

汉语口语测试 任务难度影响因素探究

聂丹◎著

口语测试任务研究关系到测试效度和信度两大核心问题，但长期



北京语言大学出版社

BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

中央高校基本科研业务费专项资金资助

Supported by "the Fundamental Research Funds for the Central Universities"

汉语口语测试 任务难度影响因素探究

聂 丹 / 著



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

图书在版编目 (CIP) 数据

汉语口语测试任务难度影响因素探究 / 聂丹著.

— 北京: 北京语言大学出版社, 2012 重印

ISBN 978-7-5619-3288-9

I. ①汉… II. ①聂… III. ①汉语—口试—影响因素—研究 IV. ①H193.2

中国版本图书馆 CIP 数据核字 (2012) 第 096101 号

书 名: 汉语口语测试任务难度影响因素探究

责任编辑: 王墨妍

责任印制: 姜正周

出版发行: 北京语言大学出版社

社 址: 北京市海淀区学院路 15 号 邮政编码: 100083

网 址: www.blcup.com

电 话: 发行部 82303650 / 3591 / 3651

编辑部 82301016

读者服务部 82303653 / 3908

网上订购电话 82303668

客户服务信箱 service@blcup.net

印 刷: 北京中科印刷有限公司

经 销: 全国新华书店

版 次: 2012 年 6 月第 1 版 2012 年 10 月第 2 次印刷

开 本: 787 毫米 × 1092 毫米 1/16 印张: 25.75

字 数: 420 千字

书 号: ISBN 978-7-5619-3288-9 / H·12

定 价: 58.00 元

凡有印装质量问题, 本社负责调换。电话: 010-82303590

序

聂丹是我的第一个博士生。她考博那年正在新疆支教，不远万里两次赴京赶考，一举中第。从此便有了我们的师生缘分。

聂丹是在职攻读的博士，平时有自己的一摊工作和教学任务，加之孩子小、家庭责任重，忙碌自不必说。我一向崇尚“因材施教”，便把聂丹划入自我要求很高的那一类学生，对她无需督促，反而要不断给她减压。事实证明，我的判断不错。不用我要求，她总是自我鞭策，精益求精。我担心她身体吃不消，反倒时常叮嘱她别太劳累。

现在呈现在读者面前的这部《汉语口语测试任务难度影响因素探究》，是聂丹在其博士论文基础上修改而成的。该书以汉语口语测试任务难度为研究对象，深入探讨影响口语测试任务难度的主要因素，研究重点是话题、体裁、输入这三大因素对任务难度的作用机理，研究目标是在任务难度与汉语口语能力之间建立联系，进而构建评估汉语口语测试任务难度的理论框架。我们知道，口语测试是检验言语习得实际水平和言语表达能力的重要手段，但由于口语测试任务设计的主观性及其操作上的复杂性，一直以来，学界对口语测试任务的研究都比较缺乏，而如何确定口语测试的任务难度、准确检测应试者的口语能力等级，尤其是语言测试实践与研究中的一大难题，这方面的理论和实证研究都十分薄弱。聂丹迎难而上，选择了这一颇具挑战性和前瞻性的课题进行积极的探索，其研究成果对于推动口语测试任务难度研究和汉语口语测试实践的科学发展，对于改进汉语口语教学，具有理论建设意义和实际应用价值。

依我所见,该书有四个主要特色:

第一,具有明显的跨学科性。该书尝试运用多学科相关理论作为研究基础,其理论建构的视角比较开阔。具体来说,它以语言测试的理论和方法为主导,综合运用了语言习得与教学、认知心理学、应用语言学等多学科的相关理论和方法,对影响汉语口语测试任务难度的因素进行了多角度的考察和研究。这一做法使任务难度研究的理论基础比较深厚,增强了研究的系统性和解释力。

第二,具有一定的开拓性。一方面,选题具有开拓性。目前在国内尚没有集中研究汉语口语测试任务难度的相关成果,该书可称得上是一个开创性研究。另一方面,在任务难度评估框架的构建思路上具有一定创新性。该书没有面面俱到地考察所有影响口语测试任务难度的因素,而是通过对历史文献的梳理分析和广泛的调查研究,提取出三个最主要的影响因素,进而深入细致地分别研究这三大因素作用于任务难度的内在规律,并分别设计了评估三大因素难度的量化模型,在此基础上,寻找口语测试任务难度与口语能力等级标准之间的对应关系,最终构建了评估汉语口语测试任务难度的综合量化分析框架。整个研究过程体现了由表及里、点面交叉、理论与实证研究结合、定性与定量分析并用的方法特点。

第三,具有较强的实证性。该书依托于大量第一手的汉语口语测试实测样本,取样丰富,材料翔实,在构建任务难度评估框架上运用了大量的实证研究手段,特别是对于同一研究目的往往采用多种手段进行研究,互相印证和补充,这样得出的结果就比较客观、可信。

第四,突出了汉语口语测试的中心地位。长期以来,有关口语测试的大量研究主要着眼于英语或其他语种的测试。而该书以汉语的口语测试为研究对象,同时将汉语作为第二语言的口语测试和汉语作为第一语言的普通话测试联系起来,共同探讨这两种测试的任务难度设计问题。这种做法具有一定的探索性,启发研究者更多地关注这两种测试的内在区别和联系,使汉语口语测试形成一个统一的体系,并有望推动国内汉语口语测试的相关研究。

聂丹的博士研究方向是普通话测试研究,通过三年的刻苦学习,她对普通话测试的研究和实践有了较多的思考和认识。同时,她又是北京

语言大学汉语水平考试中心的副研究员、硕士生导师，对汉语作为第二语言的测试和教学具有一定的理论素养和实践经验，并能够掌握丰富的测试资源。此外，她读本科和硕士研究生期间，接受过语言学及应用语言研究的系统训练，对语用学、话语分析、社会语言学等学科的理论和方法颇有了解。这样的研究背景使她能够融通运用多学科的理论和方法来研究口语测试任务难度问题，从而形成了该书的跨学科性强、理论基础丰厚、实证研究丰富、一语测试与二语测试并重等突出特点。

聂丹的博士学位论文曾获得以陈章太先生为首的答辩委员会全体专家很高的评价，一致推举为优秀博士论文。这次北京语言大学青年学者文库基金决定赞助该书出版，也是对她这一研究成果的肯定。不过，该研究还有一些问题可以进一步探讨。比如，该书所构建的汉语口语测试任务难度评估框架还可以进一步细化，有条件的话还可以继续探索影响口语测试任务难度的其他因素。另外，汉语口语测试任务难度评估框架的有效性和可行性还需要大量的实践检验，这也是后续研究的一个重要方向。对上述问题聂丹有着清醒的认识和后续研究的规划，相信这部著作会是她日后研究的一个起点。

聂丹好学多思，勤勉有加，具有较强的独立研究能力，凡事追求尽善尽美。她在工作、家庭两不误的情况下，笔耕不辍，取得了一些研究成果，这是很不容易的。我们有理由期待她今后不断取得更多的成绩，创造出更好的成果！

是为序。

姚喜双

2012年4月

内容摘要

口语测试因在保障信度和实际操作上所遇到的巨大困难，成为语言测试理论和实践中非常薄弱的一个领域。20 世纪 70 年代以来，关系到口语测试效度和信度的任务研究开始受到关注。所谓口语测试的任务 (task)，就是在测试环境下要求被试使用目标语口头完成的测试作业，其功能是用于引导被试说出可供评分的话语，以此推论被试在目标语交际环境中的口语能力。任务难度则是衡量任务质量的一个重要指标。在现有的口语测试实践中，对任务难度的评估和调控主要不是通过科学量化的手段，而是凭借设计者的主观经验，这在相当程度上影响了任务质量的提高，而口语测试任务难度研究的滞后是制约测试任务质量的关键因素。有鉴于此，本书将研究目标确定为，以汉语口语测试为对象，深入探究影响口语测试任务难度的主要因素，构建评估口语测试任务难度的理论框架。本书的基本研究路线是，依托大量第一手的汉语口语测试实测样本，将理论分析与实证研究紧密结合，以调查、统计和实测样本的话语分析作为实证研究的主要方法。我们相信，此项研究对口语测试的理论和实践发展均具有重要意义，特别是对方兴未艾的汉语口语测试的科学发展更具现实意义。

本书通过对任务及任务难度在国内外研究现状的梳理和分析，建立了由语言测试、第二语言习得、认知心理学、语言学（包括语用学）等多学科相关理论构成的研究基础。借助于 Bachman (1990) 的测试方法因素框架和 Skehan (1996, 1998) 的任务难度认知框架，综合前人研究成果，提出了影响口语测试任务难度因素的理论假设。在较广泛的问卷

调查与分析基础上,进一步发现,话题、体裁、输入这三大因素是决定汉语口语测试任务难度的关键性因素,这是因为各因素内部存在着影响信息加工过程、进而影响任务难度的若干分析性特征。在理论假设和实际调查基础上,提出了影响汉语口语测试任务难度的主要因素及其内部分析性特征假设。

基于上述假设,本书集中探讨了话题、体裁、输入这三大因素对任务难度的影响。通过理论分析和多角度的实证研究,揭示了三大因素作用于任务难度的内在运作机制,剖析了各因素内部分析性特征影响任务难度的规律,并逐一构建了衡量话题难度、体裁难度和输入难度的量化评估模型。结合汉语口语教学与测试实践,建立了汉语口语能力等级描述标准,然后将该标准与三大难度评估模型建立对应关系,以此作为汉语口语测试任务难度分级评估的依据。三大难度评估模型与汉语口语能力等级标准综合构成了汉语口语测试任务难度评估框架。

通过对具体汉语口语测试任务难度的实证检验,我们验证了该综合框架的可靠性和有效性,不仅可以用来评估汉语口语测试特定任务的难度,同时还可以评价任务难度对特定等级考试的适切性。当测算难度与考试的目标难度不符时,该框架能够提供影响任务难度的内部特征信息,据此施行有针对性的调控措施,具有一定的诊断功能和操作性。

我们相信,本研究所构建的汉语口语测试任务难度评估框架对于科学地设计、评价及改进汉语口语测试任务具有一定的指导作用,其构建策略和方法带有一定的创新性,对相关任务研究具有借鉴意义。同时我们也认识到,作为汉语口语测试任务难度研究的一部抛砖之作,该框架仍然有待进一步完善,其应用范围也有待测试实践的检验,今后还需要作更加深入的后续研究。

目 录

第一章 绪论	1
1.1 选题缘起	1
1.1.1 口语测试的发展	1
1.1.2 口语测试任务研究现状	3
1.1.3 口语测试任务研究构想	5
1.1.4 口语测试任务难度研究的必要性	7
1.2 研究思路	9
1.2.1 目标与方法	9
1.2.2 内容与步骤	10
1.3 研究意义	12
1.3.1 理论意义	12
1.3.2 实践意义	13
第二章 口语能力及其测评	15
2.1 引言	15
2.2 口语与口语习得	15
2.2.1 口语	15
2.2.2 口语习得	18
2.3 口语能力与口语测试	28
2.3.1 “说话”的构念	28
2.3.2 语言测试中的口语能力观	30
2.4 国外口语测试的实践发展	37
2.4.1 历史简述	37
2.4.2 测试组织形式	38

2.5	国内汉语口语测试的实践发展	43
2.5.1	汉语口语测试的范围	43
2.5.2	汉语作为第一语言的口语测试	44
2.5.3	汉语作为第二语言的口语测试	47
2.6	汉语口语能力等级标准描述	54
2.6.1	对外汉语教学分级目标	54
2.6.2	汉语水平测试分级描述	57
2.6.3	汉语水平等级标准和等级大纲	60
2.6.4	汉语口语能力等级标准归纳	63
2.7	本章小结	65
第三章	任务及其难度	67
3.1	引言	67
3.2	任务的含义	68
3.2.1	一般意义的“任务”	68
3.2.2	语言教学的“任务”	68
3.2.3	语言测试的“任务”	71
3.2.4	口语测试的“任务”	72
3.3	口语测试任务研究述评	75
3.3.1	研究背景	75
3.3.2	国外研究	76
3.3.3	国内研究	89
3.4	难度与任务难度	93
3.4.1	难度	93
3.4.2	任务难度	96
3.5	任务难度研究述评	97
3.5.1	国外研究	98
3.5.2	国内研究	108
3.6	本章小结	111
第四章	研究假设	114
4.1	引言	114
4.2	理论依据	115

4.2.1	测试方法因素框架	115
4.2.2	任务难度认知框架	121
4.3	实证依据	122
4.3.1	目的和方法	122
4.3.2	调查一	123
4.3.3	调查二	129
4.3.4	调查三	140
4.4	本章小结	145
第五章	话题因素	149
5.1	引言	149
5.2	话题及话题知识	149
5.2.1	话题的相关定义	149
5.2.2	话题知识及其作用	153
5.3	产出型语言运用的话题研究	156
5.3.1	关于话题与测试构念	156
5.3.2	关于话题类型	158
5.3.3	关于选择话题的原则	158
5.4	相关理论对话题设计的启示	160
5.4.1	最近发展区理论与话题适切性	160
5.4.2	近体原则与话题相关性	162
5.4.3	图式理论与话题熟悉度	163
5.4.4	关于话题难度	165
5.5	话题选择和分级实践	165
5.5.1	国外大规模外语口语测试的话题设计	165
5.5.2	汉语作为第二语言口语教学及测试的话题设计	168
5.5.3	汉语作为第一语言口语测试的话题设计	175
5.6	话题难度的实证研究	176
5.6.1	研究一	176
5.6.2	研究二	179
5.6.3	研究三	182
5.6.4	研究四	185

5.6.5	研究五	186
5.7	话题难度的评估模型	194
5.8	本章小结	196
第六章	体裁因素	198
6.1	引言	198
6.2	体裁及其由来	198
6.2.1	体裁及相关概念	198
6.2.2	体裁的由来	203
6.3	体裁分类研究	205
6.3.1	以目的为标准	205
6.3.2	以内容为标准	206
6.3.3	以时间关系为标准	206
6.3.4	以语篇结构为标准	207
6.3.5	以整体功能为标准	208
6.3.6	以认知过程为标准	208
6.3.7	四种体裁的再分类研究	210
6.3.8	本研究采用的体裁分类方法	212
6.4	体裁难度层级研究	213
6.4.1	体裁思维复杂度层级	214
6.4.2	体裁结构复杂度层级	216
6.4.3	体裁语言复杂度层级	219
6.5	体裁难度层级实践	220
6.5.1	体裁难度层级确立的重要性	221
6.5.2	国外口语测试的体裁难度层级	222
6.5.3	国内对外汉语口语教学及测试的体裁难度层级	222
6.5.4	国内普通话水平测试的体裁难度层级	224
6.6	体裁难度层级的实证研究	225
6.6.1	研究一	226
6.6.2	研究二	232
6.6.3	研究三	233
6.6.4	研究四	239

6.6.5 研究五	242
6.6.6 研究六	244
6.7 体裁难度的评估模型	249
6.8 本章小结	250
第七章 输入因素	253
7.1 引言	253
7.2 输入及其构成的不同界定	254
7.2.1 Davies 等的相关界定	254
7.2.2 Bachman 的输入构成	256
7.2.3 李筱菊的相关界定	261
7.3 本研究的输入及其构成	264
7.3.1 语境信息	265
7.3.2 支持信息	270
7.3.3 输入语言	274
7.4 可理解输入与互动理论	276
7.5 实证研究	278
7.5.1 研究一: 语境信息	279
7.5.2 研究二: 支持信息	284
7.5.3 研究三: 输入语言	288
7.5.4 研究四: 输入调整	293
7.6 输入难度的评估模型	310
7.7 本章小结	314
第八章 任务难度评估框架	316
8.1 引言	316
8.2 任务难度评估框架的确立	317
8.2.1 话题难度框架	317
8.2.2 体裁难度框架	318
8.2.3 输入难度框架	318
8.2.4 汉语口语测试任务难度分级评估	321
8.3 任务难度评估框架的应用	323
8.3.1 任务样本的选取	323

8.3.2 难度分析员的选定	326
8.3.3 方法与步骤	326
8.3.4 结果与分析	327
8.3.5 小结	332
8.4 相关建议	332
8.4.1 明确测试目的	332
8.4.2 了解目标群体	333
8.4.3 避免个人评估	333
8.5 本章小结	334
第九章 结语	335
9.1 基本观点和结论	336
9.1.1 基本观点	336
9.1.2 主要结论	339
9.2 不足与后续研究	341
参考文献	344
附录	356

图表目录

一、图

图 1.1	测试任务、被试口语表现、分数、被试口语水平的关系	4
图 1.2	口语测试任务研究内容框架	6
图 4.1	七类因素对口语表现的影响力	127
图 5.1	话题难度的 9 个分析性特征影响程度序列	184
图 5.2	话题难度的 6 个分析性特征影响程度序列	184
图 5.3	话题难度区间	195
图 6.1	不同说明模式的思维方式	212
图 6.2	叙述结构	218
图 6.3	论证结构	219
图 6.4	议论语篇结构框架	223
图 6.5	八大认知类型难度感知排序	237
图 6.6	八大认知类型实测难度排序	242
图 6.7	六大认知类型难度感知排序	247
图 6.8	体裁难度区间	250
图 7.1	任务构成及关系	253
图 7.2	调整输入的手段	310
图 7.3	输入难度区间 (二语)	313
图 7.4	输入难度区间 (一语)	313
图 8.1	话题难度区间	317
图 8.2	体裁难度区间	318
图 8.3	输入难度区间 (二语)	320

图 8.4	输入难度区间 (一语)	320
图 8.5	输入难度二次划分区间 (二语)	321
图 8.6	输入难度二次划分区间 (一语)	321
图 9.1	任务难度双重属性在研究中的体现	337

二、表

表 2.1	CLA 模型构成	33
表 2.2	描述口语交际能力的框架	35
表 2.3	口语交际能力的构成	36
表 2.4	结构主义口试与交际口试比较	37
表 2.5	HSK [改进版] 口试设计	50
表 2.6	语言要素和语言功能分级目标 (汉语言专业本科)	54
表 2.7	口语课程设置及教学目标 (汉语言专业本科)	55
表 2.8	语言要素和语言功能分级目标 (长期进修)	56
表 2.9	口语教学及成绩测试目标	56
表 2.10	HSK 系列考试分级描述	58
表 2.11	HSK [改进版] 系列考试分级描述	58
表 2.12	C.TEST 口语考试分级描述	59
表 2.13	《标准 and 大纲》的基本框架和主要内容	61
表 2.14	汉语水平、等级标准与对外汉语教学年级、HSK 的关系	62
表 2.15	汉语口语能力五级标准描述	64
表 3.1	语言任务特征框架	77
表 3.2	口语测试任务特征描述	81
表 3.3	描述口语测试任务的框架	82
表 3.4	口语测试任务类型描述	83
表 3.5	改变任务难度的方法	99
表 3.6	Skehan (1996, 1998) 任务难度认知框架	100
表 3.7	影响外语学习任务难易度的框架	109
表 3.8	外语学习任务的分级	109
表 4.1	测试方法因素框架	115
表 4.2	口语测试任务难度影响因素简表	118

表 4.3	口语测试任务难度主要影响因素的分析性特征简表	122
表 4.4	考生群体基本信息 (调查一)	123
表 4.5	话题难度	126
表 4.6	任务类型难度	126
表 4.7	影响口语表现的因素	126
表 4.8	对三种考试形式的评价	128
表 4.9	对外汉语教师群体的基本信息	130
表 4.10	学生群体的基本信息	132
表 4.11	口语测试任务难度影响因素定性分析结果	134
表 4.12	42 名教师答卷与 58 名学生答卷难度影响因素调查结果	134
表 4.13	42 名教师答卷与 56 名学生答卷 第 6 题难度影响因素排序结果	137
表 4.14	口语测试任务难度影响因素排序汇总	139
表 4.15	调查对象基本信息 (调查三)	141
表 4.16	影响口语测试任务难度的因素排序	143
表 4.17	影响口语测试任务难度的主要因素	144
表 4.18	影响汉语口语测试任务难度的主要因素 及其运作机理的研究假设	147
表 5.1	《发展汉语》话题设计	171
表 5.2	《标准和大纲》的话题内容	175
表 5.3	五组话题的平均难度	177
表 5.4	同一体裁下的话题难度比较	178
表 5.5	对五种话题名称的难度评价	179
表 5.6	三大类话题的难度	181
表 5.7	20 个话题的难度	181
表 5.8	20 个话题的难度区间分布	181
表 5.9	9 种分析性特征的被认同程度以及被选中频次	183
表 5.10	PSC “说话项” 题目难度层级设置类型	188
表 5.11	PSC “说话项” 30 个题目难度分级	188
表 5.12	PSC 三类话题难度	191
表 5.13	难度排序的理由	192