

# 汉语作为第二语言的初学者口语能力结构初探<sup>\*</sup>

王 佶

(北京语言大学 汉语水平考试中心 北京 100083)

**摘 要** 根据口语表达及汉语初学者口语能力的发展特点,拟构了初学者口语能力模型。该模型从语言的成分、情景特征以及它们相互作用的角度来定义构想,把口语表达分为听-说模式和独白模式两种表达模式,并分别从短语、句子和语段三个层面来考查口语能力。以来华学习汉语半年至一年的留学生共 275 人为样本,使用探索性因素分析和验证性因素分析研究了初学者口语能力模型与实测数据的拟合程度,结果表明拟合程度良好,模型得到了验证。

**关键词** :口语能力 ;口语表达模式 ;验证性因素分析

**中图分类号** :B842.5

**文献标识码** :A

**文章编号** :1003-5184(2008)01-0033-05

## 1 问题的提出

在第二语言口语测验的研究中,对口语能力结构的研究十分欠缺,对于汉语作为第二语言的口语能力结构问题更是鲜有人问津。多数口语测验是建立在某个语言能力等级量表之上的,也就是说,测验的编制依据的是量表中对于各个等级语言水平的描述。这样的方法虽然可以使我们对不同语言水平的考生的语言表现作出区分,但缺乏对口语能力结构的理论分析。第二语言学习者的人数分布是呈金字塔形的,初学者位于塔的底部,是人数最多的一部分。在汉语逐步向国际推广的今天,初学者的人数会以更加迅猛的势态增长。由于汉语文字系统的特殊性,初学者的书面表达能力十分有限,而口语能力却能够有长足的发展,因此探讨初学者的口语能力的结构问题就显得更加有意义。

## 2 初学者第二语言口语能力测验的理论构想

口语属于口头渠道的表达模式。这种产出模式在实际运用中有两种子模式,其一是听-说的模式。即说话人依靠听的能力来接受信息,依靠说的能力来产出信息。其二是单模式的口语产出过程,即口语独白的模式,这也是我们经常遇到的口语表达形式。欧洲语言委员会语言能力共同标准(Common European Framework, CEF)关于口语能力的描述中就明确区分了这两种模式,并且指出在评价口语时要从这两个不同的方面来进行。

另一方面,二语学习者口语水平的发展过程经历了从词语或短语表达达到句子表达再到语段表达的

三个阶段。对于多数初学者而言,口语表达的形式以短语和句子为主,部分优秀学习者会在较短的时期具备一定的语段表达能力。因而可以认为初学者的口语特点是以短语和句子表达为主,并逐渐开始具备一定的语段表达能力。因此对初学者口语能力的考查也应该涉及这三种形式。

根据口语表达及初学者口语能力的发展特点,可以把口语能力定义为学习者在口头渠道运用语言的能力,同时操作性地定义为学习者在听-说模式下的口语表达能力和在独白模式下的口语表达能力,并在此基础上提出初学者二语口语能力测验的理论模型,具体如图 1。

该模型中,听-说模式和独白模式是口语运用发生的两种具体环境,而短语、句子和语段三种表达形式可以理解为三种成分。初学者口语测验模型体现了两种口语表达环境与三种口语表达形式的相互作用。

## 3 研究方法 with 过程

### 3.1 被试

依据《汉语水平等级标准与语法等级大纲》中的有关描述,确定目标团体为接受正规现代汉语教育半年至一年的留学生,并把它作为研究对象。被试来自北京语言大学汉语进修学院一年级上和一年级下的留学生,共 275 人。其中预测样本 52 人,来自 13 个国家或地区,正式测试样本 223 人,来自 53 个国家或地区。样本的具体情况如表 1。

<sup>\*</sup> 基金项目 北京语言大学青年基金项目(06QN03)。

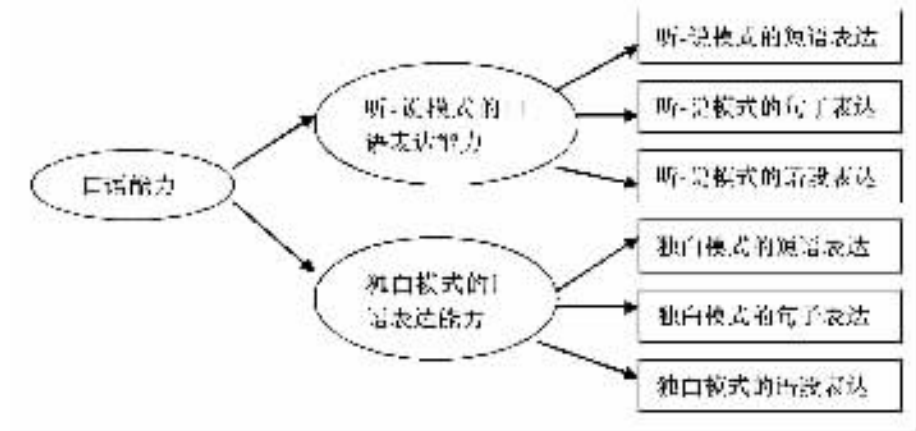


图1 口语能力模型图

表1 预测样本情况

	人数	男	女	平均年龄
一年级上	25	15	10	23.57
一年级下	27	16	11	25.34
总计	52	31	21	24.46

表2 正式测试样本情况

	人数	男	女	平均年龄
一年级上	117	65	52	21.29
一年级下	106	59	47	21.59
总计	223	124	99	21.29

3.2 研究工具

3.2.1 汉语初学者口语测验的编制

汉语初学者口语测验为半直接式口语测验,学

生持有试卷并且从耳机中听到提问,全部考试时间约20分钟,测验的结构框架如下表。

表3 测验结构框架

	题型	部分	准备时间	答题时间	题目数	预测题目数	刺激方式	反应方式
汉语初 学者口 语能力 测验	快速 回答	第一部分	2 秒	5-8 秒	10	20	听问句	口问回答 问题
		第二部分		10-15 秒	10	20		
	图片 比较	第一部分	8 秒	10 秒	10	20	看图	说出两幅图 的不同点
		第二部分		15 秒	10	20		
	听后 重复	第一题	30 秒	90 秒	1	1	听一段 话语	把听到的内 容重复说出 来
		第二题	30 秒	90 秒	1	1		
	看图 说话	第一题	2 分钟	90 秒	1	1	看一组 图片	根据一组图 片说一段说
		第二题	1 分钟	60 秒	1	1		

其中,快速问答和听后复述属于听说渠道的口语测验题型,图片比较和看图说话属于独白渠道的口语测验题型。另外,快速问答的前半部分主要以回答词或短语为主,后半部分以回答比较完整的句子为主,图片比较也是如此,听后复述和看图说话则是语段表达。

3.2.2 确定评分标准

Madser(1983)<sup>[1]</sup>针对初学者口语测验的评分作了客观化探讨,认为可以采用0/1的形式对包括问答、看图说话和复述等题型进行评分。Madsen认为初学者的发音、语法等能力比较弱,所谓的中间状态

不好评判,不适合采用分部评分(partial credit scoring),即指评分时包括对中间状态的评分。比如回答完全正确得2分,发音有问题但回答可以理解得1分,没有回答或回答不可辨认的得0分。研究采用Madsen所推荐的0/1的评分形式。对于问答题,用回答符合题意并且可以理解作为评价标准,是得1分,否得0分。对于看图说话,事先将图画切分成若干片断,对每一片段实施0/1评分,然后相加得到总分。对于复述题型,将所要复述的段落事先切分成若干要点,然后分别加以0/1评分,最后汇总得到总分。

3.3 测验构想的分析与检验

3.3.1 探索性因素分析

在理论建立与发展过程中,通过探索性分析建立模型,再用验证性因素分析来检验模型,这种程序称为交叉证实( Cross – validation ),这样可以保证量表所测特质的确定性、稳定性和可靠性<sup>[2]</sup>。另一方面,根据张建平、侯杰泰<sup>[3,4]</sup>等人的观点,不能用同一组数据既做探索性因素分析又做验证性因素分析。因此先使用探索性因素分析对预测卷的构想效度进行检验,以初步分析研究对口语能力的操作性定义,对模型进行检验和修正。

进入因素分析的变量为 6 个分测验,即快速问答第一部分、快速问答第二部分、图片比较第一部分、图片比较第二部分、听后复述、看图说话,分别记为 A1、A2、B1、B2、C、D。计算时均使用各分测验的总分。在这之前,先进行 KMO 和 Bartlett 球形检验来检验采样充足性以及变量间相关系数矩阵是否适合进行因素分析,检验结果如表 4。

表 4 KMO 和 Bartlett 球形检验

Kaiser – Meyer – Olkin 采样充足度检验		0.821
	卡方	434.761
Bartlett 球形检验		自由度 28
	显著性	0.000

一般而言,KMO 值至少应该大于 0.5 采样才算充足,0.821>0.5 检验合格。Bartlett 球形检验结果显著,说明相关系数可以用于因素分析提取因素。采用主成分分析法提取因素,具体情况如下:

表 5 探索性因素分析因素提取

因素	特征值	方差贡献(%)	累计方差贡献(%)
1	3.741	62.356	62.356
2	1.134	18.897	81.254
3	0.413	6.891	88.145
4	0.307	5.124	93.268
5	0.233	3.891	97.159
6	0.170	2.841	100.000

可以看出,只有两个因素的特征值大于 1,并且这两个因素的累计方差贡献率已经达到了81.254%,也就是说用这两个因素代替 6 个原始变量,可以概括总信息的 81% 以上,把这两个因素分别记做因素 1 和 2。两因素未经旋转的负荷矩阵如下。

表 6 两因素未经旋转的负荷矩阵

分测验	因素 1	因素 2
A1	0.757	0.487
A2	0.708	0.561
B1	0.846	-0.404
B2	0.875	-0.268
C	0.785	0.265
D	0.756	-0.525

由于各变量在因素上的负荷没有明显区别,在给因素进行命名时比较困难,这时就需要进行旋转使负荷系数向 0 和 1 两级分化。采用极大方差旋转,旋转后的因素矩阵模式如下。

表 7 旋转后的因素矩阵

分测验	因素 1	因素 2
A1	0.228	0.871
A2	0.142	0.892
B1	0.897	0.274
B2	0.826	0.394
C	0.399	0.725
D	0.912	0.124

因素 1 在 B1、B2 和 D 上有较高的负荷,而在 A1、A2 和 C 上的负荷比较低(均小于 0.4)。因素 2 在 A1、A2 和 C 上有较高负荷,而在 B1、B2 和 D 上的负荷比较低(均小于 0.4)。B1、B2 和 D 分测验考查的是独白模式的口语表达,而 A1、A2 和 C 分测验考查的是听 – 说模式的口语表达,因此可以把因素 1 叫做独白模式的口语表达能力,把因素 2 叫做听 – 说模式的口语表达能力。这说明探索性因素分析的结果同理论构想颇为一致。

3.3.2 验证性因素分析

用验证性因素分析来进一步检验实测数据是否与理论模型有较好的拟合度。研究的理论假设是汉语初学者口语能力测验包括两因素:听 – 说模式的口语能力和独白模式的口语能力,分测验一、分测验二和分测验五属于因素一,分测验三、分测验四和分测验六属于因素二。使用 Amos4.0 来检验该理论模型与实测数据的拟合程度。

要检验和评价模型是否与数据拟合,需要借助各种拟合指数。按侯杰泰、温忠麟、成子娟推荐的分类,拟合指数可以分为三大类,即绝对指数、相对指

数(或称增值指数)和简约指数,其中常用的是绝对指数和相对指数<sup>[5]</sup>。绝对指数中常用的有 $\chi^2$ 、 $\chi^2/df$ 、近似误差均方根(RMSEA)、拟合优度指数GFI、AGFI等。相对拟合指数典型的有NFI、NNFI(TLI)和CFI。这些指数在检验模型拟合度时的评价标准是: $\chi^2$ 不显著表示模型拟合得好, $\chi^2/df$ 值越接近1越好,小于5表示模型可以接受,RMSEA值小于0.1表示好的拟合,GFI、AGFI、NFI、NNFI(TLI)和CFI指数的值大于0.9表示模型可以接受。首先按照这样的规则来判断研究模型的拟合度,检验结果如表8所示。

表8 验证性因素分析拟合指数评价

拟合指数	评价标准	研究的结果( $n = 223$ )
$\chi^2$	不显著	23.771( $df = 8, p = 0.003$ )
$\chi^2/df$	< 5	2.97
RMSEA	< 0.1	0.094
GFI	> 0.9	0.984
AGFI	> 0.9	0.910
NFI	> 0.9	0.976
NNFI(TLI)	> 0.9	0.969
CFI	> 0.9	0.984

从上面的结果可以看出, $\chi^2$ 值达到了显著性水平,说明模型拟合得不好。然而 $\chi^2$ 分布作为一个渐进分布,样本容量越大分布近似得越好。所以对

于小的样本容量( $n < 50$ ),估计的 $\chi^2$ 值误差往往很大,而对于较大的样本, $\chi^2$ 都较大,导致几乎所有模型都被拒绝<sup>[5]</sup>。根据温忠麟、侯杰泰、马什赫伯特<sup>[6]</sup>提出的卡方准则,样本容量与 $\chi^2$ 的显著性阈值有关。当样本量较大时, $\chi^2$ 的显著性阈值应当从通用的0.05下调,以减少错误拒绝零假设的风险。在样本容量 $n = 200$ 时,显著性阈值应当选择0.001,研究的显著性阈值为 $0.003 > 0.001$ ,因而可以认为 $\chi^2$ 值不显著。正如前面所说的, $\chi^2$ 值容易受样本容量影响,因而并不是个十分理想的指标,在评价模型时还要参照其他指标,特别是不容易受到样本容量影响的GFI、AGFI、NFI、NNFI(TLI)和CFI等指标。研究结果显示, $\chi^2/df = 2.97 < 5$ ,表明模型拟合得比较好。RMSEA值为 $0.094 < 0.1$ ,可以认为基本合格。GFI、AGFI、NFI、NNFI(TLI)和CFI值都大于0.9表示模型拟合得比较好。因此,综合各项拟合指数,可以认为模型的拟合度比较好。

接着分析模型的参数估计值,因为即使拟合指数显示模型拟合得很好,也不排除可能有些参数的估计值根本就没有意义,因而对参数逐一检视也是评价模型的重要步骤。每一个参数估计值都有相应的标准误,这样就可以对参数进行显著性检验。当检验的结果显示参数显著不等于零,则认为假设模型让该参数自由估计是合理的。研究模型参数估计值如下。

表9 模型参数估计值

	未标准化		$p$	标准化	$R^2$	残差
	$\lambda$ 估计值	C. R.		回归估计值		
因素一→分测验一	1.000			0.906	0.820	0.180
因素一→分测验二	0.903	17.347	0.000	0.863	0.744	0.256
因素一→分测验五	1.414	13.709	0.000	0.750	0.562	0.438
因素二→分测验三	1.000			0.903	0.816	0.184
因素二→分测验四	0.950	17.995	0.000	0.874	0.763	0.237
因素二→分测验六	1.053	11.987	0.000	0.793	0.629	0.371
因素一↔因素二	4.996	8.614	0.000	0.851		

上表中,C. R.(critical ratio)是参数估计值与其标准误的比值,用以对参数进行显著性检验,其值大于2即认为有显著性。可以看出,参数检验的结果均有显著性。标准化回归估计值也都比较理想(一

般认为最好介于0.5与0.95之间)。R<sup>2</sup>介于0.5至0.9之间,表明模型的拟合度不错。如此,可以把汉语初学者口语测验模型的因素结构图表示如下。

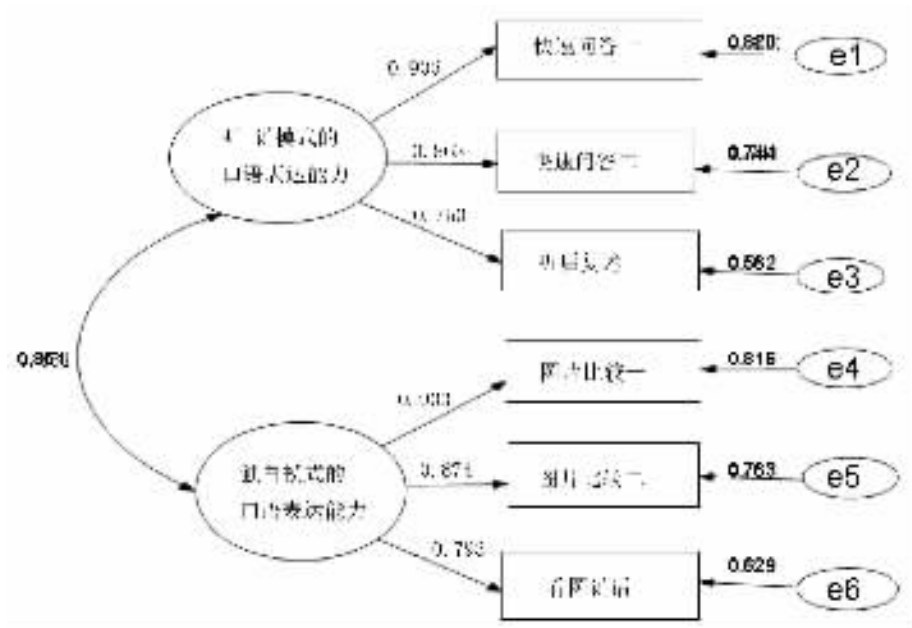


图 2 测验因素结构图

使用正式测试样本进行验证性因素分析的结果表明,不易受样本容量影响的指标 NFI、CFI、NNFI、GFI 和 AGFI 都大于 0.9,说明实测数据与理论模型的拟合程度良好。从探索性因素分析和验证性因素分析的结果来看,测验的构想效度是比较好的。这一方面说明测量工具,即汉语初学者口语能力测验有较好的效度,另一方面也说明对口语能力结构的理论构想是基本正确的。

4 讨论

4.1 关于初学者口语能力结构

口语能力作为一种相对独立的能力在第二语言的教学和学习中充当着重要的角色,口语活动的认知过程与表达方式同听力、阅读和写作是有区别的。以往的研究者从口语习得、口语训练和口语测查等方面出发构建了若干二语口语模型。Rivers 和 Temperley<sup>[7]</sup>于 1978 年提出了二语口语技能模型,该模型是以口语学习过程为基础而构建的。模型区分了技能获得( skill - getting )和技能运用( skill - using )两个方面,技能获得包括认知( cognition )和产出( production )两个因素,技能运用包括互动( interaction )因素。互动因素由接收和表达两个部分组成,在互动过程中口语表达是双向的,在产出过程中口语表达是单向的,这是两种口语产出的主要区别。上述观点可以用图 3 表示。



图 3 Rivers & Temperley 的口语能力模型图

Littlewood 从二语口语练习的角度提出了一个两维模型,包括预备交际活动( pre - communication activity )和交际活动( communication activity )两个因素<sup>[8]</sup>。模型的具体形式如下:

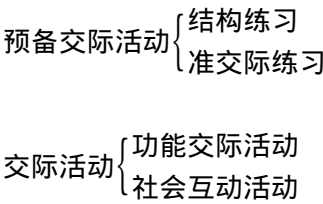


图 4 Littlewood 的口语能力模型图

Martin Bygate 则从口语测查的角度区分了两种口语技能,第一种称为机械感知技能( motor - perceptive skill ),第二种称为互动技能( interaction skill )<sup>[9]</sup>, Bygate 的观点可以用图 5 表示。

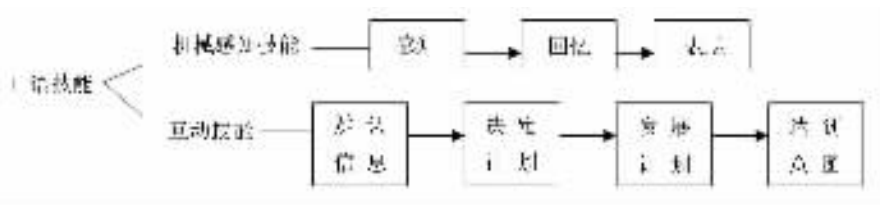


图5 Bygate 的口语能力模型图

上述的三种模型各自从不同角度分析了口语能力,虽然表述各不相同,但不难看出其中的共同点,即模型都区分了听和说互动形式的口语表达以及单向的口语表达,认为这两种口语表达能力是既有联系又有区别的能力。可惜的是,研究者并没有对以上三种模型作过实证性的检验,因而它们与实测数据的拟合程度如何人们不得而知。

根据汉语初学者的口语能力特点所拟构的初学者口语能力理论模型得到了实证研究的支持,这是在口语能力结构研究领域的一次有益的尝试。研究的结果说明口语能力并不是简单的成分组合,而是语言表达模式与表达内容相互作用的结果。

#### 4.2 关于口语能力的培养

研究的结果对第二语言口语能力的培养也有所启示。研究结果表明口语能力可以分为听—说模式的口语能力与独白模式的口语能力两个因素,在测验中前者是以听的形式输入的,而后者是以看的形式输入的。在口语训练中,教师也可以通过不同的输入途径来进行口语操练以培养学生不同方面的口语表达能力。周为京就提出,口语教学应当超越传统的听、说捆绑的中一范式,合理介入视觉输入,使听音输入和文字阅读输入并重或交替进行,充分利用不同输入模式的互补效应,综合提高学生口语的流利度、准确度和复杂度<sup>[10]</sup>。这一观点是正确的,

但周文是针对英语教学的,而在对外汉语教学中使用图片作为视觉输入要好于使用文字。这是因为汉语的文字系统不同于英语,初学者的汉字认读能力比较弱,不宜采用汉字作为输入方式,使用图片作为视觉输入方式更为合适。

#### 参考文献

- 1 Madsen H S. Techniques in testing. Oxford University Press, 1983.
- 2 姜勇.验证性因素分析及其在心理与教育研究中的应用.外国教育研究,1999(3):88-91.
- 3 张建平.一种新的统计方法和研究思路—结构方程模型述评.心理学报,1993(1):93-101.
- 4 侯杰泰.为何需要建立结构方程模型及如何尽力潜伏变量.教育研究学报(香港),1994,9:87-92.
- 5 侯杰泰,温忠麟,成子娟.结构方程模型及其应用.教育科学出版社,2004.
- 6 温忠麟,侯杰泰,马什赫伯特.结构方程模型检验:拟合指数与卡方准则.心理学报,2004,36(2):186-194.
- 7 Rivers W, Temperley R S. A Practical guide to the teaching of English. Oxford University Press, 1978.
- 8 Littlewood W. Communicative language teaching. Cambridge University Press, 1981.
- 9 Bygate M. Speaking. Oxford University Press, 1987.
- 10 周为京.语言输入模式对口语产出的影响.解放军外国语学院学报,2005,6.

## The Construct Validation of Chinese as a Second Language Oral Proficiency Test for Beginners

Wang Jimin

(Center of the Chinese Proficiency Test, Beijing Language and Culture University, Beijing 100083)

**Abstract** According to the characteristics and development of oral proficiency, we put forward a construct model of oral proficiency and design the test based on the model. This model concerns about language component, language situation and the interaction of them. The 'situation' comprises of listening-speaking mode and monologue mode, and the 'component' comprises of phrases, sentences and discourse. 275 overseas students who have received only 0.5~1 year of formal Chinese lessons serve as the subjects in the test. And the result of confirmative factor analysis supports the model of the construct of oral proficiency we hold.

**Key words** oral proficiency; mode of oral expression; confirmative factor analysis