

# 关于口语考试和评判的思考

■ 乔辉 刘庆思

**摘要** 本文通过介绍和总结外语口语考试的题型,分析了口语评判时应该注意的问题,在对口语进行评判时,除选用合适的评分标准进行评判以外,要综合考虑包括口试题型、评分信度、口试效度以及后效作用等各种因素的影响。

**关键词** 口语评判;评分标准;评分信度;效度

【中图分类号】G405

【文献标识码】A

【文章编号】1005-8427(2012)06-0020-4

在语音测试领域,关于口语测试的研究和实践是测试领域内一个较为年轻的分支。随着中国基础教育的进一步发展,尤其是2001年教育部关于英语课程标准的推行,英语口语教学和测试都面临新的挑战。正因为口语有其独特性,口语考试及其评判就显得更加复杂和多变。本文将对口语的特性、口试的题型、口试的评分标准以及评判时应该注意的问题进行探讨。

## 1 口语的特性

口试的评判对象是考生的口语。作为一种独特的评判对象,口语有其独特性。人们在说话时会自主地或非自主地用自己的言语向听者塑造自己的形象。不同的说话速度、停顿、音调、音高的变化

以及语调都是言语的一个很重要的特点。人们的言语或发音是有含义的,这也是口语考试的主要目的所在。(Luoma, 2005:10)。

口语包括很多说话时的特征,比如单音的发音、音高、音量、语速、停顿、重音和语调等。所有这些都是口语评判时应该考虑的因素。这些因素能否都放在同一个标准中进行评判以及应该更看重考生的发音还是更看重考生的表述能力等问题一直是研究的热点。

在外语学习中存在一种标准发音的说法往往受到质疑,一方面,是因为对于一门比较国际化的语言来说,很难定义哪一种发音是标准的、地道的。每一门语言都有其因地域不同而产生的差异性。有时甚至不同的目的也会造成一门语言发音

的变化,而这也给口语评判造成了很大的困难;另一方面,对外语学习者的学习过程研究发现,无论一个外语学习者如何努力,只有为数极少的人的发音能够达到说本族语的人的水平,但他们的言语作为交际来说却已远远足够。

## 2 考查口语的题型

一项初级层次外语口语考试经常包括朗读、跟读等题型,可以看出这些考试都注重对发音的考查。发音的准确度是口语考试中一个很重要的评判因素,但整个言语的可理解程度(包括语速、语调、重音和节奏)有时候显得更为重要。

随着测试研究者对口语考试的深入研究,考查口语的题型也越来越多样化。一般来说,经常在口试中见到的题型包括:朗读、重复句子、就一定的情景提问和回答问题、信息交换、连续性表达、角色扮演以及控制性访谈等(Fulcher, 2003; Luoma, 2004; 武尊民, 2002)。不同的题型在不同的口语考试中扮演着不同的角色,对保证整套试卷的信度和效度起到不同的作用。根据 Luoma(2004:48-51)的研究,可以将口试题型划分成两大类:开放式(open-ended)题型和封闭式(structured)题型。

### 2.1 开放式题型

开放式题型的目的是要求考生能灵活使用所掌握的语言来完成一项任务,如演讲和角色扮演等。开放式题型能让考生综合表现出自己的能力,且能准确地评判出考生的口语能力,但这类试题往往耗时较长,在评判时由于完全是人工评分,评分信度难以得到保证,受评分者的主观性印象给分的影响较大。

### 2.2 封闭式题型

封闭式题型中,考生的回答相对较短,朗读题、简答题等可以归为这一类。封闭式试题由于题型的限制,考查的能力也相对比较单一,单凭一两道试题很难看出考生的口语能力,在一份口语试卷中

需要综合使用。封闭式试题在评判时比开放式题型更为客观,受人为因素影响较小。对于考查考生的语音语调和语法知识非常有效。

不同的考试会将不同的题型进行组合形成一套合理的口试试卷。交际语言能力强调交谈是发生在特定的社会交往环境中,交谈双方有交际的目的和需求,这反映在口语试题上就是试题任务应与具体的交际情景紧密联系在一起,考生之间因为有交谈的需求,即信息沟,才会发生交谈(Fulcher, 2003: 51)。

不同的题型在不同的口语考试中扮演着不同的角色。就目前来说,高考中经常采用的题型包括朗读、就一定情景提问和回答问题以及连续性表达等;而大学英语四、六级考试等考查相对较高层次口语能力的考试则包含控制性访谈和讨论等题型。

## 3 口语评判应该注意的问题

口语考试的效度较高,但口语考试的信度比较低,而且组织十分困难。口语考试采用主观性试题,很难保证评分的客观性和一致性;口语考试的规模也在很大程度上受到面试形式的制约。如同其他的考试成绩一样,口试成绩必须是可信的,公平的,且对于其使用者来说是有用的,因此口语评判的评分标准、信度和效度显得尤为重要。

### 3.1 评分标准

在评分过程中,评分标准的要求应该与教学要求相吻合,应符合语言测试理论的基本原则。既要考虑到整个考试的信度、效度,又要考虑到考试的可操作性及后效作用(刘庆思, 2009)。口语测试主要是主观性测试,因此,如果几个评分员对于评分标准的掌握有差异,或者持续长时间的工作使某位评分员评分不稳定,都会给考试的信度带来影响。评分标准是衡量考生考试表现的尺度。评分标准的文字说明应该跟撰写公文一样,在不需要参照任何其他的标准的情况下,做到简洁、清楚、明确、易

懂(Council of Europe, 2001)。

目前国内外较有影响力的大型考试口试部分采用的评分方法大致可分为分项评分和综合评分。文秋芳(1999)指出分项评分侧重语言结构和语法规则的正确性,而综合评分强调语言的交际功能和交际效果。分项评分的典范是美国的托福(TOEFL)考试口试部分,它所采用的评分标准包括语音、语法、流利度等,每一项分几个等级。分项评分的信度较高,能向考生提供更加细化的反馈信息,但其同时忽略了考生的总体表现。综合评分的代表是英国的雅思(IELTS)考试口试部分。雅思口试评分标准分为9级,对各等级水平的交际总体特征进行描述。综合评分省时、快速,但其评分很大程度上依赖评分员的经验和专业素养。我国全国英语等级考试(PETS)1~5级的口试采用分项评分和综合评分相结合的评分方式,采用两位口试教师面对两名考生,由两位口试教师对考生的口语能力进行评价的形式。主持口试的教师给出一个综合分,不参与交谈的教师根据分项评分标准给出分项分(教育部考试中心,2010a;教育部考试中心,2010b)。PETS所采用的综合评分与分项评分相结合的方式可以在一定程度上避免口试成绩受口试教师主观性的影响,在一定程度上保证了口试评分的信度。

### 3.2 评分信度

对于口语评判来说,很重要的两个技术指标就是信度和效度(Luoma, 2004)。

信度通常被定义为分数的一致性,指的是一项考试的分数是否可信。如果同一考生在相近的两个时间段内参加相同的考试,他应该获得相同的分数。Bachman和Palmer(1996)归纳了导致评分信度降低的三个因素:评分人员对评分标准的理解不同;评分人员严格程度不同;与评分不相关因素的干扰,如考生外形、口音、在一些问题上的立场等。口语考试的信度需要高质量的评判工具和评判过

程来保证。考试组织者往往使用以下几种方式来保证考试的信度。

#### 3.2.1 培训口试教师

准备参加口试评判的教师在正式考试前会被考试组织者集中起来进行统一的培训。培训的主要内容为考试的要求和具体的评分标准。其目的是统一评分标准,使得所有参加过培训并获得资格的教师能够按照所规定的评分标准公平地对考生的口语进行评判。培训材料和使用的样卷由考试组织者事先聘请相关专家给出分数并写出具体的评注。经过培训的教师会参与一些口试样卷的评判,并根据打分表现决定其是否有能力进行正式阅卷。

#### 3.2.2 标准设定

标准设定是根据评分标准来确定及格分数和其他分数档次。标准设定经常采用两种方式;一种是根据同一档次的成绩较好和成绩较差的两组考生的表现来确定及格分数;另一种是通过让学科专家描述不同水平的考生在各个档次上的具体表现来确定及格分数。

#### 3.2.3 评分监控

通过监控口试阅卷过程来保证口试评分的信度主要涉及两个指标:评分员内部一致性(Intra-rater Consistency),即同一评分员在较长一段时间内对不同考生评判时对标准把握的一致性情况;评分员间的一致性(Inter-rater Consistency),即不同评分员评判不同的考生表现时的评分一致性情况。这两个指标的高低都会直接影响到考试的信度。因此,保证评分信度的一项重要内容是以各种方法将这两个指标控制在可接受的范围之内。

### 3.3 口试的效度

效度是衡量一项考试是否充分考查到应考查内容的指标。在口语评判中,对评分标准的评估是一项考试效度研究的重要组成部分(Luoma, 2004)。首先,评分标准必须与考试的目的、考试的能力构想以及试题任务相一致。Fulcher(2003)强

调应将能力构想作为制定评分标准的核心。其次,评分标准应该意义明确、通俗易懂。评分标准是否经过不断的修改和完善,评分尺度是否经过专家不断地讨论和修订都是效度检验的重要参考资料。

在口语考试的效度检验过程中,考试的实施过程和评分过程也是不可或缺的一部分。这主要表现在考试实施和评分与考试内容架构的一致性和连贯性上。这方面的效验数据包括评分员的评分报告、评分教师培训资料、评分教师的资格认证等资料。

#### 4 总结

口语考试及其评判是一项很复杂的工作,需要考虑多方面的因素。如何针对不同的考试目的选择合适的口试题型来组卷、采用哪种评分标准、如何保证评分信度等一系列的问题都需要做出综合性的全面考虑。

口语评判是英语教学和测试中的一项日常活动,既需要评判人员准确地掌握相关的评分标准,更要求评判人员有足够的责任心和耐心。口语评判方

式方法的正确选定一方面有利于提高考试的质量,另一方面也会对英语教学产生良好的反拨作用。

#### 参考文献

- [1] Luoma, Sari. Assessing Speaking [M]. Cambridge: Cambridge University Press. 2004.
- [2] Fulcher, G. Testing Second Language Speaking [M]. London: Pearson Education Limited. 2003.
- [3] 武尊民. 英语测试的理论与实践 [M]. 北京: 外语教学与研究出版社. 2002.
- [4] 刘庆思. 关于中学生英语作文批判方法的思考. [J] 中小学外语教学(中学篇), 2009(3): 1-4.
- [5] Council of Europe. Common European Framework of Reference for languages: Learning, Teaching, Assessment [M]. Cambridge University Press. 2001.
- [6] 文秋芳. 英语口语测试与教学 [M]. 上海: 上海外语教育出版社. 1999.
- [7] 教育部. 全日制义务教育普通高级中学英语课程标准(实验稿) [M]. 北京: 北京师范大学出版社. 2001.
- [8] 教育部考试中心. 全国英语等级考试第一级考试大纲 [M]. 北京: 高等教育出版社. 2006.
- [9] Bachman & Palmer. language Testing in Practice [M]. Oxford: Oxford University Press. 1996.

作者单位 教育部考试中心

## Reflections on the Testing and Marking of Language Speaking

QIAO Hui and LIU Qingsi

**Abstract:** Starting from the characteristics of language speaking, the paper introduces different test types of foreign language speaking tests and analyzes the issues concerning speaking tests marking: marking schemes, reliability and validity. It is pointed in the end that the marking of speaking tests is a complicated process which should take many factors into consideration, and which needs scorers accurately understanding the marking schemes, their responsibility and patience.

**Keywords:** Marking Speaking Tests; Marking Scheme; Reliability; Validity