

汉语测试、习得与认知探索 (续一)

——“第二届语言测试与习得研究生学术论坛”
论文选

主 编 张旺熹 邢红兵

世界图书出版公司

北京·广州·上海·西安

图书在版编目(CIP)数据

汉语测试、习得与认知探索(续一)/张旺熹,邢红兵主编. —北京:世界图书出版公司北京公司,2011.9

ISBN 978-7-5100-3804-4

I. ①汉… II. ①张… ②邢… III. ①汉语—语言学—文集
IV. ①H1-53

中国版本图书馆 CIP 数据核字 (2011) 第 144960 号

汉语测试、习得与认知探索(续一)

主 编: 张旺熹 邢红兵

责任编辑: 武传霞

装帧设计: 春天·书装工作室

出 版: 世界图书出版公司北京公司

出 版 人: 张跃明

发 行: 世界图书出版公司北京公司

(地址: 北京朝内大街 137 号 邮编: 100010 电话:
64077922)

销 售: 各地新华书店和外文书店

印 刷:

开 本: 787 mm × 1092 mm 1/16 印张: 20.75

字 数: 350 千

版 次: 2011 年 9 月第 1 版 2011 年 9 月第 1 次印刷

ISBN 978-7-5100-3804-4/H · 1232

定 价: 42.00 元

版权所有 翻印必究

编 委

(按姓氏音序排列)

柴省三 陈 宏 郭树军 黄理兵

李 慧 聂 丹 王佶旻 邢红兵

张宝林 张 凯 张旺熹

本书的出版得到北京语言大学“特色重点学科建设项目”经费支持。

出版前言

为加强研究生之间的学术交流，激发研究生们的科研热情，北京语言大学汉语水平考试中心于2008年11月30日举办了“首届语言测试与习得研究生学术论坛”，会后精选了18篇论文集结出版，论文集《汉语测试、习得与认知探索》已由北京语言大学出版社于2010年3月正式出版发行。2010年11月20日，在北京语言大学举办了“第二届语言测试与习得研究生学术论坛”。来自北京语言大学、北京大学和北京外国语大学的近百名中外研究生参加了论坛，共同就语言测试、语言习得、语言认知及语言教学等领域的相关问题进行交流 and 探讨。

本届论坛征集论文42篇，提交的论文和举办的论坛具有以下几个方面的特点：（1）涉及的研究领域广泛。本次论坛提交的论文涉及语言测试、语言习得、语言本体和语言认知研究、语言教学研究等领域，抓住了各个领域的一些热点问题；（2）研究方法丰富多样。本次论坛提交的论文采用多种研究手段，尤其是在利用语料库资源和采用实证研究方法等方面，体现出非常明显的特点；（3）论文的来源具有开放性。来自不同高校的研究生提交了论文并参加研讨，还吸引了来自日本、韩国等国和中国台湾地区的研究生参加论坛并发表自己的研究成果，体现出国际化的特点；（4）团队合作研究已成规模。本次论坛收到了多篇合作研究论文，同学们发挥各自的特长，体现了很好的团队合作精神，也大大提高了论文的质量。

论坛结束后，我们组织了论文评审委员会，对提交会议的论文

进行匿名评审，最后有 21 篇论文入选文集，其中包括语言测试研究论文 6 篇，语言习得研究论文 8 篇，语言认知研究论文 7 篇。为了保持论文集的延续性，本届论坛的论文集定名为《汉语测试、习得与认知探索（续一）》。

论文集编委会成员均为北京语言大学汉语水平考试中心的研究生导师，他们从论坛组织、论文指导、论文评审等各个环节都全力投入，对入选论文也提出了很多具体的修改意见。导师们的辛勤付出，对提高研究生的科研能力和论文写作水平都有着重要的指导意义，也是论文集能够出版的保证。

感谢世界图书出版公司郭力女士为本论文集的出版所提供的帮助，感谢本书责任编辑武传霞所付出的辛劳。北京语言大学汉语水平考试中心研究生管理办公室的李郁老师和金樱老师在论坛的组织 and 论文的收集整理过程中做了大量的工作。王正刚先生对本书进行了细致的编校。汉语水平考试中心硕士研究生王专同学为本论文集的编辑做了很多基础性工作。

谨对上述各方面人员为本论文集的出版所作的贡献表示感谢！

北京语言大学汉语水平考试中心

2011 年 3 月 8 日

目 录

初级阶段留学生语言能力的结构与

- 特点 丁慧明 周 莉 周 聪 (1)
- 任务类型对考生语言复杂性的影响 冯佼佼 (19)
- 欧洲语言能力共同参考框架 (CEFR) 对 HSK 写作测试分数
解释的启示 耿 岳 (35)
- 面试型口语考试中考官提问话语初探 马玉红 (48)
- 高级水平汉语学习者听说读写四项技能关系的研究 原 鑫 (62)
- HSK (改进版) 中级写作考试的题型研究 张 艳 梁 清 (82)
- 基于“HSK 动态作文语料库”的“得”字补语句
偏误分析 蔡晓燕 (93)
- 两种认知风格的留学生对于“控制信息输出型”课堂活动
有效性评价的实验研究 耿 岳 (117)
- 韩国学生“是……的”句式习得的实验
研究 郝瑜鑫 魏岩军 李 强 (131)
- 外国留学生汉语量词学习策略的个案研究 龙江燕 (142)
- 以泰语为母语的学生汉语趋向补语的习得
顺序研究 宋红红 辛 鑫 彭立卉 郭肖芳 黄贤淑 (159)
- 现代汉语语料库和汉语中介语语料库中动词用例的
对比分析 辛 鑫 曲学智 (178)
- 韩国留学生使用连词“还有”的偏误分析及
教学策略 徐 威 (193)
- 形声字熟悉度及形旁位置对非汉字圈
初级留学生形旁意识影响的实验研究 张力力 (212)
- 从认知的角度看成语的稳定与变化 曾小兵 邱丽娜 (233)
- “反正”类语气副词的语法化等级及其隐喻作用 高 元 (246)
- 动补结构带宾语构式考察 郭雅静 (261)



现代汉语单宾多系动词后接宾语语义类型的

配位原则	黄敬	(277)
语气副词“敢情”和“原来”的语义语用差异初探	张莹	(290)
连词“还是”和“或者”的语义差异及其认知基础	周也明	(304)
从隐性量范畴的角度看“就”的关联格式特征	佐佐木俊雄	(319)



初级阶段留学生语言能力的结构与特点^{*}

丁慧明 周 莉 周 聪

北京语言大学 汉语水平考试中心

摘 要 本文以 HSK [初级] (改进版) 考试的实考数据为材料, 运用结构方程模型的分析方法, 对考生听、说、读、写四项技能之间的关系进行考察。我们还探讨输出类技能(说、写)与输入类技能(听、读)之间的关系, 主要依赖于视觉通道的技能(读、写)与主要依赖于听觉通道的技能(听、说)之间的关系, 在此基础上来分析初级水平汉语学习者语言能力结构内部的关系。经过研究, 我们发现: (1) 听+说+读+写的四因子模型拟合程度最好, 最适宜用来描述初级水平留学生的语言能力结构; (2) 初级阶段留学生的听、说、读、写四项技能总体来看呈现分散的趋势; (3) 听觉渠道和视觉渠道对初级阶段留学生汉语技能发展的影响最为明显。

关键词 语言能力结构 结构方程模型 初级汉语

一 问题的提出

“语言能力”这一概念由乔姆斯基提出, 他认为语言能力是“说话人—听话人所具有的关于他的语言的知识”(乔姆斯基, 1965: 2)。此概念自提出以来引起了学术界的广泛讨论, 不同阶段人们对其内涵有不同的理解。研究语言能力结构不管是对语言测试还是对语言教学而言都有着深远的意义。Bachman (1990) 说“一个关于语言能力的明晰而精确的定义, 对语言测验的开发和应用而言, 是基本的要求。”语言测试作为应用语言学的一个分支, 把探究人的语言能力作为自己的终极目标。这个目标有双重意义, 一是研究语言能力, 弄清其本质及形成过

^{*} 此项研究得到教育部人文社会科学重点研究基地重大项目“汉语作为第二语言的能力标准(项目批准号: 06JJD740005)”和北京语言大学汉语水平考试中心研究生科研项目“初级阶段留学生语言能力的结构与特点”的资助。



程，进而对其进行理论上的解释；二是理论解释可直接用于语言教学，因为语言教学的现实目的就是使人获得一种语言（第二语言）的语言能力。因而，语言能力对语言测验来说，无论是在理论层面还是在实践层面都是一个核心问题（张凯，1998）。汉语作为第二语言时，学习者的汉语能力结构也会呈现出一定的特点。而目前对汉语（L2）能力的研究，尤其是国内在这方面的研究主要集中在理论层面，实证研究成果尚不多见。因而用实证的方法探讨汉语学习者的语言能力结构显得很有必要。

二 语言能力结构已有的相关研究

2.1 国外语言能力结构研究

根据 Spolsky (1995) 的观点，国外语言测验领域对语言能力的研究大致经历了三个阶段。（一）技能/成分说阶段，代表人物是 Lado 和 Carroll。Lado 的语言能力观是技能成分相结合，他认为语言测验涉及两种变量：成分和技能。成分是语言本身的各种成分，技能是听、说、读、写四种，各种成分虽可单独被研究、描写和测量，但在语言中从不单独出现，而是综合在听、说、读、写各项技能中。Carroll 提出了包括语言维度和技能维度的二维语言能力模型图，语言维度由音位/拼音、形态学、句法、词汇四元素构成，技能维度由听力、口语表达、阅读、写作四元素构成。技能/成分说对语言测试的开发和研究影响很大，分立式测验也由此盛行开来。HSK 的理论基础——它对于语言能力的基本假设，大体上也遵循了 Lado 对语言能力的定义（陈宏，1997）。（二）一元化阶段，代表人物是 John Oller。Oller (1974) 认为，听的能力要求与其他能力不一样，需要将语音、语法、词汇等方面相结合，是完成语言交际的根本前提，但是他仍然强调听与其他的能力是融合在一起的。后来，他提出单一能力假说，认为语言能力根本没有结构，而是一个单一的不可再分的能力，所有的测验测量的都是一个一般因素。他使用主成分分析法（principal component analyses）分析了 UCLA 英语分班考试（Oller, 1976; Oller & Hinofotis, 1980），得出的结果支持单一能力假说。（三）语言交际能力模型的建立，代表人物是 Bachman。20 世纪 80 年代，研究者们用探索性因素分析法（Bachman, Davidson, Ryan & Choi, 1995; Carroll, 1983）和验证性因素分析法（Bachman



& Palmer, 1981, 1982; Kunnan, 1995) 证明了 Oller 的“单一能力说”至少是过于绝对。Bachman 推出了由语言能力、策略能力、心理生理运动机制构成的语言交际能力模型。听说读写四项技能形态在心理生理运动机制这部分, 根据不同渠道、不同模式, 可以有不同的归类。详见下表:

表 1 Bachman 对四项技能形态的归类

四项技能形态	渠道	模式
听	听觉渠道	理解性模式
说	听觉渠道	表达性模式
读	视觉渠道	理解性模式
写	视觉渠道	表达性模式

一些研究表示还存在一个高阶能力 (Bachman & Palmer, 1981, 1982; Kunnan, 1995)。尽管如此, 目前语言测试学界在第二语言能力结构尤其是技能形态问题上已基本达成共识, 认为第二语言能力结构是由多个部分构成的, 每个部分又涉及不同的子能力。但是不同的研究所得出的结构模型又不尽相同, 不过这跟研究所用的方法、工具以及数据样本有很大的关系。Swinton 和 Powers (1980) 以及 Manning (1987) 采用探索性因子分析的方法对 TOEFL 纸笔考试进行了研究, 得出一个二因子结构, 独立的听力因子以及一个由结构、词汇、阅读理解混合成的因子。后来 Hale 等人 (1988) 用验证性因子分析的方法也得出了同样的结论。由于 TOEFL 纸笔考试没有口语和写作, 所以很容易得出上述结论, 也就不能完全描述二语者的语言能力。自 2005 年 ETS 推出 TOEFL-iBT 以来, 又有很多研究者作了大量研究。比较有影响的是 Sawaki 等人 (2008) 对试卷因子结构进行分析, 找到了四个独立的听、说、读、写一阶因子和一个高阶因子。这跟之前的 TOEFL 纸笔考试研究相比, 得到了更多的因子。Stricker 等人 (2005) 对和 TOEFL-iBT 试卷具有相似结构 (包括了听、说、读、写四个部分) 的 LanguEdge 测试进行分析, 得到了说、听读写两个因子。

2.2 国内语言能力结构研究

目前, 国内对语言能力结构尤其是汉语学习者能力结构的专门研究比较少, 大部分的相关文献主要集中在 HSK 考试的结构效度研究方



面。张凯（1992）使用探索性因子分析方法对 HSK 试卷的结构效度进行了探讨，试卷由听力、语法、阅读、综合四个部分组成。研究者最后找到了三个因子，分别为阅读理解、听力理解以及一个未知因子。郭树军（1995）用 Henning（1987）的方法对 HSK（试卷构成同上）的内部结构效度进行研究，得到了相似的结论。陈宏（2000）用探索性因子分析的方法也对相同结构的材料进行了分析，得到了四个因子。其中可以较为清晰辨认的包括阅读理解、听力理解以及方法因素。李慧、朱军梅（2004）采用验证性因子分析方法对 HSKJ234 进行构想效度的研究，也得到了四个因子，分别为听力理解能力、语法结构能力、阅读理解能力和综合填空能力。但在这些研究中，试卷都没有包括口语和写作部分，所以这些结果不能全面反映汉语学习者的能力结构。王倩旻（2008）立足实证，拟构了初学者口语能力模型。首先提出口语表达分为听—说模式和独白模式的理论构想，采用半直接式口语测验。施测完毕得到分数后，作者先对拟建的模型进行探索性因素分析，用主成分分析法提取到两个因素，此结果支持了理论构想；接着作者对模型进行验证性因素分析，各项拟合指标均显示模型有较好的拟合度。这两项分析结果说明，对口语能力结构的理论构想是基本正确的。马新芳（2005）以 HSK（高等）、MHK（三级）为例，对第二语言阅读能力和写作能力的关系作了实证分析。她具体研究了这两种能力水平不平衡的现象，发现“阅读水平低的人相应的写作水平一般也比较低，阅读水平高的人写作水平要滞后于阅读水平，达不到相应的阅读水平的高度；写作水平低的人相应的阅读水平一般也低。”作者通过考察口语和写作、口语和阅读的相关，探索组织能力在阅读、写作中的不同作用。进而考察证实写作能力不同于阅读能力这一假设，检验结果也支持了作者的假设。

2.3 已有研究的模型类别

上述理论与实证研究提出了众多语言能力结构模型，我们选取有代表性的如下 8 种。孰优孰劣，哪种与真实的汉语（L2）能力最为接近，我们暂无从得知，接下来本研究将对这 8 个模型进行逐一检验。



表 2 8 个语言能力结构模型

模型因子数	因子构成	理论依据	实证依据
四因子	听+说+读+写	Lado 和 Carroll 的技能成分说	Sawaki 等 (2008)
三因子	听+读写+说	——	Carroll (1983) Bachman (1995) Kunnan (1995)
	听说+说+写	Carroll, North 等	——
二因子	听说+读写	Bachman 的心理生理运动机制: 按渠道划分	——
	听说+说写	Bachman 的心理生理运动机制: 按模式划分	——
	听读写+说	——	Stricker 等 (2005)
	听+说读写	——	Oller (1974)
单因子	听说读写	Oller 的单一能力假说	Oller (1976) Oller 和 Hinofotis (1980)

三 研究方法与被试

3.1 样本

本研究的样本是 2009 年参加了 HSK [改进版] (初级) 考试所有分测试的考生, 共 49 人。他们来自意大利、法国、美国、德国、英国、保加利亚、瑞典、西班牙、蒙古国、韩国、日本、泰国、马来西亚等 13 个国家。

3.2 测量工具

本研究采用的测量工具是 2009 年 HSK [改进版] (初级) 考试试卷, 包括主试卷、口语试卷和写作试卷三个部分, 全面测量听、说、读、写四个方面。主试卷有听后看图选择 15 题, 听问题并选择应答 15 题, 听对话和问题并选择回答 20 题, 选择正确答案完成句子 40 题以及短文阅读 30 题。主试卷题目都是采用二级计分。口语试卷包括重复句子 20 题、简短回答 20 题和口头陈述 2 题。写作考试分为完成句子、看图并用指定词语写句子两部分, 各 10 题。口语考试和写作考试采用的都是 1~3 级的评分量表。





表 3 HSK [改进版] (初级) 试卷构成

试卷结构	题 型	题 数	计分方式	原始分的满分
听力理解	听后看图选择	15	二级计分	15
	听问题并选择应答	15	二级计分	15
	听对话和问题并选择回答	20	二级计分	20
综合阅读	选择正确答案完成句子	40	二级计分	40
	短文阅读	30	二级计分	30
口语考试	重复句子	20	0~3级	60
	简短回答	20	0~3级	60
	口头陈述	2	0~3级	6
写作考试	完成句子	10	0~3级	60 (两位评分员所给分数之和)
	看图并用指定词语写句子	10	0~3级	60 (两位评分员所给分数之和)

3.3 研究方法

我们采用结构方程模型 (Structural Equation Modeling, 以下简称 SEM) 来分析数据, 分别估计 8 个模型的拟合指数, 比较这 8 个模型的优劣。结构方程模型作为一种多元数据分析工具自 80 年代以来被广泛应用于生物、教育、经济、医学、心理学和社会学等领域。国外语言测试领域运用结构方程模型进行研究的成果也已有不少, 最早使用它来研究语言能力结构的是 Swinton 和 Powers (1980)。Bachman 和 Palmer 用它来研究 FSI 口试的构念效度 (1981)、交际语言能力的构成 (1982) 及交际语言能力的自我评估 (1989)。Ginther 和 Stevens (1998) 用它来调查一项西班牙语测试的因素结构。Song (2008) 也运用此方法研究第二语言听读测验理解分技能的可分性。国内语言测试领域运用此项技术进行研究的成果还不多, 吴红云、刘润清 (2004) 运用它来研究外语写作元认知能力与写作成绩之间的关系。王佶旻 (2008) 运用它对初级水平汉语学习者的口语能力结构进行了探索。我们的研究是要比较 8 个模型的拟合程度, 观察因子间的相关关系。而结构方程模型可以同时处理潜变量及其指标, 能够分析涉及潜变量的复杂关系, 能够用来比较不同的模型 (侯杰泰、温忠麟、成子娟, 2004), 很符合我们研究的需要。



上文我们已经讲过 HSK [改进版] (初级) 的试卷结构及题目类型, 整个测试是由 4 个部分组成, 共 10 种题型, 182 题。由于结构方程模型分析数据时要求研究中使用的变量个数同样本量的比值不能太大, 如果在样本小的情况下变量个数多, 那极可能得到不稳定的解。所以我们不能直接将每个题目视为一个指标, 而应将题目合并为题目组以减少指标量, 一个题目组视作一个指标, 用题目组还能使数据更接近正态分布。具体做法是: 听力理解部分的三个题型以每个题型作为一个题目组, 分别命名为 L1, L2, L3; 综合阅读部分的两个题型以每个题型作为一个题目组, 分别命名为 R1, R2; 口语考试部分的“重复句子”“简短回答”“口头陈述”分别有 20 题、20 题、2 题, “口头陈述”的题目数量过少, 与其他题型的题目数相差过大, 不宜单独作为一个指标, 考虑到“口头陈述”与“简短回答”在考查目的、答题模式等方面更为接近, 我们将二者合并, 命名为 S2, “重复句子”题目组命名为 S1; 写作考试里的“完成句子”共 10 题, 作为一个题目组 W1, “看图并用指定词语写句子”共 10 题, 作为另一个题目组 W2。

表 4 题目组构成

	指 标	题 数	计分方式	原始分的满分
L (听力理解)	L1 听后看图选择	15	二级计分	15
	L2 听问题并选择应答	15	二级计分	15
	L3 听对话和问题并选择回答	20	二级计分	20
R (综合阅读)	R1 选择正确答案完成句子	40	二级计分	40
	R2 短文阅读	30	二级计分	30
S (口语考试)	S1 重复句子	20	0~3 级	60
	S2 简短回答、口头陈述	22	0~3 级	66
W (写作考试)	W1 完成句子	10	0~3 级	60 (两位评分员所给分数之和)
	W2 看图并用指定词语写句子	10	0~3 级	60 (两位评分员所给分数之和)

3.4 数据分析

我们先统计每个被试在 10 个观测变量上的原始分得分, 将统计得到的数据输入 LISREL 8.80, 检验是否存在缺失值。发现有一个缺失值, 我们用均值代替。由于数据中是否存在相关关系尚不确定, 根据



(Jöreskog & Sörbom, 1988, 1993) 的观点, 我们选择在协方差矩阵基础上分析数据, 而非选择相关矩阵, 用 LISREL 软件计算出数据的协方差矩阵。如下表所示:

表 5 数据的协方差矩阵

	L1	L2	L3	R1	R2	S1	S2	W1	W2
L1	8.833								
L2	5.071	7.889							
L3	9.131	8.207	19.612						
R1	9.821	10.178	16.656	36.884					
R2	8.558	8.795	16.896	26.828	33.535				
S1	18.238	14.897	27.646	32.279	22.515	80.311			
S2	29.457	25.305	43.973	55.739	47.509	94.526	189.001		
W1	14.182	19.019	27.566	54.466	49.824	55.934	103.117	148.755	
W2	10.351	10.288	23.358	40.115	43.752	36.646	73.868	110.866	176.510

我们运用 LISREL 8.80 软件对上文提到的 8 个理论模型逐一进行验证, 使用极大似然估计 (ML) 的方法进行参数估计。由于复杂模型需要迭代多次才能收敛, SEM 的默认迭代次数是 250 次, 我们在输出指令中追加为迭代 500 次。

结构方程模型软件在检验模型时, 会提供众多拟合指数供我们参考。这些指数主要分为两大类: 绝对拟合指数和相对拟合指数。依据 Kelloway (1998) 的观点, 我们在检验模型、考察模型整体拟合情况时选择了 χ^2 , df, p, χ^2/df , RMSEA, GFI, AGFI, CFI, NNFI, ECVI 这 10 个指数。 χ^2 能够恰当地用来反映模型的拟合优度, 但 χ^2 的值极易受样本大小的影响。考虑到本研究中 $N=49$, χ^2 的显著性阈值 α 应取 0.01, 当 χ^2 检验显示 $p>0.01$, 即 χ^2 不显著时, 表明模型拟合数据 (Kelloway, 1998)。 χ^2/df 是卡方与自由度的比值, 通常报告 χ^2/df 比报告 χ^2 更受欢迎, 特别是在模型比较时, χ^2/df 比 χ^2 更有参考价值 (侯杰泰、温忠麟、成子娟, 2004)。Kline (1998) 指出, $\chi^2/\text{df}\leq 3$ 表示拟合较好。RMSEA 是近似误差均方根, 侯杰泰 (2002) 指出该指数受样本量影响较小, 是较好的绝对拟合指数。Steiger (1990) 认为, RMSEA 低于 0.1 表示好的拟合, 低于 0.05 表示非常好的拟合。GFI, AGFI 都是拟合优度指数, AGIF 是在模型一定的自由度条件下调整 GFI, 二者取值都在 0~1, 大于 0.9 表示拟合较好, 如果二者差异很

