



外国语言文学前沿研究丛书

# 语言测试的心理测量与统计模型

——  
朱正才 编著  
——



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

外国语言文学前沿研究丛书

# 语言测试的心理测量与统计模型

朱正才 编著



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

## 内容提要

本书主要是在作者已发表的 23 篇论文基础上完成的,专注于语言测试的心理测量学和统计学模型,分为四章:语言测试的心理测量模型及相关问题、分数等值模型、信度效度模型以及语言能力量表模型及相关研究。每章下的小节标题都是论文题目,论文发表的相关信息都标注在文章的末尾处,方便读者查找。考虑到每篇文章的参考文献对文章理解的重要性,也都一一予以保留。本书中的这些论文都密切联系中国实际,研究的都是中国大学生英语教学和测试方面的问题,以实证研究、定量研究为主,其中大量使用了心理测量和统计学的方法及模型。希望这些研究范例对中国年轻的语言测试工作者有一些参考价值。

## 图书在版编目(CIP)数据

语言测试的心理测量与统计模型/朱正才编著. —  
上海:上海交通大学出版社, 2022.8  
ISBN 978-7-313-22460-6

I. ①语… II. ①朱… III. ①语言-测试-心理测验  
-研究②语言-测试-统计模型-研究 IV. ①H09

中国版本图书馆 CIP 数据核字(2019)第 263887 号

## 语言测试的心理测量与统计模型

YUYAN CESHIDE XINLI CELIANG YU TONGJI MOXING

编 著: 朱正才

出版发行: 上海交通大学出版社

邮政编码: 200030

印 制: 上海万卷印刷股份有限公司

开 本: 710 mm×1000 mm 1/16

字 数: 296 千字

版 次: 2022 年 8 月第 1 版

书 号: ISBN 978-7-313-22460-6

定 价: 98.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 19.5

印 次: 2022 年 8 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系

联系电话: 021-56928178

在本书即将付梓之际,我想对自己的求学之路做一个简单介绍,这不但有利于读者了解我的研究特长,还能更好地理解我的这些小作品。

我大学时读的是教育管理学,之前则一直在中小学当数学老师,硕士读的是教育、心理测量与统计学,博士又转向了语言学——主要是语言测试方向。虽然我的求学经历表面上看有点令人眼花缭乱,但实际上一脉相承。教育学、心理学和语言学为我从事语言测试工作打下了理论基础,长期的教学经历能让我对考试与教学的关系有一个深刻的理解,而数学——尤其是统计学功底和编程计算能力,则让我具有较强的模型处理能力。语言测试是一个典型的交叉学科,主要涉及语言学、心理测量学和统计学,其中语言学关于语言能力的观点(即构念)是它的核心理论部分,而心理测量学和统计学解决的主要是构念的测量问题。这就像人的两条腿,走路时缺一不可。但遗憾的是,我国大部分语言测试工作者都是学语言出身,心理测量和统计学的基础很薄弱,这就给他们的研究工作带来了很大的不便。我经常能听到这样的问题和抱怨:“如何控制作文评分信度?”“如何保证分数的‘含金量’不变?”“是信度重要还是效度重要?”“如何建立分数常模?”“IRT 理论怎么这么难!”如此等等。

其实,这些问题也是我一直在苦苦思索的问题。我认为有的问题我完全有能力解决,有的问题只能暂时解决,有的问题则至今无解。在求索的过程中,我不揣浅陋,坚持把自己的一些研究成果写成论文发表出来,以便求教于同行。2019年初,上海交通大学外国语学院和上海交通大学出版社刚好有一个资助学术论文集出版的项目,部分同事和学生就提醒了我这件事,

问我能否把自己语言测试的相关论文结集出版出来。我想了想,就欣然同意了。

根据出版项目的要求,第一,入选论文必须围绕论文集的主题展开,不能包含无关论文;第二,所有论文要分出章节,自成体系;第三,论文需要适当导读,彼此之间要有衔接。

鉴于我的研究主要集中在语言能力测量问题上,因此把书名定为《语言测试的心理测量与统计模型》。本书基于 23 篇论文,分为四个专题:语言测试的心理测量模型及相关问题、分数等值模型、信度效度模型和语言能力量表模型及相关研究。

由于我一直在全国大学英语四、六级考试(CET)委员会负责心理测量和统计方面的工作——屈指一算快 30 年了,因此其中很多研究都是基于我在 CET 委员会的工作实践。我非常感激 CET 委员会的大力支持和同事们的精诚合作!自 2016 年起,我同时还参加了中国英语能力等级量表(CSE)的编制工作。因此,又有部分论文是基于这方面的工作,我非常感谢 CSE 项目的支持!CET 和 CSE 是中国英语界无数专家们的心血结晶,是具有中国特色和世界影响力的英语教学科研成果,我能参与其中,并用自己的微薄之力为其添砖加瓦,实乃荣幸之至。现在,我能有机会把我在工作中总结出来的经验教训与同行和后辈们分享、讨论,真是一件非常高兴的事情。

再次对我的论文合作者们表示感谢!

再次对上海交通大学外国语学院和上海交通大学出版社表示感谢!

朱正才

2022 年 7 月 15 日

<b>第一章 语言测试的心理测量模型及相关问题</b> .....	1
第一节 语言听力理解能力的认知结构与测试 .....	3
第二节 朗读与外语能力测量 .....	12
第三节 朗读在外语教学和测试中的作用 .....	26
第四节 大学英语分级教学模式的改革新探 .....	39
<b>第二章 语言测试中的分数等值模型</b> .....	51
第一节 Rasch 模型在大学英语四、六级考试分数等值中的应用 .....	53
第二节 大学英语四、六级考试分数等值研究 .....	66
第三节 大学英语四、六级考试分数的机助百分位等值研究 .....	76
第四节 一个基于综合印象评分法的作文分事后调整模型 .....	86
<b>第三章 语言测试中的信度和效度模型</b> .....	97
第一节 大学英语四级考试通过标准的制定 .....	99
第二节 语言测试的能力结构与因子分析法 .....	107
第三节 中国大学英语考试能力构念三十年之嬗变 .....	120
第四节 选择题的信息熵及其应用 .....	135
第五节 背景知识与听力目的对听力理解影响的实验研究 .....	140
第六节 作文网上评分信度的多面 Rasch 测量分析 .....	151
第七节 高考英语(上海卷)题库命题背景下的校本测验建设 ——基于完形填空题的案例研究 .....	165

第八节	大学英语考试的计算机化组卷探新·····	184
第九节	关于机助自适应大学英语四、六级考试 ——考试效度、信度和施测效率新的平衡·····	196
<b>第四章</b>	<b>语言能力量表模型·····</b>	<b>203</b>
第一节	制定全国统一的语言能力等级量表的原则与方法·····	205
第二节	关于我国英语能力等级量表描述语库建设的若干问题·····	219
第三节	中国英语能力等级量表描述语量表化的可行性 方案探讨·····	232
第四节	中国英语能力等级量表效度研究框架·····	243
第五节	语言能力“能做”描述的原理与方案：以 CEFR 为例·····	261
第六节	英语口语能力描述语因子分析及能力等级划分 ——制定语言能力等级量表实证研究·····	280
<b>附录</b>	<b>专业术语英汉对照表·····</b>	<b>299</b>

## 第一章

# 语言测试的心理测量模型及相关问题

---

**本章导读：**语言能力测试的构念主要包含两类问题：一类是关于语言能力的结构和语言处理过程的认知模型，解决的是语言能力是什么的问题；另一类是这样的一个语言能力应该如何去测量的心理测量模型，也可以说是“一个如何描述它的理论框架”。人的语言能力是认知能力的一种，本身看不见，摸不着，但其外部语言行为表现却是可观察的。设计语言测试的人就是要通过观察人的某些特别典型的语言行为（即语言行为样本）来推断看不见的语言能力水平。本章的四篇论文都围绕这一主题展开，被试对象局限于大学生，每篇论文都是把语言教学与测试问题放在一起讨论，这是完全符合中国国情的一种典型且非常实用的研究范式。





## 第一节

# 语言听力理解能力的认知结构与测试

### 导读

本节试图建立一个简洁而实用的听力理解认知模型,并且使用图式理论来对模型进行解释,接着在这个听力理解认知模型的基础上建立了一个心理测量模型。本节用汉语水平考试(HSK)命题来作实例,详细解释了两类模型在语言考试命题中的作用。笔者认为,掌握这两个最基本的理论模型分类方法和模型的建立方法对一个语言测试命题人员来说是最重要的理论素养。因为语言能力模型的局限性以及语言行为样本的有限性恰恰就是造成语言测试效度问题的两个根本原因,懂得了这两类模型的工作原理和工作机制就有可能找到改善语言测试效度的路径。

当代语言测试的重心正在从狭隘的语言处理能力(language processing ability)转向广阔的语言交际运用能力(language communicative ability)。语言学家们也不仅仅满足于研究语言系统本身,还要探索这个系统在交际中是如何被使用的。不但语音语调、词汇和句法结构的掌握运用,而且人们对过去事物的知识经验、人类的高级认知技能,还有周围环境的影响因素都被包括在交际语言能力测试必须考虑的问题之中。听力理解能力是交际语言能力的首要方面,因此交际听能(communicative listening ability)的训练和测试也就成了人们关注的热点。众所周知,认知心理学是以信息处理理论为基础的。从认知的观点看,语言的听力理解过程是一个信息的输入、处

理和综合运用过程,其中的现代图式理论(Modern Schema Theory)对这一过程具有很强的概括和解释能力。本文就是试图从认知心理学的角度来透视一下听力理解的内部机制,再从交际听能测试的视角提出关于一个听力理解语言能力的认知结构模式。

## 一、听力理解的信息处理过程

听力理解是一个不断聆听音响信息、理解所听声音材料意思的过程。Lado(1961)曾说:“听力理解就是对交际状态中的语言信号成分的认知控制。”Carroll(1972)提出把听力理解区分出两个阶段,即理解语言信息的语言学阶段和理解更广阔的非语言学信息的第二阶段。Gordonwells(1976)也认为,听力理解的发生不但要利用材料中的特别言语间信息,还需要利用相应的世界知识。西方许多从事语言听力理解研究的专家学者也都认识到听力理解并不仅仅是对语音语调、词汇、语法的听辨和理解,而且还有对材料所包含的各种非语言学信息的探究和理解。在真实的语言交际情景中,二者融合于统一的听话过程,而且后者是在前者的基础上发生的。从心理活动水平上说,后者处于更高的认知水平上。另外,听力理解还牵涉听力的目的或动机、发话者与听话者的社会关系、听话情境中可利用的各种非语言信息(如表情、体态等)。某些与听话内容有关的过去生活的知识和体验更是有着举足轻重的作用。正如 Kasper(1986)所说:“当输入信息与原有知识相匹配时,理解就发生了。”

为了建立听力理解测验的心理学基础,许多语言教学专家和测试专家运用现代信息论对听力过程进行了长期不懈的研究(Richards 1983; Anderson 1983,1985 等)。虽然大家在具体划分听力理解过程时仍有不少分歧,但大多数人主张听力理解过程包括两个基本阶段:第一阶段基本上是语言处理过程,第二阶段则是对第一阶段结果的运用。笔者也持这一观点,觉得对听力理解过程的分析不宜太细,并且尝试用图 1-1 对听力理解过程作一个简要的描述。

首先,外部语言信息进入“短时记忆系统”,对信息的处理主要由“中心执行者”完成,中心执行者位于短时记忆系统之中。信息的处理分“自动处

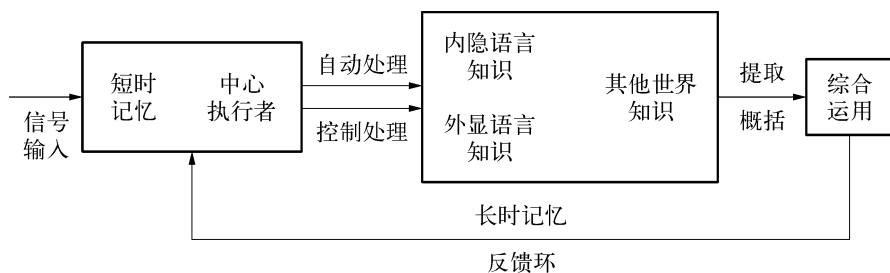


图 1-1 听力理解过程的信息处理模型

理”和“控制处理”，分别形成“内隐语言知识”和“外显语言知识”，并进入“长时记忆系统”之中。同时这些输入信息与长时记忆中的过去知识经验发生连接作用，经过提取和概括形成一个对输入信息的综合理解。整个过程是一个闭环系统，处理中的成功或错误的信息反馈到控制中心，如有必要，由控制中心进行再处理。

## 二、用现代图式理论解释听力理解过程

现代图式理论的典型代表人物 Rumelhart(1980)认为，图式理论基本上是一种关于人的知识的理论。它是关于知识是怎样被表征出来的，以及这种知识的表征如何以其特有的方式有利于知识的应用理论。按照图式理论，所有的知识都组成一定的单元，这种单元就是图式。包括在这种单元之中的东西除了知识本身以外，还有关于这些知识如何被运用的信息。

在以图式为基础的信息加工过程中，图式的活动可以来源于两个方向：一个是从上到下的驱动，也称概念驱动(concept driven)；一个是从下到上的驱动，也称材料驱动(data driven)。在前者中，一个图式可以使属于它的下一级图式活动起来，下一级图式的活动好像来源于一种预期，也称预期驱动；在后者中，下一级图式活动引起上一级图式活动。

结合前文的听力理解信息模型，可以用图式活动来描述听力理解过程。首先，人在感知系统中(譬如说听知觉)接受了某些信息，这就自动地引起了一定的低级水平的图式活动(即特征觉察器的活动)。这些低级水平的图式又可能激发一定的较高水平图式活动，因为这些较低水平的图式是较高水

平图式的一部分。然后,这些较高水平的图式就可能开始概念驱动过程,使那些尚未活动起来的下一级图式处于活动状态,同时评价这些较高水平图式本身的适合性。如果得到了肯定结果,它还会引起更高一级图式的活动。如果图式所要求的感性材料不存在时,就会成为反对这个图式活动的证据。一旦积聚了足够的这类证据,此图式活动就会暂停,而加工活动就会分配给那些更有可能的图式。

在听力理解过程中,被激活的图式使人们会对即将要叙述的内容产生一种预想。当预想同材料所叙述的内容一致时,图式将促进对材料的迅速理解;相反,当图式的预想同实际的叙述不一致时,图式将阻碍对材料的理解。在有些情况下,适合的图式还能使材料中并未出现的内容或听者漏听的内容得到补充,并建立一个完整的意义表征。某一图式一旦被有关的线索激活,还会为信息的加工储存提供一种框架,能被图式组织进来的信息将获得巩固的记忆,而与图式无关、难以组织进来的信息则很容易被遗忘甚至很快丧失。因此,图式活动能引起信息的选择性加工。

根据图式理论,听者对一篇听力文章理解不好可以有以下三种原因:①听者可能不具备适合该文的图式;②听者具备适合该文的图式,但没有被提供一定的线索,无法使这些图式活动起来;③听者可能发现某种对文章的一致解释,但这种解释并非文章的正确意义。根据皮亚杰的观点,图式并不是天生固有的,而是在生活经历中通过同化(assimilation)和调节(accommodation)两种方式发展起来的。同化过程是将感知的事件与现存的图式结合起来,调节过程是在找不到现存适合的图式解释新的情况时,创造新图式或者修改旧图式,使之能与新的情况相符合。

### 三、听力理解能力的测量

把听力理解作为一种技能来进行系统研究在 20 世纪 70 年代之前尚未得到充分重视,然而语言教师越来越相信听力理解是一种能够教授的普遍技能。

如果我们承认语言能力中间能够分离出一个听力理解能力因子,那么听力理解能力是否能再进一步划分呢?只有在对听音理解过程有了一定程

度认识后我们才有可能来回答这个问题。目前听力理解能力常常被视为一种综合性能力,即若干支属能力的总和。这些支属能力的划分不能是无序的,应该有一个统一的维度,在统一的理论框架下共同形成一个互相联系、相互补充的系统。譬如,我们是从语言的组成成分出发还是从语言运用过程出发来研究这个问题就会产生两种完全不同的分类体系。

对综合听力理解能力的认识来源于对听音者理解问题的过程进行观察和探索。如果我们基于这些观察和探索找出了听音者进行听力理解时所需的具体技能的细节,就有可能设计出重点发展这些支属技能的活动方式与练习。随着听音者在这些支属能力上的逐步提高,最终达成综合性听力理解能力。

前文已提出将听力理解过程分为两个阶段。据此我们可以区分出与之相对应的两种不同认知水平的听力理解技能:低级听力技能和高级认知技能。

从图式理论的观点来看,语音处理阶段主要是一些低级水平的声音反应图式在以一种刺激-反应的方式活动,活动的结果是激活了处于较高水平的图式,从而使听力理解过程进入第二阶段。在第二阶段,活跃着的是一些较高水平的知识图式(包括陈述性知识图式和程序性知识图式),只有当输入的信息与这些图式相匹配时,对材料的理解才会发生。

将听力理解过程划分为两个阶段,并据此区分出两级水平的听力技能有着重要意义,因为它将听力测试的目标引入了听力理解过程之中。对具体两个阶段听力理解过程中所含技能的探究是与被试的认知活动研究紧密相连的,这与 Lado(1961)仅从听、说、读、写四个维度上对语言作出语音、词汇、句法及文化背景的划分有着本质的区别。因为该理论不仅关注被试个人的行为结果,而且更加关注有关被试的知识结构、思维特点、听力策略、元认知过程等诸多方面的信息。这类研究不仅为听力测验编制提供理论依据,同时也为达成综合听力理解能力的教学寻找一条如何进行单项训练的途径,也是认知心理学为语言研究打开的一扇新的大门。

就目前的研究现状来说,一般认为在低级听力技能中主要包括辨音能力、辨明语调能力、认识语法成分能力等,在高级认知技能中则主要包括信息选择能力、预想与猜测能力、推理能力、综合归纳能力、元认知能力等。



关知识图式的数量与质量)。还允许有一些题介于两者之间。影响因素则是测试要努力加以控制和操作的变量,以使测试更接近于真实的交际情境。下面试举 HSK 命题实践中的两例进行说明。

**例 1**(男的和女的刚在公司加完班,准备回家)

男:你坐什么车回家?地铁吗?

女:都十一点一刻了,哪儿还有地铁?末班车十点半就开走了。

男:你还不知道啊,从昨天开始,地铁的运行时间延长了。末班车的时间改为十二点半了。

女:是吗?太好了。那头班车的时间是不是也提早了,不是六点了?

男:对,提早半个小时。

问:现在地铁运行时间是几点到几点?(答案:A)

A. 5:30~12:30

B. 4:30~11:15

C. 4:30~10:30

D. 6:00~10:30

这种题目在测试中经常可以看到,语言材料很有生活气息,学生似乎应该听得懂,答得出。可是仔细分析一下,在日常谈话中有很多信息是已知的,如地铁首班车是 6:00,末班车是 10:30,这经常构成谈话的背景知识。因此,在实际谈话中双方都会很自然地把注意力集中在新信息,即新的运行时间上。但作为一个听力测试题,测试目标就显得不集中。首先,听话人可能没有相关背景知识,题目中出现的时间因素都是新信息,一段小对话中出现 5 个时间,这就大大地增加了“记忆”的难度;其次,新的运行时间中,首班车的时间还要从“六点”和“提早半个小时”中推出来,增加了一种数学运算的难度。因此,作为测试题目,其难度不仅要比日常谈话大得多,而且增加的是“记忆”和“数学运算”的难度,与语言交际能力并没有很直接的关系。这类题目距离鉴定和提高学生的语言交际能力的目的就太远了。根据前文分析,笔者认为可将题目作如下修改:

男:你坐什么车回家?地铁吗?

女:都十一点一刻了,哪儿还有地铁?末班车十点半就开走了。

男：你还不知道啊，从昨天开始，地铁的运行时间延长了。末班车的时间改为十二点半了。

女：是吗？太好了。

问：现在地铁末班车是几点？（答案：D）

A. 11:30                      B. 11:15                      C. 10:30                      D. 12:30

这样一改，题眼就集中在地铁末班车的时间上，是一道测试低级听力技能的题目。

### 例 2(在出租车上)

女：师傅，你为什么不走东方路？

男：您看，东方路上堵车堵得很厉害，连自行车都排起了长队，所以我走向阳路。

女：可这样不是在走远路吗？

男：表面上看是这回事儿，不过您瞧，向阳路是机动车单行道，没有自行车，汽车可以开得很快。您放心，这样走，不但不耽误时间，而且车费也和走东方路差不多。

女：好吧。

男：国际贸易大厦到了。

女：这么快！多少钱？

男：26 元。收您 50，找您 24。

女：你估计得真准。平时到这里的车费也是 25 元左右。

问：司机为什么不走东方路？（答案：D）

A. 司机故意走远路                      B. 东方路上骑自行车的人太多  
C. 东方路是非机动车道                      D. 东方路交通堵塞

例 2 也是一个很有生活气息的题目，情境在出租汽车上，乘客和司机以行车路线为谈话的主题是很自然的。司机要向乘客解释为什么不走东方路和走向阳路要不要多付车费这两个问题。而题目要求学生回答的也正是这两个方面的内容之一。这类测试完全符合我们在教学中进行听力理解训练