

张凯 主编

语言测试概论



审订 世界汉语教学学会
总主编 赵金铭 齐沪扬 范开泰 马箭飞

商务馆对外汉语专业本科系列教材



商务
印书馆

013068444

H195

15

商务馆对外汉语专业本科系列教材

总主编 赵金铭 齐沪扬 范开泰 马箭飞

审订 世界汉语教学学会

语言测试概论



YUYAN CESHI GAILUN

语言测试概论

张凯 主编



商务印书馆

2013年·北京



北航

C1675944

H195
15

444880810

图书在版编目(CIP)数据

语言测试概论/张凯主编. —北京:商务印书馆,
2013

(商务馆对外汉语专业本科系列教材)

ISBN 978-7-100-09850-2

I. ①语… II. ①张… III. ①汉语—测试—对外汉语
教学—教学参考资料 IV. ①H195

中国版本图书馆 CIP 数据核字(2013)第 047326 号

所有权利保留。

未经许可,不得以任何方式使用。

YǔYán Cèshì Gāilùn

语言测试概论

张凯 主编

商务印书馆出版

(北京王府井大街36号 邮政编码 100710)

商务印书馆发行

北京瑞古冠中印刷厂印刷

ISBN 978-7-100-09850-2

2013年8月第1版

开本 787×960 1/16

2013年8月北京第1次印刷

印张 13 1/2

定价: 29.00 元

内容简介

本教材吸收了语言测试领域的许多新成果,从语言测试的分类、语言测试的编制和开发、语言测试的试题类型及特点、语言测试的信度和效度等方面系统介绍了语言测试的基本理论问题和比较成熟的技术问题,对语言测试的实现过程进行了全面的阐释,并对国内外一些规模和影响都较大的语言测试进行了介绍。全书将理论与实践相结合,通过大量实例对语言测试的相关理论和实际操作做了深入浅出的阐释,具有很强的针对性和可读性。通过此书,读者可以对语言测试的基本理论和方法有比较系统的了解。

本书是对外汉语专业本科学生使用的教材,同时也可供大专院校应用语言学专业的师生和其他汉语作为第二语言/外语教学工作者、研究者参考。

主 编 张 凯

编 者 (按音序排列)

陈 宏

龚君冉

郭树军

韩 阳

黄霆玮

李 慧

张 凯

赵琪凤

前 言

对外汉语教学专业的设立已经有二十多年的历史了。早在 1983 年经教育部批准北京语言学院在外语系内就设置了对外汉语教学专业,以培养对外汉语教师为主要目标。不久,北京外国语大学、上海外国语学院和华东师范大学也相继开设了类似的专业。

此后几年,该专业一直踽踽独行,没有名目。直至 1988 年,教育部颁布《普通高等学校本科专业目录》和《普通高等学校本科专业设置规定》,在一级学科中国语言文学类(学科代码 0501)下,设“对外汉语”(学科代码 050103)二级学科,这一专业才正式确立。

当初,设置这一专业,是为招收第一语言为汉语的中国学生,培养目标是将来能从事对外汉语教学及中外文化交流等工作。故该专业特点是,根据对外汉语教学对教师知识结构和能力的要求设计课程和确定教学内容。在 1989 年“对外汉语教学专业会议”(苏州)上,进一步明确了这个培养目标,并规定专业课程应分为三类:外语类、语言类和文学文化类。1997 年召开“深化对外汉语专业建设座谈会”,会议认为,根据社会需要,培养目标可以适当拓宽,要培养一种复合型、外向型的人才,既要求具有汉语和外语的知识,又要求有中国文化的底蕴;既要求懂得外事政策和外交礼仪,又要求懂得教育规律和教学技巧。这一切只能靠本专业的独特的课程体系、有针对性的教材以及特定的教学方法才能完成。

近年来,世界风云变幻,中国和平崛起。随着汉语加快走向世界,对外汉语教学事业获得蓬勃发展。目前开设对外汉语专业的高等学校已有一百三十

多所。大发展带来了丰富多彩,也伴随着不规范。对外汉语作为一个专业,既无统一的教学大纲,也无标准的课程设置,更无规范的教材。在业内对对外汉语教学的学科内涵,也还存在着不同的认识。目前,设立本专业的院校只能本着各自的理解,依据本单位的教学资源与教学条件设置课程,自编或选用一些现成的教材。

有鉴于此,在国家汉办的指导下,商务印书馆以其远见卓识,决定组织全国各高校对外汉语教学资深人士,跨校协商,通力合作,在初步制订专业课程大纲的基础上,编写一套对外汉语专业系列教材,以适应目前本专业对教材的迫切需求。

本教材以赵金铭、齐沪扬、范开泰、马箭飞为总主编,教材的编者经多次协商讨论,决定本着下列原则从事编写:

一、总结以往的经验,积成多年来对外汉语教学成果,以课程在教学计划中的地位、性质、任务和作用为依据,规定课程的基本内容,划定教学范围,确立教学要求。

二、密切关注语言学,特别是汉语语言学研究的最新进展,全面吸取汉语作为第二语言/外语教学研究的最新成果,着重体现语言规律、语言教学规律和语言学习规律。

三、教材的教学内容力求贯彻“基础宽厚,重点突出”的原则,注重基本理论、基本知识和基本技能,既要加强基础理论的教学,更要加强实践能力的培养。对课程的实践性教学环节应有明确、具体的要求,并具有较强的可操作性。

四、教材要全面显示汉语作为第二语言/外语教学的性质、特点和规律,为加快汉语走向世界,为汉语国际推广,培养外向型、复合型的人才。

五、谨守本科系列教材的属性,注意教材容量与可能的课时量相协调,体现师范性,每一章、节之后,附有思考题或练习题。特别要注意知识的阶段性衔接,为本一硕连读奠定基础,留有空间。

基于上述考虑,我们对对外汉语专业的教学内容作了权衡与取舍。本着培养目标所要求的内涵,教材内容大致围绕着四个方面予以展开,即:基础知识、专业知识、教学技能和教师素质。我们把拟编的对外汉语专业本科系列教

材组成五大板块,共 22 册。每个板块所辖课程及教材主编如下:

一、语言学、应用语言学和汉语

- | | |
|-------------|-------------|
| 1. 现代汉语 | 齐沪扬(上海师范大学) |
| 2. 古代汉语 | 张 博(北京语言大学) |
| 3. 语言学概论 | 崔希亮(北京语言大学) |
| 4. 应用语言学导论 | 陈昌来(上海师范大学) |
| 5. 汉英语言对比概论 | 潘文国(华东师范大学) |

二、中国文学文化及跨文化交际

- | | |
|-------------|-------------|
| 6. 中国现当代文学 | 陈思和(复旦大学) |
| 7. 中国古代文学 | 王澧华(上海师范大学) |
| 8. 中国文化通论 | 陈光磊(复旦大学) |
| 9. 世界文化通论 | 马树德(北京语言大学) |
| 10. 跨文化交际概论 | 吴为善(上海师范大学) |

三、汉语教学理论、第二语言习得理论与实践

- | | |
|----------------|--------------|
| 11. 对外汉语教学导论 | 周小兵(中山大学) |
| 12. 第二语言习得研究 | 王建勤(北京语言大学) |
| 13. 对外汉语本体教学概论 | 张旺熹(北京语言大学) |
| 14. 对外汉语教学课程论 | 孙德金(北京语言大学) |
| 15. 双语与双语教育概论 | 关辛秋(中央民族大学) |
| 16. 华文教学概论 | 郭 熙(暨南大学) |
| 17. 世界汉语教育史 | 张西平(北京外国语大学) |

四、对外汉语教材、教学法与测试评估

- | | |
|----------------|-------------|
| 18. 对外汉语教学法 | 吴勇毅(华东师范大学) |
| 19. 对外汉语教材通论 | 李 泉(中国人民大学) |
| 20. 语言测试概论 | 张 凯(北京语言大学) |
| 21. 对外汉语教学模式概论 | 马箭飞(国家汉办) |

五、现代教育技术在对外汉语教学中的应用

- | | |
|----------------|-------------|
| 22. 对外汉语教育技术概论 | 郑艳群(北京语言大学) |
|----------------|-------------|

本系列教材主要是为对外汉语专业本科生编写,也可供其他对外汉语教学工作者、研究者参考,同时也可以作为大专院校语言文学类专业的课外参考书。

目前,汉语国际推广正如火如荼,汉语作为第二语言/外语教学也面临着巨大的机遇与空前的挑战。我们愿顺应时代洪流,为汉语国际推广尽绵薄之力。大规模、跨地区、跨学校地组织人力进行系列教材的编写,尚属首次,限于水平,疏忽和不妥之处在所难免,敬祈专家、读者不吝指正。

(学大旦夏)解恩桐

(学大旦职新工)李海平

(学大旦夏)潘张莉

(学大育部京北)靳树昌

(学大旦职新工)徐成良

(学大山中)刘小慎

(学大育部京北)陈彦生

(学大育部京北)袁卫华

(学大育部京北)金蔚冰

(学大育部京北)冯羊关

(学大南夏)照一萍

(学大西国长京北)平西华

(学大旦职新工)魏良昊

(学大育部京北)梁一萍

(学大育部京北)周一华

(学大育部京北)刘一萍

李文科 赵金铭 齐沪扬

2007年6月5日

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

学大旦职新工

目 录

绪论	1
第一章 语言测试的分类	6
第一节 按用途分	6
一 水平测验	6
二 成绩测验	7
三 能力倾向测验	7
四 分班测验	8
五 诊断测验	8
第二节 按测验方式分	9
一 分立式测验和综合式测验	9
二 笔试和口试	10
三 纸笔测验和计算机辅助测验	12
第三节 按评分方式分	13
一 主观测验	13
二 客观测验	14
第四节 其他分类	14
一 常模参照性测验和标准参照性测验	14
二 标准化测验和非标准化测验	15
第二章 语言测试的编制和开发	17

第一节 规划和设计	18
一 确定目标团体	18
二 确定测验目的	20
三 确定测验内容	22
四 确定测验方式、卷面构成和指导语	23
五 确定评分方式和分数体系	26
第二节 操作和实施	28
一 编写测验说明和题目细则	29
二 训练命题员	32
三 命题和拼卷	35
四 聘请专家组对测验进行审核	36
五 预测和实测	37
第三节 分析和检验	38
一 题目分析和修改	38
二 信度、效度分析	39
三 后效研究	39
第三章 试题	41
第一节 客观性试题	41
一 多项选择题	41
二 是非题	50
三 配伍题	51
第二节 半客观(半主观)性试题	55
一 完形填空题	55
二 成段改错题	60
第三节 主观性试题	61
第四章 预测、选题、拼卷	63
第一节 预测的要求和方式	63

一 被试样本	63
二 预测的条件	64
三 预测的方式	65
第二节 题目分析	67
一 难易度	67
二 区分度	73
三 选项的分布	84
四 题目参数的二重性	86
第三节 题目的修改	87
第四节 选题、组卷、施测	90
一 选题标准	90
二 组卷和施测	93
第五章 分数和信度	95
第一节 分数	95
一 原始分数	95
二 百分等级	96
三 Z 分数	97
四 导出分数	98
五 正态化 Z 分数及导出分数	101
第二节 信度	103
一 信度的基本概念	103
二 测验信度的估计	105
三 影响测验信度的一些因素	109
第六章 效度	111
第一节 效度的基本概念	111
一 测量的有效性	111
二 理论的有效性	112

4 语言测试概论

三 应用的有效性	113
第二节 内容效度	115
一 内容效度检验的基本特征	116
二 内容效度检验的一般程序及定量方法	117
三 内容效度本身的局限	118
第三节 效标关联效度	119
一 调查效标关联效度的步骤	120
二 共时效度和预测效度	121
三 效标关联效度的计算	123
四 效标关联效度的局限	124
第四节 构想效度	125
一 什么是构想	126
二 检验构想效度的经验性方法	130
第七章 项目反应理论和概化理论	139
第一节 项目反应理论	139
一 项目反应理论的出现	139
二 项目反应理论的基本原理	141
第二节 概化理论	155
一 概化理论的出现	155
二 几个基本概念	156
三 概化理论的基本方法	160
第八章 国内外几种语言测试	163
第一节 托福	163
一 托福考试的性质和用途	163
二 试卷结构和题型设计	163
三 分数体系	168
第二节 托业	169

一 托业考试的性质和用途	169
二 试卷结构和题型设计	170
三 分数体系	173
第三节 雅思	174
一 雅思考试的性质和用途	174
二 试卷结构和题型设计	175
三 分数体系	177
第四节 汉语水平考试	177
一 汉语水平考试的性质和用途	177
二 试卷结构和题型设计	178
三 分数体系	184
四 考试规模	187
五 发展趋势	188
第五节 实用汉语水平认定考试	188
一 实用汉语水平认定考试的性质和用途	188
二 试卷结构和题型设计	190
三 分数体系	195
四 考试规模	196
参考文献	198
后记	202

绪 论

1961年,Robert Lado 写了一本书,叫作 *Language Testing: the Construction and Use of Foreign Language Tests*,这是全面讨论语言测试问题的第一本专著。从此,语言测试成了应用语言学里的一个专门的研究领域。

虽然 Lado 的书出版于 20 世纪 60 年代,但语言测试方面的实践,在 20 世纪初就已经出现了。20 世纪 60 年代以前的语言测试,除了在技术上有一些革新外,如使用多项选择题,在理论上并无建树。

20 世纪 60 年代,除了 Lado 的那本书外,语言测试的历史上还出了一件事,那就是 TOEFL 的出现。当时,从世界各地到美国求学的国际学生急剧增加,美国的高校行政部门感到,需要一个能够测出外国学生英语水平的测试。为了满足这种需要,应用语言学中心组织了一次研讨,会议最后决定,开发一个新型的语言测试,这就是后来的 TOEFL (Testing of English as a Foreign Language)。研制和开发 TOEFL 的工作交给了 ETS (Educational Testing Service, 美国教育考试服务中心)。

从 20 世纪 60 年代至今,语言测试的实践和理论研究已经走过了半个世纪,然而它也面临着一些困难,这些困难,大致体现在两个方面:一是语言能力问题,一是技术手段的限制。

一 语言能力问题

语言测试是一种测量活动,它所用的工具叫作语言测验。从本质上讲,语言测验是一种测量工具,这和尺子或天平是测量工具一样的。只不过尺子或

天平所测的东西,在物理学家看来似乎是明确的,而语言测验所测的东西却大可商榷。

如果我们问:语言测验测的是什么东西?大家也许会不假思索、异口同声地说:语言能力!但如果我们再问:语言能力是什么东西?历史文献告诉我们,对这个回答的回答是其说不一的。

关于语言能力问题,TOEFL 最初的研制者们(如 Carroll、Harris、Palmer 等)似乎并不认为是个多大的问题,Lado 的那本专著,也没有给予充分的讨论。Lado 及 Carroll 等人认为,语言能力,也就是语言测验所测的东西,无非就是语言的各种要素(如语音、词汇、语法)以及听、说、读、写等语言技能。在那个时候,人们甚至没有为“语言能力”找一个专门的术语,只是用 ability 这样的普通名词来指称。然而没过多久,事情开始变得复杂起来了。

20 世纪 70 年代末,Oller 对语言能力进行了一些研究。Oller 大概是受了 Spearman 关于智力研究的启发。20 世纪上半叶,英国心理学家和统计学家 Spearman 提出,人的智力是由两个因素构成的,一个叫作“一般智力因素”,简称“g 因素”(g 是 general),一个叫作“特殊因素”。一般智力因素与人的所有活动都有关,而特殊因素只与少数特殊的活动有关。智力测验,测的就是人的一般智力因素。

Oller 使用 Spearman 发明的因素分析方法,对语言测验(主要是 TOEFL)进行了大量的研究,然而其结论却不那么让人信服。Oller 用因素分析对测试的数据反复计算,结果显示,他所分析的所有的语言测验都只有一个单一的因素。由于语言测验所测到的只有一个单一的因素,Oller 便认为,所谓的“语言能力”是一种不可再分的单一的能力。Oller 根据他的这个观察结果,提出了他的“单一能力假说”(Unitary Competence Hypothesis, UCH)。除了提出这个假说外,Oller 等人还发现,在所有测验中,语言测验和智力测验的相关最高,根据这个现象,他们甚至认为,语言能力和智力就是同一种能力。

Oller 的“单一能力假说”固然和 Spearman 的“一般智力因素”很像,但多数人从直觉上认为这是不可能的。

20 世纪 70 年代,Hymes 提出了交际能力理论,应用语言学界很快就接受

了这个理论。1980年, Canale 和 Swain 提出了一个交际能力模型; 1983年, Canale 又对这个模型做了改进。到了 20 世纪 80 年代末、90 年代初, Bachman 提出了一个更为复杂的交际能力模型。迄今为止, Bachman 的模型是最复杂的交际能力模型。这个模型提出后, 一些人认为, 该模型已经很完备了, 今后的工作就是如何用统计手段验证这个模型。

Bachman 的模型固然复杂, 看起来也很完备, 但是, 他的整个模型至今仍处于有待验证的状态。不仅 Bachman 的模型是这样, 其他人提出的各种各样的语言能力模型也都没有得到经得起推敲的检验。因此, 语言测试面临的第一个困难, 就是语言能力问题尚未解决。

二 技术手段的限制

语言能力究竟是什么? 语言能力的组成成分究竟是什么? 这种问题固然是高深的理论问题, 但绝不是只靠纯粹的思辨就能解决的。

语言测试的手段, 绝大多数人都很熟悉: 测试者采用书面或口头的形式, 向被试提出一些问题, 这些问题也许是开放性的(如请被试描述一个事件), 也许是封闭式的(如多项选择题); 被试则根据测试的要求回答这些问题。然后, 测试者根据被试答对题数或答题质量给出一个分数。

当然, 现代语言测试不会停留在给出一个分数这一步上, 它还会使用一些科学上常用的方法对整个测验和分数做进一步的分析(这些分析方法后面会介绍), 其分析结果表现为题目的难度、信度、效度等各项指标。通过这些分析, 测试者和被试及社会各界都可以知道测验的质量如何。

语言测试最重要的指标就是效度。问一个语言测试是否有效, 就是问它是否测到了所谓的语言能力; 问一个测试有多高的效度, 就是问它在何种程度上测到了所谓的语言能力。

那么, 如何把一个测试的效度展示给人看呢? 到目前为止, 我们所用的手段极为有限——主要是使用一些多元统计方法。先举一个简单的例子。

假定现在有一个测试, 我们想证明它是有效的, 该怎么办呢? 我们的做法是, 首先找一个已被公认为有效的测试, 然后找一批被试, 让他们把两个测试