



华夏英才基金学术文库

曾用强 著

计算机辅助英语口语考试研究

——基于广东省机助高考英语口语考试改革的研究



科学出版社

 華夏英才基金學術文庫

计算机辅助英语 口语考试研究

——基于广东省机助高考英语
口语考试改革的研究

曾用强 著

科学出版社

北 京

内 容 简 介

本书介绍了英语口语考试的发展历史,重点讨论了计算机化英语口语考试的设计、效度验证、评分模型以及考务工作。全书分五个章节,每个章节由两部分组成:关于计算机化口语考试的理论和关于广东省计算机化高考英语口语考试的实证。五个章节的内容主要涉及英语口语考试的现状与发展、计算机辅助口语考试的开发、计算机辅助口语考试的信效度、英语口语考试的评分和计算机辅助考试考务工作。

本书可供从事语言教学和语言测试研究的英语教师、研究人员和研究生阅读和参考。

图书在版编目(CIP)数据

计算机辅助英语口语考试研究:基于广东省机助高考英语口语考试改革的研究/曾用强著. —北京:科学出版社,2011

(华夏英才基金学术文库)

ISBN 978-7-03-031113-9

I. ①计… II. ①曾… III. ①计算机应用-英语-口语-高等学校-入学考试-研究-广东省 IV. ①H319.9

中国版本图书馆 CIP 数据核字(2011)第 091082 号

责任编辑:卢秀娟 / 责任校对:朱光兰

责任印制:钱玉芬 / 封面设计:陈 静

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

骏 立 印 刷 厂 印刷

科学出版社发行 各地新华书店经销

*

2011 年 6 月第 一 版 开本:A5(890×1240)

2011 年 6 月第一次印刷 印张:4 3/4

印数:1—2 000 字数:144 000

定价:25.00 元

(如有印装质量问题,我社负责调换)

目 录

第 1 章 英语口语考试的现状与发展	1
1.1 口语考试的基本模型	1
1.2 常用的口语考试形式	2
1.3 口语考试的发展趋势	9
附 1 广东省普通高考英语口语考试的发展	12
第 2 章 计算机辅助口语考试的开发	18
2.1 考试开发的基本模型	18
2.2 计算机化考试的开发步骤	20
2.3 设计阶段的基本问题	22
2.4 开发阶段的基本问题	34
2.5 实施阶段的基本问题	37
附 2 广东省高考英语计算机辅助口语考试的设计	39
第 3 章 计算机辅助口语考试的信效度	48
3.1 信度	48
3.2 效度	52
3.3 信度和效度的关系	58
3.4 影响计算机化考试的因素	60
3.5 计算机化考试的效度验证	64
附 3 广东省计算机辅助高考英语口语考试(COET)的效度 验证	74
第 4 章 英语口语考试的评分	90
4.1 口语考试的评分过程	90
4.2 口语考试的评分模式	91
4.3 口语考试的评分误差	93
4.4 口语考试评分误差的控制	100
附 4 基于网络评分的误差控制模式	106
附 5 口语考试中不同评分模式的差异性研究	113

第5章 计算机辅助考试考务工作	123
5.1 考试场次的安排	123
5.2 考试的报名	126
5.3 考试场地的要求	127
5.4 考试的安全性	129
5.5 考试过程的控制	131
5.6 评卷与分数发布	132
附6 广东省普通高考英语口语计算机辅助考试考务工作实 施办法	133
附7 广东省普通高考英语口语考试统一指令	138
参考文献.....	141

第 1 章 英语口语考试的现状与发展

1.1 口语考试的基本模型

所谓口语考试,就是向考生提供一系列任务,要求考生按照任务的要求完成口头的语言产出(即,考试行为),然后考官(或评分员)按照特定的评分标准评定考生语言产出的等级或得分。所以,口语考试的实施过程有两个主要阶段:①任务激活考生的语言产出(语言能力的表现),即考试过程;②评分员对考生语言产出的评定(语言能力的表示),即评分过程。

口语考试包括任务、考生和考官三个主要组成部分,三者之间存在互动的关系:考生和考官互动完成口试任务。口试任务规定了考生语言产出的主题和内容;考官的主要任务则是引导考生按照任务要求产出相应的语言,并对考生的表现水平做出评判。所以,在口语考试过程中,任务是制约和影响考生语言能力发挥的主要因素之一,其次是考官因素。

口语考试任务包括以下几个方面的信息:①语言输入;②语言产出的目标要求;③任务的参与者;④任务的完成方式;⑤评分模式等。表 1.1 归纳了设计口语考试任务时所需考虑的主要因素。

表 1.1 口语考试的任务设计

参 数	表现形式
语言输入	文本、图片、录音、视频
语言产出的目标要求	自由式产出、半自由式产出、限制性产出
任务的参与者	单考生、双考生、多考生
任务的完成方式	独立完成、交互式完成(与考官交互、考生之间交互)
任务的完成条件	预先准备、即时
语言产出的记录方式	实时、录音、数字录音、录像
评分模式	根据评分时间、评分标准、评分分数、评分人数以及评分方式可进一步细分

口语考试的任务设计首先要设计一个语言交际情景,然后组合表 1.1 中各个参数的不同表现形式。口语考试的情景要真实,贴近考生的生活、学习或工作场景。同一种情景可以与不同的参数值组合,而设计成不同的口语考试任务。比如,以下的口语任务都是根据情景“你要陪一位美国客人参观校园”而设计的:

口试任务设计 1:

观看介绍校园的视频(或阅读相关的文字材料,或校园平面图),准备 1 分钟后,向主考(扮演美国客人)介绍校园。

注:“考生向主考介绍校园”可以设计成自由介绍,或按提限定性地介绍等不同的形式。

口试任务设计 2:

(1) 5 分钟的时间阅读一份校园地图和部分建筑的中文介绍;

(2) 向主考(扮演美国客人)介绍校园(限时 1 分钟);

(3) 根据地图和中文介绍的信息,用英语回答主考的提问。

口试任务设计 3:

三个考生扮演美国客人,先观看介绍校园的视频(大约 5 分钟),然后互相谈论校园(包括就视频的内容提问、发表评论,或回答其他考生的提问等)。

考官观察考生的表现行为,并根据他们的口语产出的数量和质量进行评分。

1.2 常用的口语考试形式

在前一节,我们简述了口语考试任务设计的基本模型,下面我们将介绍国内外目前流行的考官与考生面对面的口语考试形式。按照考官在口语考试过程中的作用,这些考试可以归纳为三大类:单向、双向和互动的口语考试。

1.2.1 单向的口语考试形式

单向的口语考试形式是指,考生按照考试要求直接向考官作口头表达,主考即时评分。这种形式包括:朗读、口头报告、看图说话、复述

故事、讨论交流等。

1. 朗读

考生随机抽一篇短文,准备1分钟后(或即时)开始朗读。主考根据考生的语音、语调和语速对考生的朗读进行评分。

变换形式: 模仿朗读

考生先听一段录音或观看一段视频,准备1分钟后(或即时)开始模仿朗读。主考根据考生的语音、语调和语速的模仿近似度进行评分。

2. 口头报告

考生从所提供的若干主题中,(随机)选择一个主题,然后就这个主题作2~5分钟的口头报告。考官根据口头报告的思想内容、语言表达、流利程度以及语音语调等进行评分。

口头报告的主题必须具体,确保考生能够就此发表观点和看法。比如:“高速公路是否限速”可能比“如何改进高速公路的交通安全”更适合作为口头报告的主题。

变换形式: 提供指令/描述/解释

考生作简短的准备后开始描述一个常见的物体或日常生活中的某个过程等。比如:

- How do you make a good cup of tea or coffee?
- Describe a bicycle.
- Describe how to prepare your birthday party.
- Give instructions for using a public pay-phone.
- Explain how you would advise someone to look for a job.
- Describe how people in your hometown celebrate the Spring Festival.

3. 看图说话

向考生提供若干幅相近的图片,要求他/她根据自己的水平和爱好从中任意选择一幅进行描述。考官的评分是基于考生所提供信息的完整性,以及语言表达的准确性和流利度。看图说话的常见的变换形式

有 4 种。

变换形式 1: 使用几幅类似的图画

向考生提供几幅类似的图画,这些图画之间只是存在细微的差别,要求考生描述其中的一幅画,主考猜测他/她所描述的图画,并评判描述的准确度。

变换形式 2: 连画成故事

首先向考生提供几幅分散的图画或照片,要求考生对它们进行排序,然后编一个故事。这些图画或照片允许有多种不同的排序方法,目的是让考生更好地发挥想象力和创造力。

变换形式 3: 看视频续编故事

考生观看一段视频后,根据已有的故事情节续编故事。主考根据考生的内容、语言、流利和语音等进行评分。

变换形式 4: 看视频重构故事

考生观看一段无声的视频后,准备 1~2 分钟,根据视频的画面信息重构一个故事。主考根据考生的内容、语言、语音和流利度等进行评分。

4. 复述故事

考生首先听一段录音,或阅读一篇短文,或观看一段故事视频,然后就所听、所读或所看的内容,复述或概括其中要点。主考根据复述或概括的内容、语言、语音及流利度进行评分。

5. 讨论交流

两个或更多的考生一起参加考试。考生完全自己组织控制讨论。考试任务通常包括从书面文本获取信息,并通过讨论对特定的问题达成某种决定或观点。当需要应用多方面的文本或信息来源时,可能要求考生在讨论开始之前先阅读这些材料,甚至是在考试的前期阶段。讨论的一个重要特征就是没有唯一的答案,否则一旦回答了问题之后考生就只能坐等下一个问题的回答,而不是连续地交谈。考生预先应该知道,评分是基于他/她的表达方式、观点的论证以及对其他人的评价等,而不只是他们所表达的事实性信息。评分可以是即时的,也可以

是录音之后的评分。以下是讨论交流的几个话题示例:

- 为具有特定兴趣爱好的学习者选择一门培训课程;
- 根据宣传册为某个家庭或个人设计一个假期;
- 从某份职务的应聘者中选择最佳的候选人;
- 讨论某家(真实或虚构的)公司开发新产品或进军新领域的论证报告;

- 为城市交通“治堵”出谋献策等。

变换形式 1: 根据小型讲座的录音(或视频)展开讨论

4~8 人组成一组,一起听(看)5~10 分钟的录音(视频),内容主要涉及一般性题材,比如:电台新闻报道等。考生可以做笔记,但是不允许照读。他们首先从录音(视频)中获取某观点和看法,然后就此展开讨论。

变换形式 2: 就个人私物展开讨论

向一组考生提供某私人的一些小物品,比如:票证、证件、照片、信件、发票、通讯簿、日记等。他们的任务就是通过讨论这些小物品,形成对物主的认识,比如:他/她的年龄、职业、社会地位、兴趣爱好、生活方式等。

这种形式特别适用于测验考生假设、推测和谈判的语言使用能力,尤其当考生了解这种讨论并没有所谓的“正确”答案,他们会感到放松、有自信发挥自己的想象力,来完成这项任务。

1.2.2 双向的口语考试形式

双向的口语考试形式是指,考生按照主考的提示提供口头表达。主考的提示是预先设计好的,比如:提问的问题等。这种口语考试的组织形式主要包括:填写表格、情景反应、问题回答、补全对话、句子复读、句子改错等。

1. 填写表格

考生和主考共同完成表格的填写。主考提问,考生回答问题,并填写表格。问题通常涉及考生的个人信息,职业环境以及语言需求等。

表格或由主考,或由考生填写。另一形式是,考生在口试开始之前先花2~10分钟填写表格,在口试时主考将利用表格的信息提问考生。

变换形式: 使用问卷调查

填写表格只是提问考生一些事实性信息,比如:考生的个人信息等。但是问卷调查可以提问特定领域的问题,比如:个人的兴趣爱好、对社会热点问题的看法、工作(学习)计划等。

2. 情景反应

主考向考生提供一些日常生活中可能发生的情景。提供方式可以是书面的,也可以是口头的,或两者并用。考生设想自己处在各个不同的情景中,然后根据情景作出反应。情景设计得简单、易懂,所要求的语言不能复杂,一般只需两三句话就能回答,比如,咨询某信息、表示道歉、谢绝邀请等。

3. 问题回答

考生根据阅读短文的内容或客观事实,回答主考的提问。

变换形式 1: 使用确定的问句类型,比如:

- yes/no 问句
- 反意问句
- either/or 问句
- 简单的事实性问句
- 描述性问句
- 叙事性问句
- 思辨性问句
- 虚拟性问句
- 观点论证问句

变换形式 2: 暗示提问

提问与回答都是口语能力的基本技能。主考只是简单地说 Now I'd like you to ask me a few questions. Ask me about my name, job, family and interests.

4. 补全对话

向考生提供一段不完整的对话。他有几分钟的准备时间,然后补全对话。比如:

A: Can you tell me how to get to the station?

B:

A: And where does the bus stop?

B:

A: Oh yes, I can see it.

要求考生填入的单词可以是开放式的,即没有标准答案。比如:

A: What did you do at the weekend?

B:

A: What was the weather like?

B:

也可以是完全限定的。比如:

A:?

B: I'm thirty one.

5. 句子复读

考生先听一组句子,然后准确地复读这些句子。这些句子可以由考官朗读,也可以是预先录音的。句子长度可以从一个词(如:Hello!)到15个,甚至20个词。超出20个词的句子就少见。

这种考试方式简便,可以较好地地区分各种不同水平层次的考生,但是缺乏交际性和真实性。

评分的最简单方式就是,评判每个句子是否准确,然后累计准确数。这种评分模式不需要对评分员作专门的训练。较为复杂的评分需要考虑部分正确的复读,比如意义相同,但是用词不同。比如:要求复读的句子是 He said he was hoping to come this weekend, but as it turned out he wasn't able to,考生的复读是 He said he was hoping to come this weekend, but in fact he couldn't.

变换形式 1: 复读句子的长度由短到长递增

复读先从一两个词开始,然后考生每正确复读一(或两)句,句子的长度随之增加。最后的得分取决于考生所能准确复读句子的长度。

变换形式 2: 复读句子限定于特定的语言范畴

复读句子可以限定在特定的语言范畴,比如:

- 大纲规定的某一语法结构或功能意念的句子
- 特定范围的词汇
- 音似词
- 最容易受母语干扰的音
- 不同的语音语调等

评分时应特别注意观察这些特征是否得到准确的复读。

6. 句子改错

考生先听或读一个有语病的句子,然后要求找出错误,并改正错误。错误可以多类型、难度层次也不同,如:一般的语法错误,或更难的习语和搭配错误等。

这种形式也可能要求考生改正一篇短文中的若干错误,而不仅仅是一个句子。短文中的每个句子不一定都包含错误。错误也可能允许有多种改正方法。

1.2.3 互动的口语考试形式

互动的口语考试形式是指,主考与考生互动,共同完成特定的语言交际任务。这类口试形式需要情景设计,比如:角色扮演等。

角色扮演(role-plays)与其他口试形式有很多相似之处,考生扮演的角色是预先设计好的,而不能自由扮演自己的角色。考试设计者构思一个情景,定义好应该表达的思想内容,考生的任务实际上就是按照这样的设计用自己的语言填补必要的信息。

考试前几分钟考生获知自己要扮演的角色,以及所要完成的交际任务。比如:以下几个例子解释了角色扮演的角色设计:

- 一位男生在一次晚会上遇见一位女生
- 律师和被告
- 警察和报警人

- 商场经理和一位投诉的顾客
- 酒店服务员和一名旅游者
- 医生和病人

设计角色扮演的口试任务时,需要注意的是,应根据考生的特征设计更具体化的情景。为了平衡每个角色的难度及产出的语言,有时有必要对口试任务作更具体化的说明,比如:“至少提问三个问题”或“概括你的内容,确认你已经理解对方的意图”等。

变换形式:使用电话

电话是角色扮演的十分真实的工具。它特别适合于考查考生语用能力,比如:如何表达确认、要求重复、拼写姓名或地名等。利用电话设计角色扮演比较真实,容易激活考生的语言产出,考生扮演打电话者和听电话者。

1.3 口语考试的发展趋势

任何考试的发展都取决于两个领域的发展,其一是涉及考试内容方面的领域,如:语言教学理论以及语言测试理论的发展;其二是涉及考试手段方面的领域,如:科学技术等领域的发展。下面重点讨论语言教学理论、语言测试理论和科学技术的发展对口语考试发展可能产生的影响。

1.3.1 语言教学理论

语言是一套知识,这就是20世纪40年代以前外语教学和测试体系的语言观的内涵。到了40年代,人们开始感觉到,关于语言规则的知识不等于语言本身,于是有人提出,教语言,教的应该是“语言”,不应该是“关于语言”(Moulton 1961)。以美国语言学家布龙姆菲尔德(Bloomfield 1933)等为代表的结构主义语言学派,第一次对语言的形式系统作了客观的、科学的分析。主要观点可以归纳为:语言是一套形式结构,一套符号系统。同时,他们受心理学的行为主义影响,他们认为学语言就是要获得操作这套符号系统的技能,用行为主义心理学的

术语说,就是训练对刺激作出正确反应的一套语言习惯。结构主义语言学和行为主义心理学的结合,产生了结构主义教学道路。同时,在测试方法上,形成了心理计量结构主义测试。教学语言观的内涵从知识转变为技能。但是这一流派受到了来自生成语法学家乔姆斯基(Chomsky, 1957, 1959)的强烈抨击。他从心理学和语言学两方面驳斥了结构主义,提出了能力(competence)的概念。能力和技能是截然不同的概念。一种新的语言教学体系也随之诞生——交际语言教学。这个体系认为,学语言不仅仅是学语音、词汇和语法知识,也不仅仅是训练操作形式符号的技能,而是获取人与人交际的一种能力。能力不排除知识和技能,但是也不等于知识加技能,而是把知识和技能包容进去的、性质起了变化的综合体。交际语言教学不停留于教知识、教技能,而是教能力。同时,在测试领域,相对应的外语测试(交际语言测试,communicative language testing)提出语言测试所测试的,也应该是能力。

从语言教学理论的发展观来分析,口语考试必然成为未来语言测试中的一个重要组成部分。

1.3.2 语言测试理论

语言测试理论论述语言测试使用 and 开发必须考虑的问题,以及如何控制语言测试的质量问题,比如:考试的效度和信度等问题。语言测试理论随语言教学理论的发展而发展,语言测试理论的发展主要体现在测试内容和测试方法两大方面。在测试内容方面,语言测试经历了从重语言形式到语言意义的转移,也就是,现代的考试以学生在实际的语言交际情景中应用语言知识和技能的能力为主要测试内容;在测试方法方面,语言测试开始从传统的标准化测试(standardized tests)向行为测试(performance tests)转移,在试题设计时,真实性(authenticity)成为一个重要的标准,也就是,考试任务越来越接近实际情景中的语言使用,此外,传统的考试往往把语言分解成知识和技能进行分项测试,现代的考试开始按照交际任务设计综合考试任务,要求考生必须综合应用听、说、读、写的技能来完成特定的一项任务。

1.3.3 科学技术的发展

科学技术的发展为测试走向现代化提供了物质基础,尤其是多媒体、计算机和网络技术的迅速发展,使计算机化或基于网络的考试进入一个实际应用的阶段。近些年来,许多考试开始实现计算机化考试。与传统的纸笔考试相比,计算机化考试具有自身明显的特征和优势。

计算机化考试的最基本特征就是,它从相关专业学科(如:语言学和语言教学)获得考试的内容,从现代科学技术(如:计算机和网络技术)获得考试的形式。计算机化考试是专业学科内容和科学技术手段有机结合的产品。所以,它具有一些传统纸笔考试所没有的特征和优势。

具体地讲,与传统的纸笔考试相比,计算机化考试的特征主要体现在以下几个方面:

(1) “持续性”的考试模式:传统的纸笔考试只能在固定的时间进行,考生只能被动接受考试的时间和地点;计算机化考试可以随时组织,考生可以自己安排考试时间。

(2) 适宜性:传统的纸笔考试是预先命制好试题,考生必须回答全部的试题;计算机化考试可以让考生只回答与其能力相适应的测试试题。

(3) 提高考试的效度:借助多媒体或其他科学技术,计算机化口语考试可以开发一些创新性的题型,提高试题的真实性,使考试任务更接近真实的语言使用任务,从而提高考试的效度。

(4) 提高考试的信度:面对面口语考试由于采用人工现场评分,误差较大。计算机化口语考试通过存储考生的考试数据,考后专门组织专家评分,在很大程度上保证了评分标准掌握的一致性,从而保证考试的信度。

口语考试的具体任务设计都是按照考试的目标要求和用途,科学合理地组合以上的具体表现形式。一次口语考试可能包括一个或多个这样设计的任务。

随着测试理论和语言学习观的发展,以及现代科技的不断进步,现代口语考试任务的设计越来越强调考试任务的真实性,即最大程度上

真实体现现实语言使用的情景。根据图 1.1 所归纳的主要参数,我们可以清楚地意识到,现代口语考试任务的设计开始呈现以下的特征:

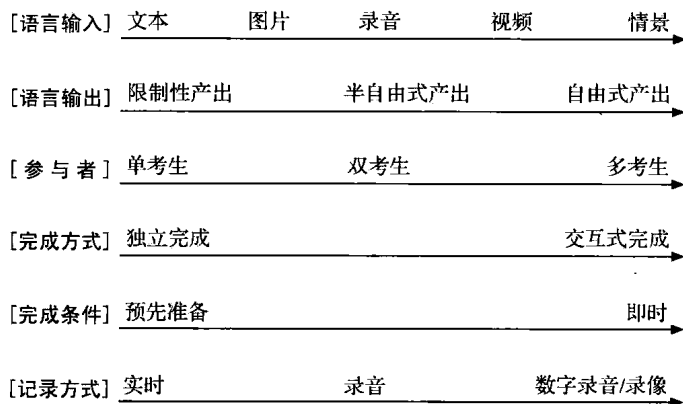


图 1.1 口语考试的发展趋势

通过图 1.1,我们可以预见,未来的口语考试将特别强调语言交际的实际情景,要求在实际的语言交际情景中考生可以自如地与其他参与者进行语言交流,现代科学技术(比如:计算机、网络和多媒体等)开始在口语考试中扮演越来越重要的角色。

附 1 广东省普通高考英语口语考试的发展

1. 广东省英语高考历史简要回顾

自 1977 年我国恢复高考制度以来,国家教委(现国家教育部)为了顺应国际英语测试的改革方向,经过多年的试点和研究,于 1985 年开始在广东省试行高考英语标准化考试——MET(matriculation English test)。经过 4 年的试点后,MET 于 1989 年推广到全国。1991 年国家教委在总结多年 MET 试行的基础上,推出了新的高考英语模式——NMET(national matriculation English test)。到 1995 年全国除上海外全部参加了 NMET 高考。1996 年广东省率先在高考英语中增加听力测试,但成绩不计入总分。自 1997 年起,广东省采用了全国