



国家汉办1998—2000年科研项目成果丛书

对外汉语教学语音测试研究

毛世桢

主编

叶军

副主编

中国社会科学出版社

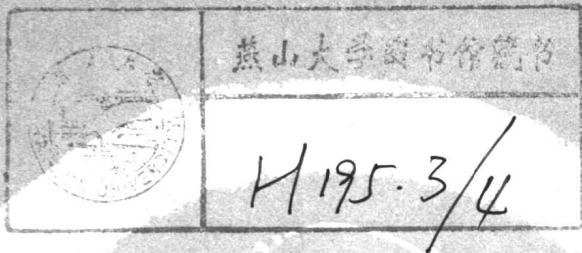
DUIWAI HANYU JIAOXUE YUYIN CESHI YANJIU

国家汉办1998-2000年科研项目成果丛书

对外汉语教学语音测试研究

DUIWAI HANYU JIAOXUE YUYIN CESHI YANJIU

28



05
10

毛世桢 主编 叶军 副主编

中国社会科学出版社



0380927

图书在版编目 (CIP) 数据

对外汉语教学语音测试研究/毛世桢主编. —北京:
中国社会科学出版社, 2002.8

ISBN 7-5004-3474-X

I. 对… II. 毛… III. 汉语 - 语音 - 对外汉语
教学 - 教学研究 IV. H195.3

中国版本图书馆 CIP 数据核字 (2002) 第 049714 号

责任编辑 周用宜
责任校对 张莲芳
封面设计 王 华
版式设计 张汉林

出版发行	中国社会科学出版社		
社 址	北京鼓楼西大街甲 158 号	邮 编	100720
电 话	010—84029453	传 真	010—64030272
网 址	http://www.casspw.com.cn		
经 销	新华书店		
印 刷	北京奥隆印刷厂	装 订	三河鑫鑫装订厂
版 次	2002 年 8 月第 1 版	印 次	2002 年 8 月第 1 次印刷
开 本	850 × 1168 mm 1/32		
印 张	9.625	插 页	2
字 数	250 千字	印 数	1—3000 册
定 价	20.00 元		

凡购买中国社会科学出版社图书, 如有质量问题请与本社发行部联系调换
版权所有 侵权必究

前 言

对外汉语教学，作为一个国家和民族的事业，其肇始可以追溯到中华人民共和国建国之初，而发展成为一个学科则是在1978年“北京地区语言学科规划座谈会”上由吕必松先生首先提出来的。从那时起，应用语言学家和对外汉语教学工作者们戮力奋进，在对外汉语教学学科建设的推动下，卓有成效地促进了该项事业的蓬勃发展。

然而在丰收的喜悦之中，我们不无遗憾地发现语音教学（包括语音测试）成了天鹅群中的一只丑小鸭。有的教学系统回避了语音教学，有的教学系统只进行了应景式的语音教学，有的教学系统虽然进行了探索和尝试，却因科研力度不足，难以获得显著的系统性教学成果。林焘先生近来把他的忧虑之情坦诉出来（参见焦立为“现代语音学的方向——访林焘教授”，《语言教学与研究》2000.2），应该引起我们这些后学者的重视。

语音问题在对外汉语教学系统中本不该被忽视。从教学内容方面来看，它是跟词汇、语法、语用等平起平坐的一个重要的语言要素，它又是跟听、说两项言语技能密切相关的一种语言能力。而从语言系统的角度来看，语音不仅占据了系统中的一个层面，而且它作为语言的结构形式和表达形式，又跟词汇、语法、语用乃至文字都发生紧密的联系。语音教学之所以在各类语言教学系统中不被重视，大抵有两方面的原因。其一是语音问题相对复杂：语音现象是一种稍纵即逝的振动，认知一般只能以“耳

治”的方式实现；现代语音科学被分割为语音学和音系学两门学科，如何将这两方面的研究成果融会到语音教学中去是个难题；语音能力中不仅包含智慧技能还包含动作技能，Stevens 因此认为语音教学“比起语言学来更像体育活动”；学能的差别性在语音上表现得最为明显。其二是学术界的某些观念所起的导向作用：理论语言学家崇尚描写的研究范式，批判所谓的“语言纯正论”（参见弗罗姆金和罗德曼《语言导论》，北京语言学院出版社1994年），其背后的哲学思想是“凡是合理的就是存在的，凡是存在的就是合理的”；一些应用语言学家在倡导交际法的过程中认为，不要求学习者获得“完美的发音”（perfect pronunciation），语音的教学目标似乎只是“可接受性”（acceptability）。

我们对此的态度是，不去纠缠理论观念上的争论，而去研究如何提高对外汉语语音教学的实效。其中相关的问题根据我们初步的考虑应有以下一些方面：语音的教学目标、教学任务、教学内容、教学进程、教学方法、学习者的语音认知和获得规律，等等。

我们首先研究对外汉语语音测试的问题，是出于以下的考虑。

美国著名教育学家 R. 加涅等认为：“教育设计旨在促进各种能力的习得。能力是通过学习成绩的提高体现的。……学生成绩可以显示各种能力是否被习得，因此这种有计划的教学结果是由学生的成绩组成的。……（智慧技能、认知策略、信息、动作技能和态度）这些类别里的成绩目标，如果应用于某一教学课程，还可以做进一步的分析，以便发现其学习所依赖的前提。而这些前提反过来又可以作为确定每节课的顺序、这些课本身设计的基础。”（参见加涅等《教学设计原理》，华东师范大学出版社1999年）可见教育测量不仅是教学系统中不可或缺的一个环节，而且对整个教学系统的设计也至关重要。如果没有一套科学的评

价系统，我们对以往任何一种语音系统的批评都是武断的。即使这种批评实际是合理的，也是缺乏分析性，因而难以令人信服。所以，我们把事情倒过来做，先探索语音测试的各种问题。虽然这样做实际上也是有问题的，因为测试系统本来应该隶属于教学系统，但是任何事情总得找出它的头绪。假如我们把研究的范围主要限制在标准参照性的测试上，而且不妄图解决所有的问题，那么这样的研究应该还是可行的并且是有意义的。

其次，在对外汉语教学系统中已经设计出一套经过实践证明是卓有成效的、通用性的测试系统——HSK（汉语水平考试）。在这套考试系统中测量属性的不平衡主要反映在语音测量的薄弱上。那么，参照 HSK 的设计原则，加强语音属性的测量，也就成为可能。

最后，当前在全国范围内正在推广和实施“普通话水平测试”（PSC）。这套测试系统，在两个方面跟“汉语水平考试”形成对照：（1）HSK 的适应对象是外国学生，而 PSC 的主要适应对象是母语说话者（native speaker）——中国人；（2）HSK 中语音测量属性所占比重最小，而 PSC 中语音测量属性的比重又最大。这两套语言测试系统刚好处于两端，形成互补，这就为我们研究留学生语音测试系统提供了重要的标杆。

我们就是本着这样的想法，投入到对外汉语语音测试的探索中去。其间，承蒙国家对外汉语教学领导小组办公室的信任，立项资助。在整个研究过程中，我们深感自己知识贫乏、功力不足，虽然竭尽全力完成了这个项目的研究，但质量究竟如何、疏漏究竟有几多、错误究竟有多严重，心里颇不踏实。我们就是怀着这种惴惴不安的心情把这份研究结果呈示出来，以求得各位同仁和方家的指教。

另外，在该研究项目的实施过程中，我们还得到华东师范大学社科处、对外汉语系的领导和有关人士的支持和帮助，他们的

支持和帮助是完成此项课题的根本条件。我们还感谢对外汉语系98级研究生胡晓春、99级研究生袁斐、丁玉华、施云、梁建芬同学的积极参与，尤其是梁建芬同学还协助我们做了大量事务性工作。更让我们感佩的是参加试测的那些外国留学生。他们忍受了半个小时的口试煎熬，为我们的对外汉语教学事业做出了令我们永远铭记的贡献。他们是高山千惠美、远藤志保、横田丰初艺、高木友子、池田阳子、小池祈穗、牧坂麻依子、铃木奈绪美、伊藤尚论、永井秀和、小田正范、高惠卿、朴相和、朴钟贤、任少英、赵贞媛、陈咏珊、禹相哲、金大勋、李熙日、许虎镇、陈福盛、陈泰圭、金荣洙、崔闰植、辛东俊、崔允馨、金来玄、刘锐克(Eric)、艾摩尔、帕布尔、Laura，等等。于此也一并致以谢忱！

毛世楨 叶 军
2000年9月4日

目 录

前言	(1)
第一章 语音测试概说	(1)
一、教育测量	(1)
二、语言测试	(5)
三、语音测试	(9)
第二章 语音能力与语音运用	(19)
一、语音能力	(19)
二、语音行为	(27)
三、语音能力与语音运用	(37)
第三章 普通话音系及其运用	(45)
一、普通话的音段音位	(46)
二、普通话的声调	(52)
三、普通话的音节	(54)
四、音联	(56)
五、韵律	(62)
第四章 普通话的音位负载	(66)
一、音位负载及其分析方法	(66)

二、分析数据及讨论	(70)
三、根据音位负载制定教学和测试的对策	(77)
第五章 当前对外汉语教学中的语音教学现状	(80)
一、现行对外汉语初级教材中的语音教学	(80)
二、各类教材中的语音教学法	(100)
三、深度和进度	(107)
第六章 对外汉语语音测试的目标	(116)
一、对外汉语教学语音测试的参照性质	(116)
二、达意和表情	(119)
三、标准、规范	(124)
四、个性化、得体性	(128)
第七章 容错度与口语语音测试	(140)
一、语音容错度的社会心理依据	(140)
二、语音测试中的容错度因素：定量和定性	(142)
三、普通话语音的容错度分析	(144)
四、对外汉语语音测试中有关容错度的处理意见	(152)
第八章 难点的测试和测试的重点	(153)
一、难点的测试	(154)
二、测试的重点	(195)
三、对以英语为母语的应试者的测试重点	(198)
四、对以日语为母语的应试者的测试重点	(198)
五、对以韩语为母语的应试者的测试重点	(199)
六、对有某种汉语方言背景的海外华人、华侨 应试者的测试重点	(200)

第九章 对外汉语语音测试语音错误和语音缺陷分析 ·····	(205)
一、划分语音错误和语音缺陷的理论依据·····	(205)
二、对外汉语语音测试语音错误和语音缺陷的 主要类型·····	(210)
第十章 对外汉语教学语音测试的试题和试卷构成 ·····	(216)
一、各类教材中的语音练习题和测试题·····	(216)
二、语音测试试卷的编制·····	(226)
第十一章 实测和试卷评价 ·····	(244)
一、有关组织实测的一些问题·····	(244)
二、有关试卷编制的一些问题·····	(247)
三、有关评分办法和评分标准的一些问题·····	(250)
四、试卷的信度、效度和区分度分析·····	(253)
五、由测试引发的关于对外汉语语音教学的思考·····	(263)
附录 ·····	(267)

第一章 语音测试概说

一、教育测量

教育测量是教学过程中必不可少的一个环节。所谓“测量”，就是根据某个标准或尺度，对客体进行数量化的赋值。当然，人们进行测量都有一定的目的，对于教学过程来说，测量的目的性更加突出、更加明确。因此，教育测量又跟教育评价交织在一起。B.H. 肖邦（Choppin, 1992）指出，评价和测量的共同点是都要运用测验的手段。我们在此对语音测试进行研究，其间测量和评价是交融在一起的，所以我们将不对测量和评价做刻意的区别。

教育测量，是通过对某个或某些进行特定学习的学习者实施观测，获得所需的数据，以反映、说明、解释学习或教学的成效。所以，教育测试对教学有极其强大的反馈作用（Bachman 称之为“washback”，胡壮麟译作“反溅作用”）。有时，这种反馈作用可能会过于强大，导致目前教育界正口诛笔伐的“应试教育”。然而，应试教育的形成，过失不在测试，而在于教学组织对测试的误解。当然，教育测试也往往对受教育者的学习成果或实际能力加以评定，但是所有的测试都是有限的，总只是在某一方面进行测量并加以评定。社会上的某些人把某项测试成绩当作学生或工作人员的总体评价，从而作为安排社会角色的依据，这

是他们对教育（包括教育测量）的无知。认识这一点在语音测试这方面显得特别重要，因为语音能力跟语言其他层面的能力的相关性相对较小。我们对语音测试进行研究，绝没有要夸大语音测试的作用，恰恰相反，我们试图给语音测试以合理的定位。国际著名教育学专家加涅（Gagné, 1999）指出：“能力是通过学生成绩的提高来体现的。尽管众多的学习是在学校之外进行的，且许多成绩是学生自己努力的结果，学校仍有责任组织并提供指向于具体目标的教学，从而使在缺乏组织的情况下无法实现的目标得以实现。”他的意思是说，测试甚至对学习具有导向作用，使学习目标更加明确而自觉。另外，测试通过组织者（多数是学校）的协调，可以使自学和各种类型的教学形成一致的教学目标。

教育测量包括三个要素：（1）事物的属性，（2）数据，（3）测量的规则和依据（王孝玲，1989）。就语音测试来说，事物的属性是基本确定的，即语音能力。可是细细想来，我们对语音能力的认识还十分有限，还需要开展深入的研究。即使我们能对语音能力有完整的、深刻的理解和认识，也还需要通过有效的测试手段去进行观测。教育测量，说到底是一种心理测量。测试设计者必须通过有效的诱导手段，让应试者完成某些行为，而这些行为为足以反映、显示被试者的能力或心理倾向。就是说，我们必须通过有效的手段来获取适当的行为样本。因此，测试设计是一种科学性和艺术性高度结合的工作。最后，我们还应在语音测试中考虑应该怎样给试题赋值，如何建立量表，也就是说，我们要努力让通过测试得到的数据具有足够的解释能力。

测试根据不同的目标和要求，可以分成各种类型。仅从跟语音测试有关的方面看就有以下一些类型。（1）学能测试，是测量学习者在学习之前的学习能力倾向。正如吕必松教授（1989）所指出的，（我国）“至今还没有进行语言学能测试。在来华学习汉语的学生中，有些人学习第二语言的条件很差，即使主观上很努

力，也难以学好。这部分学生不但使任课教师增加了负担，而且给教学组织工作带来了困难，同时使国家白白浪费了一部分奖学金。要解决这个问题，必须研究和实行语言学能测试”。(2) 成绩测试，也称课程测试或教育测试。这个类型的测试是当前运用得最为广泛的，但是其中不乏经验型的、粗放型的成绩测试。(3) 水平测试，是测量被试者当前的实际能力，而不考虑他是如何获得这种能力的。对外汉语教学已有标准化的水平考试——HSK，但是这个测试（初、中等）缺少语音能力的测量项目。(4) 诊断测试，是为判断被试者在某种能力或能力的某个方面的特殊优点或缺点而设计编制的测试。根据诊断测试的结果，可以有效地调整教学重点，也有利于开展个别化的教学。(5) 标准化测试，是由测试专家精心设计编制的测试，并经过实践证明具有较高信度和效度的测试。(6) 教师自编测试，是由个别或部分教师自己根据教学需要制定的测试。一般说来，这类测试信度和效度不是很高，但理想的自编测试也最好依据或参照标准化测试来设计。(7) 常模参照测试，是“将被评对象的水平进行比较的教育评估”（D. 文森特，1985）。这种测试有两个主要特征：① 试题内容分布在教学内容中；② 具体被试的成绩是建立在与该被试所在群体的平均分数的比较之上的。而这个群体的平均分数就是“常模”。(8) 标准参照测试，“是在跟一组规定明确的能力标准对比之下，对被试人的测验成绩做出解释而制定的”（Popham，1978）。这种测试的关键在于建立一个测量领域规范。那么，试题内容就应落实在这个领域中。而最后成绩的评定也应跟这个领域规范加以对比来得出。此外，测验还可分为个别测试和集体测试，客观评分测试和主观评分测试，笔试和口试，等等。每一种测试都有其特定的适用范围，也都有其特定的测试原则和操作方法。我们所做的对外汉语语音测试研究工作，不可能兼顾各种类型的测试。

任何测试都需要编制试卷来实施，所以在测试的设计中试卷的构成是一项重要的具体任务。试卷的构成，首先在内容上必须覆盖所有的测量目标，而且要给具体的测量目标安排适量的试题。试卷的构成，还应考虑到测试的可操作性：一方面要编制有足够长度的试卷以确保测试的信度，另一方面也要考虑到平抑测试的成本而压缩试卷的长度。每一份试卷，都包含一定数量的试题。由于所有的能力测量都具有一定的综合性，这些试题往往呈现出不同的题型，这样方能获得必要的行为样本。另外，为了测出被试者的能力高下，这些试题应有一个适当的难度（亦有人认为应称做“易度”）分布。从这里可以看出，编制试卷也是一项技术性很强的工作。

L. F. Bachman (胡壮麟, 1997) 提醒人们，不要误信会有一种“最好的测试”。这也就是说测试的质量是通过比较显示出来的。而测试的质量，主要反映为信度、效度以及试题的区分度和难度。测试的信度，反映该测试的可靠性，“表现为同一个测验在不同时间上所得结果的一致性”（《教育大辞典》7）。测试的效度，则反映测试的有效性，也就是说该测验对所需测量的内容测到了多少。试题的区分度，是指对于被试者反应的区分程度或鉴别能力。通俗地讲，也就是能力强的被试者与能力差的被试者，能不能在该试题的得分上显示出来。试题的难度，表现为被试群体能做出正确反应的人数。所有被试者都能正确反应的试题难度指数为1，即最容易；所有被试者都不能正确反应的试题难度指数为0，即最难。当然这两种难度的试题就测验来说几乎没有意义，比较合适的难度指数是0.5。在实际的测验中，试题应有不同难度的分布，而且这种分布要根据测试目的和测试类型加以调整。

二、语言测试

语言，是我们这个地球上只有人类才具有的一种特殊的能力。它是伴随着我们这个物种的形成而同时建立起来的。它与人类的本能或其他后天获得的能力，也都有所不同。正因为如此，人类在语言科学的建立和发展方面显得特别笨拙，至今我们对语言的认识还十分有限，十分浅薄。与此相应的是，我们对语言获得和语言教学也很不自觉。在教育学家所做出的学习类型中，语言学习无法对号入座。而语言学习作为一个特殊的学习类型，我们也无法确定它的性质和特点。我们还只能在心理语言学领域里探索（其中包括大量的假说）语言获得的规律，而在应用语言学领域里试图建立语言教学的原则、操作规范和教学法。

语言测试，既然是语言教学的一个重要环节，它的设计水平和操作水平必然跟语言教学相称。许多见识不凡的应用语言学家都表达了这样一种观点：“任何一种语言测试方案，都是一定的语言观、语言习得和语言教学原则的体现。”（刘珣等，1986）这种情况的存在是不难理解的，因为教育测量必须规定所要测量的事物的属性。人们只有感知、意识到事物存在着某种属性，才会自觉地设计、制造适当的工具去测量它。实际上，我们的测量工具总是存在着需要改进之处。因此，现行的语言测试在表面效度上或多或少都存在着一些问题，Bachman 认为不存在“最好的测试”也许是基于这方面的考虑。因此，语言测试的发展和完善，在很大程度上取决于语言科学研究的进步。

早期，在语文学研究的影响下外语教学主要采用语法翻译法，与此相应，当时的测试项目也集中在语言结构和翻译能力方面，试题类型主要有语法形式的填空、语法分析、改写句子、翻译和缩写等。20世纪40年代，在行为主义心理学和结构主义语

言学的指导下,听说法占据了语言教学的主导地位。跟听说法配套的测试可称做“分离式”测试。这种测试,一方面重视言语行为,因此对听、说、读、写四项言语技能分别进行测试;另一方面,分离式测试还对由结构主义语言学发现的各层面的语言要素进行分别测试。跟语法翻译法时期的测试相比,分离式测试显示出严谨细致的科学分析精神。自从 Hymes 提出要重视“言语交际能力”(communicative competence)以来,语言学研究越来越注意语言系统的运用特点和社会环境的影响。受到语言研究这种转变的影响,语言教学也推出了交际法,语言测试也在基本原则和基本框架方面发生了巨大的变革。应用语言学家认为语言及其运用中存在大量的可变规则(或称“软”规则),“因此,语言测试必须更具兼容性,内容更丰富”(阿兰·戴维斯,1997)。适用于交际法教学的语言测试是一种交际式的测试,它特别重视被试者的言语交际能力。因此,其试题跳出句子的范围,更多地进入话语、语篇的领域。试题设置交际目标、语境等条件,不仅测量被试者运用语言的正确性,还测量他们运用语言的得体性。李筱菊(1997)把语言测试的发展跟语言观的发展联系起来,划分语言测试的三个发展阶段。她把这三个阶段称为“三代”,也许是看到了其间的批判、继承、发展关系。第一代被称做“科学前语言测试”(prescientific testing),这类测试的主要测量目标是知识性的。第二代是“心理计量结构主义测试”(psychometric - structuralist testing),这类测试的主要测量目标是对于语言形式结构的操作技能(skill)。第三代则是“交际语言测试”(communicative language testing),也被称做“心理语言学 - 社会语言学测试”(psycholinguistic - sociolinguistic testing)。这类测试的主要测量目标是语言能力(competence),不仅包括前两代所测量的知识和技能,更重要的是还应测试言语交际能力。语言学研究还在飞速发展,语言教学(应用语言学)的探索也沿着科学性、实用性的方

向进行，我们相信语言测试也将进一步发展、完善。而且可以预见语言测试的发展，除了有教育测量理论和技术进展的影响，更多得益于语言学和应用语言学的进步。

另一方面，正如 Upshur (1971) 所指出的，语言获得研究和语言测试研究是互动的。单凭直觉我们就能发现，测试的结果对教师和学生都有重大的意义。

测试的结果，常常会迫使教师和学生改变下一阶段的行动方向或行动计划。事实上，语言测试对语言教学的影响是极其深刻的。S. 皮特·科德至少归纳出以下几点：(1) 语言测试能够对教学设计、教学过程进行有效的评价。实际上，在进行教学设计时总有很强的“预测”成分，对这些预测成分的第一个实证就是测试。(2) “语言测试是一种衡量工具，它们是用于学生的”，而“衡量学生的知识，不仅是评价学生本人，也是评价教师和教材的一种手段”。(3) 测试本身是一种实验手段，而“实验是对假设进行检验的过程”，因此，“设计这些测试正是应用语言学家的活动之一”。在我们看来，语言测试在实测过程中向我们提供了大量的行为样本，为语言获得理论的建设提供生动、鲜活的语料。其中，被试者的错误反应固然可以作为偏误分析(EA)的语料，其实每一位被试者在应试过程中所显示的所有反应均可作为构建中介语(interlanguage)的有效材料。此外，语言测试实际上也可被视为教学目标和教学任务的具体体现，对大部分语言教师来说，从标准化测试的试卷中能够接受比教学大纲文本更加明确的教学要求。

还有一点值得一提，现代语言教学理论从认知心理学的观点出发认为，语言获得的主体是学习者（一个著名的表述方式是：“语言是学会的，而不是教会的”），因此学习者的动机和态度对语言学习至关重要。其间，语言测试也会对学习者的学习态度产生深刻的影响，这种影响可以是积极的，也可以是消极的。仅从