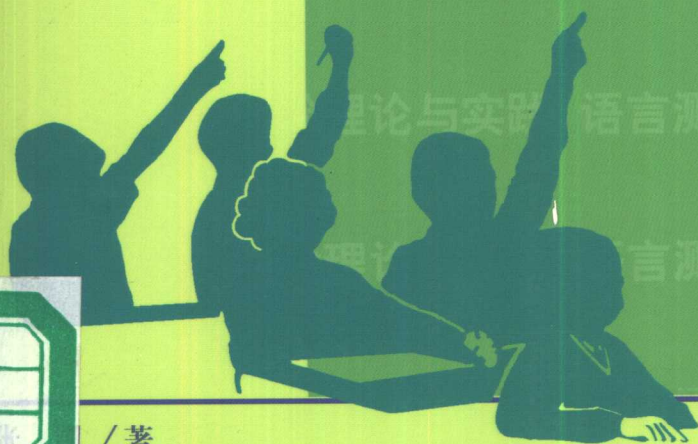


语言测验

理论与实践

LILUN YU SHIJIAN



张 / 著

北京语言文化大学出版社

汉语教师培训教材

语言测验理论与实践

张 凯 著

北京语言文化大学出版社

(京) 新登字 157 号

图书在版编目 (CIP) 数据

语言测验理论与实践/张凯著.

- 北京: 北京语言文化大学出版社, 2002

ISBN 7-5619-1101-7

I. 语…

II. 张…

III. 语言 - 水平考试 - 研究

IV. H0

中国版本图书馆 CIP 数据核字 (2002) 第 045557 号

责任印制: 乔学军

出版发行: 北京语言文化大学出版社

社 址: 北京海淀区学院路 15 号 邮政编码 100083

网 址: <http://www.blcup.com>

<http://www.blcu.edu.cn/cbs/index.htm>

印 刷: 北京北林印刷厂

经 销: 全国新华书店

版 次: 2002 年 9 月第 1 版 2002 年 9 月第 1 次印刷

开 本: 850 毫米×1168 毫米 1/32 印张: 6.125

字 数: 130 千字 印数: 1-3000 册

书 号: ISBN 7-5619-1101-7/H·02087

定 价: 10.00 元

发行部电话: 010-82303651 82303591

传真: 010-82303081

E-mail: fxb@blcu.edu.cn

出版说明

随着中国经济的快速增长及中国加入世贸组织，汉语的重要性已经为越来越多的人所认识。无论是对外汉语教学还是国内少数民族的汉语教学，近年来都得到了迅猛发展，教学规模不断扩大，学生人数迅速增加，这就要求有一支懂得第二语言教学规律的高素质的教师队伍。但目前从业的汉语教师大多数没受过汉语作为第二语言教学的专业教育，在教学实践中沿袭的是母语文教学的老路子，这种情况远远不能适应汉语教学发展的需要。因此，对现有汉语教师进行在职培训是解决这一问题的切实可行的办法，也是当务之急。

近十年来，汉语作为第二语言教学的学科建设不断加强，理论研究不断深入，汉语教学理论、语言习得理论和作为第二语言学习和教学内容的汉语研究都取得了丰硕成果，这些都为汉语教师通过培训学习进一步提高素质，并运用理论知识提高教学效率和教学质量提供了可能和保证。

汉语作为第二语言教学是不同于母语文教学的专门学科，这一点早已成为人们的共识。汉语作为第二语言教学的教师也应该具备不同于从事母语文教学教师的知识结构。具体来说，21世纪的汉语教师不仅要具备扎实的汉语基础知识，熟悉教学法并能熟练运用课堂教学技巧，还要懂得第二语言教学理论、语言习得理论、语言测验理论等，并能运用这些理论知识来指导教学实践。应该说，不

具备这些知识的教师，肯定不能胜任今后的汉语教学工作。

正是基于以上认识，北京语言文化大学出版社出版了这套汉语教师培训系列教材，第一批共5本，包括《现代汉语概论》、《汉语课堂教学技巧》、《语言测验理论与实践》、《汉语作为第二语言教学简论》和《汉语普通话语音教程》（新疆版）。读者对象为从事汉语作为第二语言教学的教师，包括对外汉语教师、对国内少数民族学生进行汉语教学的教师及海外汉语教师。这套教材除个别是在原有基础上改编之外，大多为新著，其内容反映了最近几年来汉语教学领域的一些新的研究成果，它们既有侧重本学科基础理论的，也有侧重具体教学技巧的，基本涵盖了汉语作为第二语言教学教师所应具备的主要知识。特别值得一提的是，这套教材的作者都是既有多年的汉语教学经验，又在本学科的理论研究上颇有建树的专家，相信这套教材会帮助广大汉语教师提高素质、改进教学，积极应对21世纪的新挑战。

北京语言文化大学出版社

2002年6月

使用说明

本书是介绍语言测验的入门教材，可供应用语言学专业的本科生和研究生使用，也可供汉语教师培训之用。本书共有九章，如果把它当作课堂讲授的底本，讲授这门课约需 40 学时。如果把它当作自学教材，读者应该在读完本书后，按书后的参考文献，扩大自己的读书范围。

语言测验是一个理论和实践并重的领域，无论是课堂讲授，还是自学，学习者都不要死记书中的定义和公式，而应该寻找机会亲自实践，把书中的内容应用到自己编制的测验中去，从而加深对语言测验理论的理解和体会。

引 言

语言测验是什么？这是关心语言测验的人首先应该问的问题。答案很简单，语言测验是一种测量工具。测量工具是我们很熟悉的东西，日常生活中往往要用到。尺子，无论是学生尺、木匠尺，还是卷尺、卡尺，都是测量长度的工具。秤，无论是杆秤、台秤，还是弹簧秤、电子秤，都是测量重量（质量）的工具。钟表，无论是闹表、手表，还是电子表、原子钟，都是测量时间的工具。语言测验和那些工具一样，也不过就是一种测量工具而已，它的作用是测量人的语言能力。尺子、钟表等测量工具是我们熟悉的，语言测验虽然也不陌生，但我们很少想到它也是测量工具，尤其是我们往往想不到它是一种重要的科学测量工具——至少对语言学习来说是如此。

开发或编制语言测验的目的是什么？答案也很简单：开发或编制语言测验就是制造测量工具。但是，对于这个说法，很多人可能没有思想准备。作为语言教师，编写测验是常有的事，但很少有人想到，编制测验就是设计和制造测量工具。和制造尺子之类的测量工具不同，编制语言测验要复杂得多。当然，我们不是说制造一把尺子、为尺子找到科学依据、制定科学地使用尺子的规则等就是简单的、轻而易举的事。我们只是说，制造尺子过程中的问题基本上

早已被前人解决了，我们现在只要按一定的规矩去做就行了；语言测验的麻烦在于，在尺子那里已经解决了的问题，大部分在我们这儿还没有解决。因此，编制语言测验，亦即设计、制造一种能够测量人的语言能力的科学工具，基本上是在没有先例的情况下从头做起。

使用语言测验就是使用测量工具，和使用尺子一样。但它也有和使用尺子不一样的地方。由于语言测验的制造过程还鲜为人知，也由于语言测验所测到的东西不是很明确的，我们对这个工具的了解往往就是很肤浅的。有时候，我们并不能正确地使用语言测验这种工具。因此，认识语言测验，不仅是研制语言测验的人的事，也是使用语言测验的人以及跟语言测验有关的人的事。生活中的例子也能说明这一点。作为一个普通人，我们的职业大都不是制造和研究秤这种测量工具的，我们大都也不是以使用这种工具为职业的（如市场上的小贩）。但这并不意味着我们可以不了解秤这种工具。如果我们不会看秤，市场上的小贩很可能搞出缺斤短两的把戏，我们上了当还一无所知。语言教师和语言测验的关系也是这样的。也许你不是语言测验的专家，也许你对编制语言测验没有兴趣，但你却不能不了解语言测验。如果你了解语言测验，当有人误用测验，或对测验分数作出错误解释（这二者都相当于小贩的缺斤短两）时，你就可以识破它。

无论我们把语言测验看成是一项技术，还是一个学科，它都是理论和实践并重的。在开发、研究和使用的语言测验的过程中，我们既要精通相关理论，又要重视实践。说要精通理论，是因为语言测验不仅和语言理论、测量理论、统计理论密切相关，而且语言测验本身就有重要的理论意义。说要重视实践，不仅因为语言测验的开发、使用和研究中包含了大量的实际操作，还因为语言测验是要在社会生活中使用的，因此，它的效果如何，是要在社会实践中检验的。

本书第一章介绍语言测验的基本原理，在这一章，我们主要介

绍语言测验中的基本概念和基本方法，最后还要介绍一下语言测验的分类。第二章将概略地介绍开发和研究语言测验的基本步骤。第三章到第五章，我们将比较详细地介绍语言测验中的操作程序和方法，这包括题目的编写和修改、题目的预测和分析以及施测等实际操作。第七章和第八章主要讨论语言测验中的理论问题，这包括信度和效度理论、分数的合成与解释等。最后一章讨论决策和后效问题。

本书是一本通俗教材，读者对象主要是语言教师，不是专门从事语言测验开发和研究的人，所以，本书的目的可以用“四个初步”来概括：通过本书，读者可以初步了解语言测验的原理和理论，初步掌握一些简单的技术（如命题、题目分析、分数转换等），初步接触语言测验中的深层理论问题，初步关注伴随着语言测验的社会效应。

目 录

引言	(1)
第一章 语言测验的基本原理	(1)
第一节 语言测验的作用和目的	(1)
一、选拔	(2)
二、诊断	(2)
三、评估	(3)
四、研究	(3)
第二节 怎样用科学方法测量语言能力	(4)
一、测量的定义	(4)
二、测量的四个要素	(6)
三、四种测量水平和四种量表	(9)
四、测量对象及其性质	(11)
五、测量工具	(12)
六、测量中心	(12)
七、怎样使不可测的东西变为可测的	(17)
第三节 测验的种类	(22)
一、按用途分	(22)
二、按测验方式或评分方式分	(25)
三、按参照性质分	(29)
四、测验的标准化	(29)
思考题	(30)

第二章 规划、设计和实施	(31)
第一节 规划设计	(32)
一、确定目标团体	(32)
二、确定测验目的	(34)
三、确定测验内容	(35)
四、确定测验方式和卷面构成	(36)
五、确定评分方式和分数体系	(38)
第二节 操作和实施	(39)
一、编写测验说明和题目细则	(40)
二、训练命题员	(43)
三、聘请专家组就有关问题向其咨询	(44)
第三节 施测、信度、效度及后效研究	(45)
思考题	(46)
 第三章 客观性试题的编写和修改	(47)
第一节 客观题的类型	(47)
一、多项选择题	(48)
二、完形填空	(53)
三、配伍题	(53)
四、是非题	(54)
五、半客观性试题	(54)
第二节 适用于不同技能的多项选择题	(55)
一、听力理解试题	(55)
二、阅读理解试题	(58)
三、语法试题	(60)
四、词汇试题	(63)
第三节 题目的修改	(65)

思考题	(66)
第四章 主观性试题的编写	(67)
第一节 口试	(67)
一、口试的方式	(67)
二、口试的题型	(69)
第二节 作文	(73)
一、单句写作	(73)
二、排句序	(76)
三、限制性作文	(77)
思考题	(78)
第五章 预测、题目分析和正式施测	(79)
第一节 什么是预测	(79)
第二节 题目参数的二重性	(80)
第三节 预测的要求和方式	(81)
一、被试样本	(81)
二、预测的条件	(82)
三、预测的方式	(83)
四、评分方式	(84)
第四节 题目分析	(84)
一、难易度	(84)
二、区分度	(85)
三、选项的分布	(90)
第五节 题目的修改	(92)
第六节 选题、组卷、施测	(95)
一、选题标准	(95)

二、组卷和施测	(97)
思考题	(98)
第六章 测验的信度	(99)
第一节 信度的基本概念	(99)
一、误差的四个来源	(100)
二、真分数	(101)
三、三个假设	(101)
第二节 相关系数	(103)
一、相关的基本概念	(103)
二、积差相关系数	(105)
第三节 内部一致性信度	(107)
一、再测信度	(108)
二、复本信度	(110)
三、分半信度系数	(112)
四、K20 公式	(113)
五、阿尔法系数	(114)
第四节 影响信度的因素	(114)
一、被试的同质性	(114)
二、测验的时间限制	(115)
三、信度和题数的关系	(116)
四、信度和题目难度的关系	(117)
五、信度和区分度的关系	(117)
思考题	(118)
第七章 分数、常模和等值	(119)
第一节 分数	(119)

一、原始分和导出分数	(119)
二、标准差和标准分数 (z 分数)	(121)
三、常用的导出分数	(122)
四、百分位	(123)
五、标准误差和分数的解释	(126)
第二节 常模	(130)
第三节 分数的等值	(131)
一、什么是等值	(131)
二、等值的步骤	(133)
三、常用的等值方式	(134)
思考题	(137)
 第八章 测验的效度	(138)
第一节 效度的基本概念	(138)
第二节 内容效度	(139)
一、内容效度检验	(140)
二、一个例子	(141)
三、检验内容效度的定量方法	(142)
四、在语言测验中建立内容效度的困难	(142)
五、内容效度本身的局限	(143)
第三节 效标关联效度	(143)
一、共时效度	(144)
二、预测效度	(145)
三、效度系数的计算	(145)
四、效标关联效度的局限	(147)
第四节 构想效度	(148)
一、什么是构想	(148)

二、为什么要提出构想	(150)
三、构想、测验、理论三者的关系	(151)
四、构想效度的本质	(152)
五、证明构想效度的经验性方法	(153)
六、语言测验处在什么阶段	(155)
思考题	(158)
第九章 决策和后效	(159)
第一节 决策	(159)
一、测量和决策的区别	(160)
二、教学外的决策	(161)
三、教学内的决策	(164)
四、制定标准时应该注意的问题	(166)
第二节 后效	(167)
一、什么是后效	(167)
二、什么是后效研究	(169)
思考题	(170)
参考文献	(171)
后记	(177)

第一章 语言测验的 基本原理

本书的名字叫“语言测验理论与实践”，也就是说，它讲的是语言测验。“测验”，我们往往喜欢把它叫做“考试”。在日常生活和学习中，“期中考试”“期末考试”“汉语水平考试”“托福考试”，都是我们常常提到的。但在这里，我们为什么不叫“考试”而叫“测验”呢？这是我们首先要解释的一个问题。

第一节 语言测验的作用和目的

中国是考试的故乡，考试在中国已有 2000 年以上的历史。说到考试，我们会想起科举考试。科举考试和我们现在说的测验是有区别的。科举考试的目的只有一个：选拔人才。现代的语言测验的作用与目的和传统科举考试是不一样的。**语言测验的作用只有一个：科学地测量出学习者的语言能力。**语言测验是一种测量工具，和任何测量工具一样，它有自己的专门的作用。比如说，秤的作用是

称出重量，尺子的作用是量出长度，手表的作用是量出时间。语言测验也一样，它的作用就是测出“语言能力”。

测量工具的作用和测量目的是有区别的。**所谓测量目的，是我们使用测量工具时要达到的目的。**秤的作用是称重，但为什么要称重呢？在市场上，称重是为了公平交易；在医院里（称体重）是为了了解健康状况；在（举重的）运动场上，是为了找出力气最大的人。由此可见，“称重量”在不同的情况下是有不同的目的的。语言测验的作用只有一个，即测量语言能力，那它的目的有哪些呢？换句话说，我们为什么要知道人的语言能力呢？在以下四种情况下，我们需要知道人的语言能力。

一、选拔

这里所说的“选拔”是广义的。中学生考大学时，大学需要知道他的外语或汉语水平；进了大学（或任何高一级学校）的学生可能是学过第二语言的，但他们的第二语言水平可能是不同的，为了有效地组织教学，我们可能要把水平不同的学生分到程度不同的班级，如高、中、低班；学生毕业的时候，我们可能要知道他的第二语言水平是否达到了毕业所要求的水平；学生走出学校找工作，用人单位一般要挑水平高的人。以上这些都是所谓的“选拔”，即在一群人中选拔（第二语言）好的。在各种各样的选拔中，语言测验为我们提供了必要的信息：谁语言水平高，谁语言水平低。这是使用语言测验的目的之一。

二、诊断

在教学过程中，我们需要知道学生的语言水平是否有提高。经过一个学期（或一个月）的学习，学生的语言水平提高了吗？班里所有人提高的幅度都一样大吗？是否有人退步？这样的信息也要由