不同	13
思考与练习	13
案例与分析	14
延伸阅读	14
第二节 测试开发与设计	15
案例导入	15
基础知识	15
一、测试规范编写	16
二、测试规范的样例	20
思考与练习	24
案例与分析	24
延伸阅读	25
第三节 测试试题的编写	25
案例导入	25
基础知识	26
一、多项选择题的编写	27
(一)题干的编写方法	28
(二)选项的编写	29
(三)多项选择题质量检验标准	31
二、其他题型及编写	33
思考与练习	35
案例与分析	35

延伸阅读	35
第二章 听力测试与评估	37
第一节 听力的性质	37
案例导入	37
基础知识	37
一、听力能力的定义和分类	37
二、影响听力理解的因素	38
(一)听力理解能力与语言知识	38
(二)听力理解能力与背景知识	39
(三)听力理解能力与记忆能力	39
三、听力策略的教学	40
(一)培养听力速度	40
(二)培养预测能力	40
(三)培养跨越障碍的能力	42
(四)把握不同听力材料特点	43
(五)培养理解关键信息的能力	44
(六)培养总结概括信息的能力	45
思考与练习	46
案例与分析	46
延伸阅读	46
第二节 听力测试与评估的方法	47
案例导入	47
基础知识	47
一、利用图画测试听力	47
二、多项选择题	49
三、听后写	56
思考与练习	59
案例与分析	59
延伸阅读	59
第三节 听力测试与评估的试题设计	60
案例导入	60
基础知识	60
一、语料的选择	60
二、问题的设计思路	62
三、设计听力测试题目时的注意事项	68
思考与练习	69
案例与分析	70

延伸阅读	70
第三章 阅读的测试与评估	71
第一节 阅读的性质	71
案例导入	71
基础知识	71
一、关于阅读	71
(一) 阅读目的	71
(二) 阅读方式	72
(三) 阅读策略	73
(四) 阅读模式	74
二、阅读教学	77
(一) 影响学习者在阅读测试与评估中表现的因素	77
(二) 针对阅读测试的教学策略	78
思考与练习	82
案例与分析	82
拓展阅读	83
第二节 阅读测试与评估的方法	83
案例导入	83
基础知识	83
一、配对题	83
二、判断题	86
三、词语填空	86
四、句子排序	88
五、选错句	89
六、多项选择题	89
思考与练习	93
案例与分析	93
拓展阅读	93
第三节 阅读测试与评估设计	93
案例导入	93
基础知识	94
一、阅读材料的选择	94
(一) 阅读测试材料的难度	94
(二) 阅读测试材料的代表性	95
二、阅读理解的问题设计	97
(一) 选词填空题及完型填空题的设计	97
(二) 篇章理解(句子、语段、语篇)多项选择题的设计	98

思考与练习	102
案例与分析	103
拓展阅读	104
第四章 口语测试与评估	106
第一节 口语测试与评估的基础	106
案例导入	106
基础知识	106
一、口语测试与评估的理论	106
(一)关于口语能力	106
(二)口语测试理论	109
二、口语测试与评估的设计与实施	110
(一)设计口语测试与评估的基本原则	110
(二)口语测试与评估的基本形式	111
(三)口语测试与评估的评分	112
三、针对口语测试与评估的教学策略	114
思考与练习	115
案例与分析	115
延伸阅读	116
第二节 口语测试与评估的方法	116
案例导入	116
基础知识	117
一、口语测试与评估的题型	117
(一)大规模考试中的口语测试题型	117
(二)其他口语测试题型	121
二、口语测试与评估的试题设计	125
三、口语测试与评估的评分标准	127
(一)整体评分与分项评分	127
(二)评分标准的制定	136
(三)口语测试的反馈	137
思考与练习	138
案例与分析	139
延伸阅读	140
第五章 写作测试与评估	142
第一节 写作测试与评估的理论	142
案例导入	142
基础知识	142

一、写作能力的定义.....	142
二、写作教学策略.....	144
(一)影响学习者写作测试与评估表现的因素	144
(二)写作教学策略	146
思考与练习.....	148
案例与分析.....	149
拓展阅读.....	149
第二节 写作测试与评估的方法.....	149
案例导入.....	149
基础知识.....	150
一、两种模式：间接测试与直接测试.....	150
(一)间接测试	150
(二)直接测试	150
二、写作中的常见题型.....	151
(一)HSK 中的题型	151
(二)其他题型	154
思考与练习.....	157
案例与分析.....	157
拓展阅读.....	158
第三节 写作测试与评估的设计.....	158
案例导入.....	158
基础知识.....	159
一、设计写作测试题的要点.....	159
二、具体的题目设计.....	161
(一)面向初级水平的学习者	161
(二)面向中高级水平的学习者	161
三、写作评分标准的设计.....	162
(一)写作的评分及评分标准	162
(二)评分标准的制定	166
思考与练习.....	169
案例与分析.....	170
延伸阅读.....	170
第六章 测试成绩分析与报告.....	171
第一节 基本成绩分析.....	171
案例导入.....	171
基础知识.....	171
一、量表.....	171

二、集中量数	173
三、差异量数	178
思考与练习	179
案例与分析	180
延伸阅读	180
第二节 成绩分析报告	181
案例导入	181
基础知识	181
一、描述性统计分析	181
(一)偏态和峰度	182
(二)正态分布	183
二、相关分析	185
三、导出分数	189
思考与练习	190
案例与分析	191
延伸阅读	191
第七章 信度与效度	193
第一节 信度	193
案例导入	193
基础知识	193
一、经典测量理论及信度估计	193
(一)信度的理论基础	193
(二)信度的分类	195
二、项目反应理论及信度估计	199
思考与练习	204
案例与分析	204
延伸阅读	204
第二节 效度	205
案例导入	205
基础知识	205
一、关于效度	205
二、效度研究的框架	207
(一)基于内容的证据	207
(二)基于反应过程的证据	209
(三)基于内部结构的证据	209
(四)基于与其他变量之间关系的证据	212
(五)基于测验结果的证据	214

三、影响效度的因素	215
思考与练习	215
案例与分析	215
延伸阅读	216
第八章 题目分析	218
第一节 基于经典测量理论的试题质量分析	218
案例导入	218
基础知识	218
一、试题质量分析的核心概念	218
二、题目分析软件示例	221
思考与练习	224
案例与分析	225
延伸阅读	226
第二节 基于项目反应理论的试题质量分析	227
案例导入	227
基础知识	227
一、试题质量分析的核心概念	227
二、题目分析软件示例	230
思考与练习	235
案例与分析	235
延伸阅读	236
第九章 标准化测试	237
案例导入	237
第一节 国家汉办的标准化汉语测试	237
基础知识	237
一、HSK 简介	237
二、HSKK 介绍	244
三、YCT 简介	246
四、BCT 简介	251
思考与练习	260
案例与分析	260
延伸阅读	261
第二节 其他国家的标准化汉语测试	261
案例导入	261
基础知识	261
一、SAT Chinese with Listening Subject Test	261

(一)听力理解	262
(二)运用	263
(三)阅读理解	263
二、The AP Chinese Language and Culture Exam	264
三、中国语检定考试	271
四、IGCSE	276
(一)IGCSE 的课程大纲、课程目标及测试目标	276
(二)IGCSE 的考试简介	278
(三)小结	289
思考与练习	290
案例与分析	290
延伸阅读	291

第一章 语言教学、测试与评估

第一节 测试的理论与分类

【案例导入】

郑老师今年刚到国外一所中学孔子学院课堂任教。期末考试即将来临，郑老师按照该中学的要求完成了试卷设计，然而中文部的教学主管对郑老师所设计的试卷颇有微词。他认为郑老师尚未完全理解该中学一直以来秉持的教学理念，郑老师所设计的试卷体现了传统语法翻译法的理念，要求郑老师根据该学校的语言教学特点重新斟酌设计。郑老师该如何进行调整呢？

【基础知识】

一、语言测试理论

由于语言由不同的技能或能力组成，使用中的语言呈现出十分复杂的现象。在设计语言测试前，就需要对由各种语言能力组成的语言测试模型及各语言能力之间的相互关系进行定义。这个模型不必与我们脑中所谓的语言能力模型一致，尽管也许可能是一致的。这个模型的主要功能在于对测验设计者所重视的语言能力特点进行说明。对于确定哪些语言能力能或者应该被测试来说，这仅仅只是起点而已。这对于保证测试结果的可解释性及有用性也是重要的。测试模型所确定的心理特征也叫做特质(trait)或构念(construct)。

受语言学、语言教学和心理测量学的影响，语言测试大致经历了以教什么就测什么为特征的前科学语言测试阶段；强调标准化和客观化的心理测量学——结构主义阶段；强调单一语言能力和提倡综合测试法的社会语言学阶段；以及80年代以后流行的以测试语言运用能力为目标的交际语言测试阶段。90年代以来，交际教学法不再像过去那样时髦，基于任务的语言教学渐得人们的追捧。受此影响，基于任务的语言测试(task-based language assessment)随之产生。

这三个阶段也被称为前科学阶段、心理测量与结构主义相结合的阶段和心理语言学与社会语言学为基础的阶段。这一发展过程与语言学理论的发展过程，即从历史比较语言

学、结构主义语言学、生成-转换语法理论发展到心理语言学和社会语言学的过程基本一致。它与外语教学法的发展过程即从语法翻译法、听说法到交际法的发展过程也是一脉相承的。从时间上看,这反映了后者对前者的影响。

1. 前科学阶段测试

20 世纪 40 年代以前的语言测试统称为前科学语言测试。在此时期,语言教师对于什么是语言这个最基本的问题没有科学的认识,只是把语言作为一门包括语法知识、词汇知识、语音知识的符号体系去教,把语言看作一门知识,是这个时期外语教学和测试体系的语言观的内涵。测试中对技巧、技能的考查没有量度依据,测试结果主要依据教师的主观判断。传统的英语测试方法十分重视语法规则、词形变化和词语的用法,通常考语法、词汇、阅读和翻译,重视书面用语,一般只有笔试,没有口试,没有听力题。命题人不大注意测试的可信性、客观性、科学性之类的问题,不懂得用统计学的方法来收集、处理数据。

传统的英语测试常见的题型有填空,以英语为例,例如填写名词单复数形式、定冠词和不定冠词,动词的时态、语态;语法分析,例如在句子中某个语法成分下面画线,指出句子是简单句还是复合句;改写句子,例如改句子中的人称代词,把动词从一般过去时变成过去时,用学过的词语替代句子中的某个词语;翻译,例如把英语翻译成母语,或者把母语翻译成英语;书面答问,例如回答有关课文内容的问题;缩写,将一篇较长的文章缩写成短文,或用一段文字来阐述短文内容以及命题作文等。听力、口语被认为是附加的,不是测试的必要组成部分。

2. 心理测量—结构主义语言测试

心理测量—结构主义测试(the psychometric-structural approach)是语言测试史上的第一个现代测试模型,萌芽于 20 世纪早期,成熟于 20 世纪 50、60 年代。

20 世纪 40 年代至 70 年代中期,人们开始意识到:关于语言规则知识的学习并不等于语言学习。以布龙姆菲尔德(Bloomfield)为首的美国结构主义语言学家第一次提出语言是一套形式结构,是一套习惯。以斯金纳(Skinner)为首的美国行为主义心理学家认为:语言行为是一连串的刺激—反应过程。因而,语言就是要获得操作这套符号系统的技能,训练对刺激做出正确反应的一套语言习惯。

在这样的理论基础上,形成了心理测量—结构主义语言测试。心理测量—结构主义语言测试属于能力测试,以推测应试者的语言能力,并根据推测出的能力再推测应试者在非测试环境中的语言运用为目标。在这一测试模型中,主要把语言作为一个形式系统来检测,因此称作系统参照式(system referenced)测试。其假设为可分性语言能力假设,即语言可以分解成语言技能(包括听、说、读、写技能)和语言成分(包括语音、语符、语法、词汇)。这一语言测试模式提倡分离式测试,主张应该考查语言的应用,而不是有关语言的问题;考查的结构,应该是口语中确实使用的结构。基于语言测试的设计理念,分离式测试通常把语言测试按照语言技能分成若干部分,例如听、说、读、写四个部分,分别进行测试。由于语言构成在不同方面有层次之分,语言测试也分方面和层次。例如,可以分成语音、词汇、句子、短文等,或者分成识记、理解、应用、分析、综合和评估。由于语言项目数目巨大,而一次语言测试所能包括的测试项目有限,因此抽样就必不可少。考生

运用测试中语言项目的能力构成其测试语言能力,为检测考生的测试语言能力需要设计试卷。因此,在设计试卷前,命题人就可以根据需求来确定怎么考,考什么。在命题之前,命题人已确定了试卷的内容比重、范围及层次。这也使得在分离式测试试卷中,通常各题在内容方面是孤立的,每一个题都有其具体的目标。

这一语言测试模式的早期倡导者多为美国的语言测试专家,以 R. Lado、J. Carroll、D. Harris 为主要代表人物。由美国教育考试服务中心(ETS)开发,于 1964 年开始实施的托福(TOFEL)英语水平考试是其测试实施的典型代表。这也使得在这一语言测试模型的发展之下,语言测试形成了设计、命题、评分、分析的完整科学流程,开始讲究测试的科学性,注重应用统计学对题项及整份试卷进行评价。应当说,现有的语言测试中,采用分离式测试的较多。

3. 心理语言—社会语言学测试

从 20 世纪 70 年代中期起,语言测试界开始重视所谓总体综合法的研究,这一时期被称为心理语言学—社会语言学时期。实际上,综合法是针对采用以分离式测试为主的方法而提出来的。有的批评者就来自于心理测量学—结构主义语言测试的内部。他们呼吁在语言测试中不仅有分立式试题,还有像作文、听写等综合式试题。分离式测试把语言技能加以分解,而综合测试正好相反,认为语言不但是一个可以分解的体系,更是一种动态的、具有创造性的功能体系。语言使用过程中冗余度很大,因此不能说缺少了哪一个语言点语言交际就无法实现。使用语言不但要能够按照语法规则构造出合格的句子,而且还必须具有在不同的语境中合理地使用这些句子的能力,这就对语言测试的效度提出了新的要求。这一点是孤立地测试语言点的分立式试题做不到的。

针对心理测量学—结构主义语言测试,有批评者质疑其作为测试目标的可分性语言能力和主要测试手段的客观性分立式试题。其中的代表人物是美国教育家 J. Oller(1979)。他提出“整体语言能力假设”,主张使用完形测试(cloze test)和听写测试(dictation tests)取代其他测试方法。Oller 使用了当时流行的实证研究统计方法因素分析技术(factorial-analytic technique)和主成分分析(principal component analysis),他认为影响语言测试成绩的诸因素中存在一个主要成分来作为其整体语言能力假设的实证证据。

这一时期采用得比较多的是综合题,如完形填空、综合改错、听写、口试、作文等。综合法的主要问题是难以做到评分的客观性和一致性。虽然有些综合题型采用客观题形式,如完形填空等,但采用客观题形式后所考核的仍然只是理解能力,对考核综合语言能力仍有一定的局限性。

4. 交际语言测试

尽管心理测量—结构主义语言测试与心理语言—社会语言学测试看起来针锋相对,但实际上两者仍有相似之处。这两类测试都是能力测试,都是间接测试。这是因为这两类测试中应试者的测试行为和人们的交际行为完全不同,却通过应试者的非交际性语言运用行为来推测其语言水平。也就是说,这两类测试并不能说明应试者语言与应用的情况。交际语言测试理论的产生和发展与语言学理论的发展密切相关。社会语言学中的语言功能、语言变异和社会文化原则,功能语言学中的系统功能原则以及语用学中的言语行为原则等为交际语言测试理论的产生提供了总的理论根据。交际能力理论为交际性语言教学与测试奠

定了直接的理论基础。

20 世纪 60、70 年代,形式语言学继续发展,但研究者的兴趣从语言形式扩展到了语言交际。乔姆斯基(Chomsky)在 1965 年提出了语言能力和语言行为的概念。1972 年,社会语言学家海姆斯(Hymes)提出了交际能力的概念,认为交际能力包括四个社会文化因素,即语法的正确性、语言的可行性、语言的得体性和语言的可接受性。在 Hymes 交际能力概念的基础之上,Canale 和 Swain 在 80 年代初提出了一个既具有语言学基础又符合教学实际的交际能力理论模式。该模式将交际能力定义为至少包括知识与技能的四个方面:(1)语法能力:这种能力涉及的是正确理解和表达所说话语的字面含义所需的语言知识;(2)社会语言能力:包括说出和理解语言的能力,而这些话语及其内容在语言环境中必须是恰如其分的;(3)话语能力:即将意义与不同类型的既连贯又可接受的口头和书面内容融合在一起的能力;(4)策略能力:指在交际过程中的应变能力。该模式于 80 年代在应用语言学界颇具权威。

1990 年 L. F. Bachman 在 *Fundamental Considerations in Language Testing* 一书中提出的交际能力模型,是里程碑式的交际语言能力模型,是交际英语测试设计的理论基础。Bachman 认为交际语言能力(communicative language ability, CLA)由三部分构成:语言能力、策略能力和心理生理机制。语言能力包括以下组成部分:(1)语言组织能力(organizational competence):①语法能力(grammatical competence),包括词汇、词法、句法、语音学、书写。②语篇能力(discourse competence),包括词语联结知识、修辞知识。(2)语言使用能力(utilizing competence):①功能能力(illocutionary competence),包括达意功能、操纵功能、探索功能、想象功能。②社会语言能力(sociolinguistic competence),包括方言或变体知识的敏感性、语域知识的敏感性、自然地地道语言的敏感性、文化典故和比喻知识。策略能力是在语言交际时运用各种语言知识的心理能力。心理生理机制是语言交际时作为物质现象如声、光等所涉及的心理和神经活动过程。

韦尔(Weir)详细阐述了交际语言测试的理论原则,同时介绍了交际语言测试题目类型。交际语言测试的主要特征是:①重点是意义;②语境化;③语言活动带有可接受的目的性;④同真实情况一样的语言;⑤使用的语言材料;⑥文本处理有真实性;⑦反应结果不可预见性;⑧以互动为基础;⑨考生在真实心理状态下展示语言能力;⑩根据交际结果判断成绩。

交际语言测试是从交际法教学方法发展起来的。交际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运用语言的能力上,即要着重了解受测者用目标语进行交际的能力。交际语言测试是运用测试(performance test),即根据应试者的测试行为直接推测其在非语言测试环境下运用语言的情况的测试。

交际语言测试一般在话语这一层次,而不是在单词或者句子层次。这样做,便于设置所需要的语境,迫使受测者要联系上下文考虑问题。交际语言测试运用较多的主观题型,如面试、作文等。这些题型有较高的交际性,但信度相对较低。如果说心理测量学—结构主义测试主要依赖于命题人的专业水平以保证测试质量,而交际语言测试更多地依赖的则是评分员的专业水平以保证测试质量,这也使得测试过程复杂,耗时