

## 现行英语口语测试优劣分析

张 涛

摘要: 对 OPI ,SOPI , COPI 三种目前流行的英语口语测试进行多角度横向比较,对大规模口语测试中采用 COPI 方式提出看法,并从口语测试的本质以及 COPI 的效度方面加以阐释。

关键词: 英语口语测试; COPI ;交际能力;效度

Abstract :Horizontal comparison is made from multi - angle for three current popular oral English tests which are OPI , SOPI , COPI . The paper presents opinions on using COPI in large scale oral test . It also expounds the nature of oral English test and the validity of COPI .

Key words :oral English test ; COPI ; Communicative Competence ; validity

语言测试是用来衡量一个人掌握一种语言水平的方式,在语言教与学的过程中扮演着重要的角色。无论教师采用何种教学模式,都要对学生获得的知识及其学习进程进行评价。口语测试是语言测试的一个组成部分,它与常见的阅读、语法和词汇的考试不同,具有较强的主观性。而如果测试的主观权重过高,就会失去测试结果的公正性与真实性。因此,余姚有一种标准化的考试系统来定义并实施口语测试,使得口语测试走向更为科学、健康的发展道路,同时为指导口语教学以及进行人才选拔提供参考。OPI (口语能力面试, oral proficiency interview) ,SOPI (模拟口语能力面试,simulated oral proficiency interview)和 COPI (计算机辅助口语能力面试,computerized oral proficiency instrument)三种口语测试方式,各自有其优势与不足,本文将着重加以分析比较。

### 一、 OPI ,SOPI 和 COPI 比较

目前,根据口语测试的组织方式,通常可分为直接测试,半直接测试和间接测试三种。直接测试就是考官与考生面对面交谈,由考官为考生的口语表现进行评分(如 OPI) 。半直接测试中,考官和考生面对面进行交流,考试过程全部录音,考官当场不进行评分,而是在考试结束后,把磁带寄给考试中心统一评分(如 SOPI) 。间接测试则不需要考官与考生面对面交谈,通常由录音机或借助计算机录下考生的口语表现,测试结束后,由考试中心统一评分或是由计算机统一评分(如 COPI) 。OPI 是由 FSI (Foreign Service Institute)在上世纪 50 年代提出的一种口语测试方法。属于一种标准化语言测试程序。测试的方法是通过测试受试者的定性性强项与弱项来衡量受试者语言整体生成能力。同时,测出在语言功能上所维持的等级。OPI 测试反映了以下特点:第一,它是一种标准化测试,也就是说,教师事先将问题预设,从而保证受试公平性。

第二,根据话题的难度及其口语表现分成四级十等,施测者根据受试者的反应不断调整话题难度,控制整个面试过程,最终不断接近受试者的真实水平。第三,以受试者为中心,话题的提出,对话的展开都是围绕受试者进行的,所有的过程都是为了获取其真实的语言表达状况。第四,OPI 是常模参照测试,而非尺度比较测试,也就是说,考官给出的成绩是以单个考生为依据,不参考同组的其他考生的水平状况的。口试题目的题型有(1)朗读; (2)复述; (3)看图说话; (4)口头报告; (5)回答考官的问题等。OPI 测试同时存在着如下问题:首先,有的学者认为,在 OPI 测试中,面试( interview)与实际的口语交流(conversation)存在比较大的差异。主要表现为:在 conversation 中,对话双方的信息交流是双向的、均等的;而在 interview 中,考官在口语互动中拥有主控权,信息的流动依考官而定,这就增加了考生的压力,加大了考生的焦虑,影响考生口语水平的充分发挥。另外,直接口语测试具有较强的主观性,考生的口语水平会受到诸多因素的影响,如交际压力,认知复杂性和编码复杂性等。从测试成本角度考虑,在采用直接测试方式的 OPI 中,考官必须要经过培训,才能使得测试成功实施。而考官的培训通常要投入巨大的人力、物力,增加了考试的成本,而在大规模口试中,由于各地考官的口语水平参差不齐,很难有十足的把握断定某一成绩能真实地反映考生的语言能力,口试的信度无法保障。所以,从测试的效度和信度两个层面来评价,OPI 这种测试方法只适合小规模测试。1985 年,

美国应用语言学中心(CAL)开发了 SOPI ,旨在改善 OPI 测试的局限性,提高口试的信度并使得大规模口语考试成为可能。SOPI 是一种采用磁带(tape - mediated)的测试方式。在进行口语测试时,考官和考生面对面进行交流,考试过程全部录音。在 SOPI 中,主要的题型有提问和简单描述。一套完整的题目包括 15 个任务,共 45 分钟,其中考生作答时间约占 25 分钟。可以说,SOPI 在测试规模和评分体系方面相对于 OPI 都有了很大的改善。利用这一测试方式可以同时测试大批考生,考官不当场评分,而由考试中心根据考生的磁带来统一评分避免了评分的主观性,增加了口试的信度。但这种测试方法也存在一些问题。例如:在任何地方,任何层次的考生所接受的考试材料都是一模一样的,分毫不差。这种情况看似公平,但考生由于无法选择话题,如果对话题不熟悉,势必会造成考生的心理压力,影响其口语水平的正常发挥;其次,对于同一话题没有设计难度等级,考生必须在规定的时间内来作答,考生对话题难度及准备时间没有选择权;第三,磁带评价系统缺乏灵活性,有的学者认为让考生面对录音机进行考试的方式是冷漠的、没有人情味儿的。由于以上两种测试方式的局限性,以计算机技术为基础的、用于大规模测试的评价系统——计算机口语水平系统(COPI)应运而生,它是对以磁带为依托的 SOPI 的改进。首先,由于借助了计算机的技术优势,COPI 的任务库储存量大,话题广泛,同一任务设计了适合不同性别、年龄和职业的人,考生可以通过考试前输入的个人信息与题库中的任务进行匹配,在计算机的帮助下选出最适合自己的话题。其次,COPI 允许学生选择任务/话题的难度等级。计算机根据自我评价的结果,提供一份样题,考生可以选择同样题难度相等、高于或低于样题的正式试卷。第三,COPI 允许学生自己掌握准备时间和作答时间。在听到一个任务指令后,考生可以就某一任务/话题进行准备。当按下“我准备好开始讲话”按钮时,计算机就会开始录下应试者的口语表现,当考生结束某一任务/话题时,他只要按下“我已完成讲话”按钮即可。第四,COPI 采用计算机评分,减少人为因素所导致的评分不公正现象,比较客观。考生的口语水平等级由计算机根据考生所选择的话题难度等级及其口语表现而定,共分四级十等:D(优秀)优秀+,C(高级)高级+,高级和高级-,B(中级)中级+,中级和中级-,A(初级)初级+,初级和初级-。当然,COPI 测试中存在的缺乏人机的互动、系统中的任务是否具有真实性等不完善之处,还有待于进一步开发,最终使得技术能够更好地为人类服务。

## 二、大规模口语测试实施 COPI 的可行性分析

根据 COPI 的特点,并结合我国受试群体规模普遍较大的现状,一些学者提出采用计算机辅助口语测试。理由是:“随着参加口试人数的逐年增加,目前所采用的口语测试方法在测试内容、组织方式等方面的问题会日益突出。而机助测试可以利用计算机的技术优势,从其测试的界面设计、标识符号、语音识别、以及试题的字号、字体颜色等各方面进行人性化的设计,可以在某种程度上缓减考生的压力,有助于考生口语水平的充分发挥。更重要的是,机助口语测试可以满足不断增加的考生的人数,无需培训考官,任何考生可以在任何时间、任何地点,只要有 COPI 测试系统,他就可以完成口语测试。而且,计算机评分系统的客观性,可以减少人为因素所导致的评分不公正现象。”笔者认为 COPI 虽然能够解决上述问题但是仍不能作为大规模口语测试的手段。原因如下。

### 1. 口语测试的本质

语言测试是伴随语言教学应运而生的。西方语言学界的测试观大致经历了三个阶段。20 世纪 40 年代之前,语言学界缺乏理论基础,教学主要依靠教师内省,即凭借教师经验,把语言当作由语音、语法和词汇组成的知识系统去教。受这种观点制约,测试目标变成了对语言知识掌握多少的衡量。但实践证明,即使说话者掌握足够多的语言知识同样会造成交际失败。随后受结构主义和行为主义影响,语言被看作是由一整套符号组成的形式结构系统。这时,语言不再是知识,而是一种可以通过不断刺激反应而获得的技能。虽较第一阶段有所发展,但同样忽略了情景因素,而且不能立足于实践。直到 20 世纪 50 年代,由美国交际语法派语言

学家 Swain , Bachman 等人开始提出语言是一种交际能力,才使得语言教学得到了较为全面的发展。Bachman 认为语言是一种交际能力,由语法能力、 社会语言能力、 语篇能力和交际策略组成。它们缺一不可,共同决定交际者的语言表达能力。这时,它们已经把语言场景、 语用、 交际策略等非语言知识因素提高到一个新的高度。至此,语言真实性的问题开始受到广泛关注。口语测试的本质就是考查说话人在真实语境当中的语言表现能力。作为语言测试的一部分,在交际策略、 语用等方面显得更为突出。因此,要想真正地获得一个人的口语交际能力表现,就要涉及到上述非语言本体的能力考查,否则测试结果是没有可信度的。

## 2. COPI 的效度缺陷

信度和效度是衡量一个测试成功与否的重要指标。其中,效度又叫有效性,即判断某测验在何种程度上测出了它宣称要测的东西。根据 Bachman 提出的语言交际理论,COPI 存在以下自身缺陷。首先,口语测试的目的是测定学生的口语交际能力,测试内容以信息交流和意义表达为出发点,其原则是创造真实的交际情景和交际语境,保障口语测试的交际性,提高测试效度;制定合理的评分标准,采用科学的评分方法,挑选合格、 公正的考官,提高测试信度,保证口语测试的质量。而 COPI 依靠计算机进行评价,是“人机对话”而非“人际对话”。虽然能让更多的人得到口语测试的机会,并且解决了测试成本问题,但是得到的成绩与口语测试目标是南辕北辙的,无法达到测试目的,测试的效度也就大大降低了。其次,COPI 声称拥有广泛的话题储备,能考查被试对不同话题的表现,并且根据话题难度和被试表现进行评价,但是掌握虚拟的话题场景并不等于获得了处理实际交际场景的能力,很多非语言因素(如交际对象,交际场合)也会影响交际成败。因为没有完全相同的两次交际,再丰富的话题库也无法穷尽人类全部的交际场景,所以这种做法也是不可行的。最后,COPI 在设计中允许学生自己掌握准备时间和作答时间,一旦准备好,就开始录音,结束后对上述录音进行评价,这同样违背了语言的真实性特征。真实的口语交际是线性的,而且交流的双方在面对面的情况下,不可能一方提出话题,等待对方经过“准备”之后再重新交际。实际上,COPI 如此设计恰恰忽略考查受试处理话题的能力。在真实交际中经常出现偶发事件,如话题中断,转移或者回避,而如何处理这些情况正是口语交际能力的体现。

综上所述,以上三种测试在施测方式、 完成测试目标、 考查被试能力角度基本达到测试目的,可以说为实现口语测试标准化作出了贡献。但又同时存在者各自的局限性。我国目前的现状是,很多语言测试由于受到测试人数,成本的制约,至今仍不能引入口语测试,所以还处于起步阶段。引入计算机辅助口语测试固然有其积极意义,但一味满足口语测试规模需要,让更多的考生获得所谓的“口语测试成绩”,反而不能真正推动口语教与学的发展。

参考文献:

- [ 1 ] Kormos , Judit . Simulating Conversations in Oral - Prof iciency Assessment : A Conversation Analysis of Role Plays and Non - Scripted Interviews in Language Exams [ J ] . Language Testing , 1999 (16) :163 - 188.
- [ 2 ] Malabonga , Valerie. Dor ry M. Kenyon & Helen Carpenter . Self - assessment , Preparation and Response Time on a Computerized Oral Prof iciency Test [ J ] . Language Testing ,2005 (22) :59 - 92.
- [ 3 ] 张蓉.大规模英语口语测试的思考[J] .浙江万里学院学报,2008 , (5) .
- [ 4 ] 张凯.语言测验理论与实践[ M ] .北京:北京语言大学出版社,2002.