

不同时间计划条件下的 独白性口试任务

Monologic Oral Test Tasks
under Different Planning Conditions

—— 兰春寿 著 ——

世界图书出版公司

福建师范大学外国语言与文学研究中心资助出

**Monologic Oral Test Tasks
under Different Planning Conditions**

**不同时间计划条件下的
独白性口试任务**

兰春寿 著

世界图书出版公司

西安·北京·广州·上海

图书在版编目(CIP)数据

不同时间计划条件下的独白性口试任务:英文/兰春寿
著.—西安:世界图书出版西安公司,2008.9

ISBN 978-7-5062-9840-7

I. 不... II. 兰... III. 英语—口语—教学研究
IV. H319.9

中国版本图书馆 CIP 数据核字(2008)第 116674 号

不同时间计划条件下的独白性口试任务

作 者 兰春寿

责任编辑 陈康宁

封面设计 Look. Book 飞洋设计机构

出版发行 世界图书出版西安公司

地 址 西安市北大街 85 号

邮 编 710003

电 话 029-87214941 87233647(市场营销部)
029-87232980(总编室)

传 真 029-87279675

经 销 全国各地新华书店

印 刷 西安市建明工贸有限责任公司

开 本 880×1230 1/32

印 张 8

字 数 220 千字

版 次 2008 年 9 月第 1 版 2008 年 9 月第 1 次印刷

书 号 ISBN 978-7-5062-9840-7

定 价 30.00 元

☆如有印装错误,请寄回本公司更换☆

序

考试的信度和效度是语言测试的两大基本概念,而近年来语言测试界的很多研究更多的是关注语言测试,特别是大规模、高风险的语言测试的效度问题。在国内,计算机化高考英语口语考试已经在广东省实行了。随着计算机化语言测试的发展,其效度问题,尤其是计算机化口试的效度问题已引起语言测试研究者们的关注。在计算机化口语考试中,任务计划条件是口语考试命题人员必须要考虑的一个因素。从 Fulcher(2003^①)的口语能力构念框架看来,任务计划条件是一种与口语能力构念无关的因素(construct-irrelevant factors)。根据 Messick(1989^②)的语言测试效度理论,在计算机化口语考试中与口语能力构念无关因素的影响应尽可能降低,从而最大程度地保证学生的口试结果真实反映他们的口语能力水平。为此,首先要明确的问题是施与考生不同的任务计划条件是否影响考生的口试结果?换言之,考生在不同的任务计划条件中的口试结果所推论出的口语能力是否有效?兰春寿博士论文的研究正是以此为出发点验证口试任务在不同时间计划条件下的效度问题。

兰春寿的研究体现了语言测试研究领域所关注的热点问题,即测

① Fulcher, G. 2003. *Testing Second Language Speaking*. Pearson Education Limited.

② Messick, S. 1989. Validity. In R. L. , Linn (Ed.), *Educational Measurement* (3rd ed). New York: Macmillan.

试的效度在考试任务中的验证。论文研究针对的是具有重要社会影响的高考英语口语,其研究结果不仅对相关领域的学术研究有一定的启示,也对我国大规模的英语考试,特别是英语口语的设计和改进行具有积极的实际意义。

兰春寿对课题所涉及的相关理论和研究有较充分和准确的理解和把握,对测试效度、口试任务设计与交际策略使用、口语能力等语言测试学相关概念,以及话语产出、(语言)信息处理、记忆能力等心理学或心理语言学方面的理论和研究状况及对课题的影响都做了综合的评述。

论文研究采用了与传统的效度研究方法不同的基于论证的效验方法(argument-based validation approach)来验证该口试任务在不同条件下的效度。这一方法基于效验即论证(假设检验)的思想,与效度整体观一脉相承。效度整体观认为效度是从总体上评价所提议的测试分数的解释和使用的合理性。效验的对象是分数的解释(包括推论和决定),既要对所提议的测试分数解释提供足够的证据,也要考虑到其他可能的解释,并证明其不合理性。一般来说,效验论证包括两个步骤:第一,提出效验观点,包括两方面的问题:试图对测试分数做哪些解释和使用?还有哪些与预期相反的解释和使用?第二,收集有关证据支持所提议的解释与使用,反驳相反的观点。收集效度证据的方法之一是试题或任务的实验分析:定量的方法可以探究在多大程度上相关的因素影响了题目的难度和区分度,定性的方法可以研究受试在测试中使用的策略,也可以使用话语分析法研究受试的“产品”(产出的语言),并将其语言和语用特征与构念定义加以比较。基于论证的效验法的优点主要包括:(1)适用于任何种类测试的解释和使用;(2)尽管解释性论证不能对效度做出任何绝对决定,但可以在整体上提高该分数解释的合理性,因为大多数推论和假设都经过检

验,得到支持或改进。鉴于此,兰春寿在围绕其研究的中心选题所提出的研究问题的基础上提出了四种相关的效验假设,并以多种类型的数据(如:口语评分分数、口语语篇标注编码、问卷调查等)和多种角度的分析(如:分数统计分析、语篇准确度、流利度、复杂度分析、交际策略分析等)展开效验论证。通过对两组英语水平不同的考生在两种不同的时间计划条件下所表现的口语能力分析,探讨了该口试任务的效度问题以及相关影响因素。研究思路清晰,研究方法合理,研究结果可靠。

值得一提的是,为进一步获取有关证据支持所提出的效验假设,论文研究采用了话语分析法分析考生口试话语的准确度、流利度和复杂度的语言特征以及交际策略的使用特点。话语分析作为一种语言研究方法,形成于20世纪70年代,之后日益受到国际语言学界的高度重视。作为一种语言研究方法,话语分析被许多语言学者认为能弥补传统语言研究的许多不足,因此被用以研究各种语言。近十几年来在国外语言测试研究领域有不少的语言测试研究者将话语分析的理论、方法应用于口试效度验证研究,并因此而解决了不少口试效度验证方面的问题。将话语分析应用于口试效度验证方面的研究应该是得益于效度概念的演变发展。自从20世纪30年代提出以来,效度概念经过不断演变,在近半个多世纪取得了重大进展。如同在教育测量的其他领域,效度已经不再被看作是一套“可测量的”现象,而是一种基于从许多不同领域获取证据的复合的整体的判断。在这种重新界定的概念中,所需要的是从过程和结果两方面来充分理解考试。话语分析的研究方法所具有的特殊价值正是体现在这里。

此外,作者对所获数据进行了详细的分析和解释并做出了一些结果与先前研究相同,另一些则与先前研究有别的有说服力的结论。作者围绕论文的中心研究问题对研究结果进行了较充分的阐述和讨论,

并与已有的相关同类研究进行了比较和分析。从中可以看出论文的一些研究结果在相关研究领域具有一定的创新性,也为使用基于论证的效验方法对某一语言测试任务的分数解释和使用进行论证提供了榜样。总之,兰春寿的研究体现了相关研究领域的前沿性,在理论上为效验探索提供了经验,在实践上为我国的大规模计算机化英语口语任务设计提供了参考。

曾用强

2008年8月于广州

前 言

本书是在作者博士论文的基础上修改而成。本书主要研究对考生在不同时间计划条件下的独白性口试任务中的表现所作推论的有效性。本书在综述有关任务条件的研究、信息处理模型,语言测试效度验证等方面研究的基础上提出了效验假设。通过实验从口试评分分数,考生话语的流利度、准确度和复杂度等语言特征分析,交际策略分析以及问卷分析等方面所获取的数据对两组英语水平不同的考生在两种不同的时间计划条件下所表现的口语能力的研究,探讨了该口试任务的效度问题以及相关的影响因素。

本书共7章。第1章引言,介绍研究的背景、目的、关键术语的操作化定义以及研究问题等;第2章文献综述,从口语测试理论模型的扩展、口试任务条件的研究以及语言信息处理模型等方面对相关理论和研究进行了阐述和总结;第3章介绍语言测试中基于论证的效验方法,阐明该效验方法的优点以及通过该方法所获取的效验证据对本研究的口试任务效度验证的意义,并在此基础上提出了四个效验假设;第4章描述实验设计过程,相关数据的收集和分析方法;第5章结合研究问题进行实验数据的分析;第6章对研究结果进行讨论和解释。第7章是结论和启示,总结了研究结果并分析这些结果对口试效验研究和口语教学的启示,同时指出本研究存在的局限性以及将来进一步完善研究的方向。

回顾自己博士论文的研究和撰写过程,总会想到这么一首诗:“静伴山河翠,禅房花木深;心中有万卷,何时不歌声。”这是笔者2002年9月到广东外语外贸大学访学期间认识的太原师范学院外语系的王维老师,获悉笔者要准备博士研究生入学考试时,通过短信发来的一首诗。当时正逢非典时期,考试延期,王老师的这首诗让笔者感触颇深,并用以自勉。在随后艰辛的学习日子里,每当博士论文研究过程中遇到挫折,心情浮躁之时,这首诗总能让笔者找到一种无形的精神力量,鼓励自己:坚持!坚持!再坚持!适逢自己的博士论文即将出版,笔者在此对王维老师表示衷心感谢。

论文研究在选题时得到了我的导师曾用强教授的悉心指导。在广东外语外贸大学的桂诗春教授、亓鲁霞教授、刘建达教授等也对选题提出了许多建设性的意见,使笔者深受启发,不至于在研究过程中走弯路。

实验研究中的计算机化口试任务设计得到了广东启明科技发展有限公司的大力支持。公司免费提供计算机化口试系统软件供笔者研究使用,工程师卢经先生帮助设计实验中的口试任务,并在实验过程中提供技术支持。实验在福建师范大学外国语学院进行,学院的杨本生老师帮忙安排语言实验室和设备调试,保障实验的顺利进行。实验结束后,曾美英、李玉平、张昌宋、方艳、邱志芳、刘茵、洪梅和蔡巧莺等老师在百忙中抽出宝贵的时间帮助笔者对口试录音进行评分。

当然,在论文实验设计、撰写和修改过程中离不开导师曾用强教授的适时帮助和建议。他丰富的学识和缜密的思维总能在我的论文研究和写作过程中遇到难题时给我启发。广东外语外贸大学的博士生导师吴旭东教授也对论文提出了建设性的修改意见。本书的修改还得益于论文答辩委员会的张文霞教授(主席)、桂诗春教授、何莲珍教授、何广铿教授和刘建达教授对论文所提出的许多批评和建议。笔者

在广东外语外贸大学学习期间还得到了学友胡春雨、黄莹、徐蔓菲、张巨文、袁传友、左红珊、陈念宁、张新玲、李光敏、王永庆、李清华、王敏、李光泽、李亮等的鼓励和帮助。

笔者对这些师长和朋友们的关心和帮助表示衷心感谢。

本书的出版得到福建师范大学外国语言与文学研究中心的资助。笔者单位的领导也对本书的出版给予大力支持和帮助。林大津院长、黄远振副院长、李荣宝副院长以及薛常明副院长等都非常关心笔者的博士论文撰写以及出版事宜。尤其是薛常明博士对本书的研究提出许多宝贵的建议。笔者在此一并表示诚挚的谢意。

在笔者攻读博士学位期间,家人的关爱与支持才使得论文研究得以顺利完成。笔者的妻子兰善萍女士既要顾及繁忙的工作又要照顾年幼的孩子,是她的理解和辛劳付出激励自己克服困难完成论文的撰写。年迈的岳父母对笔者的学习也极力支持,是他们对我家庭的照顾使我能够全身心投入学习和研究中。在此,对所有关心我的亲人表示深深的敬意。笔者永远怀着一颗感恩的心。

由于笔者的学术功底和理论水平有限,本书虽经多次修改,仍难免有错漏和缺陷。对于书中的任何错漏和缺陷由笔者承担全部责任,恳请学界同仁和读者朋友批评指正。谢谢!

兰春寿

2008年8月于福州

ABSTRACT

A recent innovation in language testing in China involves the use of computerized oral tests. Studies have shown that task planning conditions have particular relevance to the development of a computerized test of speaking ability in which monologic production tasks are commonly adopted. The inclusion of planning time during task performance is considered likely to affect task difficulty and hence a potential source of construct-irrelevant difficulty or easiness. Therefore, any attempt to provide test takers with a certain amount of task planning time cannot proceed until it is justified by the language outcome.

This study, accordingly, examined the validity of making inferences from test scores elicited from test takers' monologic oral task performance under different task planning conditions. The present study focused on four assumptions which have important implications for validity. The four assumptions are: (1) there will be no differences in the oral test scores awarded to the same test takers performing a monologic oral test task under two different planning conditions; (2) there will be no differences in linguistic features (fluency, accuracy and complexity) when test takers perform a monologic oral test task under two different planning conditions; (3) there will be different use of communication strategies by test takers in order to achieve their communicative goals across two different planning conditions; and (4) the interactive effect of English proficiency and task planning conditions will result in

different task performance. Using a counterbalanced design for task planning conditions and English proficiency level, this study compared the performances of 120 full-time undergraduate English majors, of which sixty were freshmen and the other sixty were seniors. The test was administered in a language lab where a computerized oral test system was available. Test takers were required to produce oral narratives from video clip prompts that had been designed to differ in task planning conditions. The data were gathered both by subjectively rating the speech samples directly using analytic rating scales for fluency, accuracy, and complexity and by counting discourse features through use of discourse analysis procedures.

Results showed that: 1) high-level test takers achieved significantly higher complexity scores under the strategic planning condition than under the pressured planning condition, while low-level test takers were awarded significantly higher scores in terms of fluency, accuracy and complexity under the strategic planning condition than under the pressured planning condition; 2) compared with the pressured planning condition, providing an opportunity for strategic planning was beneficial to both low- and high- level test takers in producing significantly more syllables, meaningful syllables, error-free clauses and correct verb-forms, with the exception of syntactic complexity and lexical variety, two linguistic measures of complexity; 3) test takers tended to employ achievement strategies to produce more accurate speech and to rely on reduction strategies to achieve more fluent use of language. In particular, under the pressured planning condition they relied more on reduction strategies to achieve their communicative goals; 4) there were no interactive effects of English proficiency levels and task planning conditions on any one of the three areas of monologic oral test performance, namely,

fluency, accuracy, and complexity.

In conclusion, the evidence in this study only provided support for the third assumption, that is, test takers employ different communication strategies under different planning conditions in order to achieve their communicative goals. These differences are a reflection of the differences in test takers' strategic competence, suggesting that, at the level of task planning, the relationship between strategic competence and language competence is closely related. On the other hand, it reflects differences in how test takers at different levels of proficiency utilize their strategic competence. However, the lack of evidence in support of the other three assumptions requires us to conclude that test-takers' discourse produced under different task planning conditions differed in many respects which would, thus, preclude making inferences about their oral abilities on the basis of their monologic oral task performance under different task planning conditions. Consequently, when designing monologic oral test tasks the provision of a certain amount of task planning time to test takers needs to be considered carefully in connection with their levels of English proficiency. The test scores thus yielded needs to be interpreted with particular reference to task planning conditions.

Contents

序	I
前 言	V
ABSTRACT	IX

Chapter 1	Introduction	1
------------------	---------------------	----------

1.1	Background to the Study	1
1.2	Rationale for the Present Study	3
1.3	Operationalizing Task Planning Conditions	7
1.4	Operationalizing the Speaking Construct for Monologic Oral Production Subtest	10
1.5	Language Ability and Test Performance	14
1.6	Key Research Questions	16
1.7	Organization of the Dissertation	23

Chapter 2	Review of Literature	25
------------------	-----------------------------	-----------

2.1	Model of Oral Test Performance	25
2.2	Studies on Task Planning Conditions	33
2.2.1	Theoretical Perspective of Research on Task Planning Condition	34

2.2.1.1 Attention and Noticing	34
2.2.1.2 Working Memory	36
2.2.1.3 Focusing on Form	37
2.2.2 Research on Task Planning Condition in SLA	42
2.2.2.1 Within-Task Planning	42
2.2.2.2 Pre-Task Planning	44
2.2.3 Research on Task Planning Condition in Language Testing	46
2.2.4 Measurements on Linguistic Features of Task Performance	49
2.3 Task Planning and the Use of Communication Strategies	54
2.4 Summary and Conclusion of the Literature Review	60

Chapter 3 Validation in Language Testing 63

3.1 Messick's Framework of Validity	63
3.2 Kane's Argument-Based Approach to Validity	68
3.2.1 Interpretive Argument	69
3.2.2 Evaluating the Interpretive Argument	70
3.2.3 Advantages of an Argument-Based Approach to Validation	73
3.3 Discourse Analysis and the Validation of Second Language Speaking Tests	75
3.3.1 Investigating Participant Behavior in Oral Interviews	77
3.3.2 Comparing Oral Interview Behavior with Natural Conversation	80
3.3.3 Investigating the Effect of Test Format on Oral Test Performance	81

3.3.4 Planning Condition and Produced Monologic Discourse	82
3.3.4.1 Effect of Planning Condition on Fluency	82
3.3.4.2 Effect of Planning Condition on Complexity	83
3.3.4.3 Effect of Planning Condition on Accuracy	83
3.4 Assumptions for Argument-Based Validation in the Study	85

Chapter 4 Methodology **89**

4.1 Data Collection	89
4.1.1 Design	89
4.1.2 Test Takers	90
4.1.3 Tasks	94
4.1.4 Questionnaires	96
4.1.5 Procedures	97
4.2 Scoring the Monologic Oral Production	98
4.2.1 Rating Scales	98
4.2.2 Rating	100
4.3 Transcription of the Recorded Speech Samples	100
4.4 Coding Linguistic Features	102
4.5 Identifying and Coding the Use of Communication Strategies	109
4.5.1 Reduction Strategies	110
4.5.2 Achievement Strategies	111
4.5.3 Other Strategies	114
4.5.4 Inter-Rater Reliability for Coding Communication Strategies	115
4.6 Conclusion	117

Chapter 5 Research Results **119**

5.1 Test Score Analysis	119
5.1.1 Inter-Rater Reliability	119
5.1.2 Independence of Sub-scales	121
5.1.3 Research Question 1	122
5.2 Analysis of Linguistic Features	128
5.2.1 Quantity of Language Produced	129
5.2.2 Factor Analysis for the Linguistic Measures Used in the Present Study	130
5.2.3 Research Questions 2, 4a and 4b	132
5.2.3.1 Fluency Measures	133
5.2.3.2 Accuracy Measures	135
5.2.3.3 Complexity Measures	137
5.2.3.4 Research Question 4a	139
5.2.4 Summary of Main Results for Linguistic Features	142
5.3 Use of Communication Strategies	143
5.3.1 Research Question 3a	144
5.3.2 Research Question 3b	144
5.3.2.1 The Relationship between Communication Strategies and Accuracy Measures	146
5.3.2.2 The Relationship between Communication Strategies and Fluency Measures	148
5.3.2.3 The Relationship between Communication Strategies and Complexity Measures	149
5.3.3 Research Questions 4c and 4d	151