

计算机化大学英语水平测试口试题型设计研究

冯陵琳

(湖南大学外语与国际教育学院, 湖南省, 长沙市, 410000)

摘要: 本文在对常见口语测试题型研究的基础上, 根据测试目的和学生特点为湖南大学英语水平测试口语考试设计了四种口语题型的框架和内容。通过实证研究证明这四种题型适合口语水平测试并能较好的反映大学生口语水平。

关键词: 英语水平测试; 题型; 计算机化口语测试

中图分类号: H0

文献标识码: A

为同国际教育接轨, 湖南大学测试中心接受学校教务处任务主持开发“湖南大学计算机化大学生英语水平测试”。根据教学改革要求和大学生提高英语口语水平的迫切需求, 口语测试成为整个水平测试不可或缺的一部分, 因此如何公平公正地鉴定大学生口语水平被提上了议事日程。本研究以 Bachman 的交际语言能力理论为指导, 探索性地研究常见的口语测试题型, 在分析了水平考试的目的和本校学生特点的基础上设计口语考试的题型和命题规范。本研究除了对预测试成绩进行统计分析外, 还对考生进行了问卷调查, 问卷由 24 道选择题和 1 道开放题组成, 问卷内容主要包括考生对测试的态度, 考生的计算机熟悉度和他们自身的口语水平评价, 考生对考试题型以及题型组合的认可度等内容。分析结果显示大多数考生对本次考试题型基本认同, 同时也对考题提出了需要改进的地方。

一、口语测试题型研究

国内外学者对英语口语做过不少研究但较多涉及口语考试的信度效度研究, 测试方法和评分方法研究, 专门关于口试题型的研究非常有限。很多学者一致认为图片能够自然的激发学生灵感和想象力, 创造合理的情景引导学生产生有意义的口语。其中 HSU(2009)发现了一幅图片的看图说话题型的缺陷, 他认为多张同一主题的图片集合更能引导学生说出更长的更复杂的话语。同时 Jungok Bae & Yae-Sheik Lee(2010)指出了在运用图片刺激学生进行口语表达之前, 必须确保不同图片的难度值相等。他们研究最重要的贡献之一就是提出了一系列如何确保不同的图片具有等值的难度的方法。Carswel et al.(1998)有一项重要发现即图表中变化趋势多少可以预测图表题的复杂度, 因此我们可以通过控制图表变量来控制题目的难度。Xiao Mingxi(2010)也通过控制一些关于图表特点的变量比如学生对图表的熟悉程度来研究英语口语中图表题。国内学者金艳(2002)用定性研究方法对试题的难易度, 题型受欢迎程度等方面进行了研究。此外吕长竑(2006)把口试任务分为 A,B,C,三类并对这三类任务及其任务组合进行了效度研究, 对口语任务的选择具有指导意义。本文将参考不同口试题型的研究设计本校英语水平测试口试部分的题型和命题规范, 并将依据出题规范所撰写的样题进行一次预测试。我们将根据预测试的考试成绩和考试后对考生的问卷调查结果提出修改意见将其完善。

二、研究方法

1 研究对象

本次研究按照出题规范上对题型的规定编写样卷，并于 2015 年 1 月 11 日对本校 61 位来自 4 个班级的非英语专业大一学生进行了口语预测试。研究数据来源主要是考生的考试成绩、考生问卷调查的答案。

2 研究步骤

本研究采用定性和定量相结合的研究方法。首先按照命题细则里设计的口语机考题型对 61 名学生进行预测试，然后对这 61 名考生的考试成绩进行统计分析，包括相关分析和因子分析以及考试结束后的问卷调查结果分析，所有数据研究都采用 SPSS 统计软件计算。

3 研究问题

本研究的主要目的是研究考生对不同口语题型的主观反应和态度以及本研究中题型是否能够有效地反映学生的口语能力。

三、口语测试系统及题型设计

1 口语测试系统

本次口语机考采用的是湖南大学英语分级测试考试系统。考场为多功能数字媒体教室，为每位考生配备一台电脑，一个带话筒的耳机。考试过程中有两名英语老师，一名技术人员。考生直接输入身份证和学号登录电脑桌面上的考试界面，在监考老师指令下统一作答，答题时间控制在 10 多分钟。考试结束将由两位经过系统培训的评分员单独评分，评分实行网上人工阅卷并在指定的考试中心统一集中完成，评分过程中每部分试题评分标准自动显示在电脑屏幕上，评分方法采取三读的阅卷形式，尽可能减少了评分的不一致性即如果两位评分员所评分数相差一个等级以上，则由三阅评分员给出最终分数。评分标准借鉴新托福机考口试分别从话题展开程度，语音语调和流利度及语言使用的准确性三个方面进行评分。口语评标准简洁清晰、界定明确、易于把握。

2 题型研究与设计

表 1 用表格形式列举了目前常见的口语测试题型，表中归纳了不同口试题型的测试方法及各自的优缺点。研究不同口试题型的特点对于后面题型筛选，设计有一定的指导作用。

表 1 常见口语测试题型

测试题型	测试方法		优点	缺点
朗读	短文	非交际性	方便评分	不能测出其综合口语能力
复述	短文	非交际性	听、说或读、说结合，符合日常言语交流活动	复述材料长短，内容控制困难

问答	问题	非交际性	出题容易, 操作简单	缺少互动, 问题的选择
看图说话	图片, 图表	非交际性	较好地测试叙述等常规技能	图片选择容易不清晰, 杂乱
看文字说话	阅读材料	非交际性	结合读和说多种技能, 考察综合能力	材料选择难
听音频说话	音频	非交际性	结合听和说多种技能, 考察综合能力	听力难度, 长度需要严格控制
看视频说话	视频	非交际性	呈现更为丰富的信息	设备要求高, 需要额外的视频播放设备
评论/口头作文	话题	非交际性	能够考察较高语言能力	内容抽象需较高理解力
小组讨论	话题	交际性	可以考察讨论过程中应变技能	讨论过程难以控制, 容易受外在因素影响
角色扮演	剧本	交际性	交际性很强	多名考生间配合较难

首先,本次口语机考的定位是水平测试,即测量学生当下的英语水平和英语综合运用能力是否达到某种水平。水平测试的立足点是现在和未来,因此测试的内容以及其难度应该由测试目的和学生特点决定而不应仅仅着眼于学生所学教材,我国各种等级考试例如四六级口语测试和英语专业八级测试口试都属于英语水平测试的范畴。

本口语考试的设计参考了杨惠中(1990)提出的的大学英语口语考试设计原则:1)提供真实的交际情景通过需求分析确定考生需要使用英语的场合。2)创造真实的交际语境,考生在说话的时候要以一定的角色、对一定的对象说话、达到一定的交际目的。3)语言能力是交际能力的基础,设计口语考试必须以信息为出发点,设计交际情景时应尽量体现交际功能。

基于对不同口试题型分析基础上,根据本校水平测试的测试目标设计了四种考试题型如表2所示,它们分别是热身问答题,描述图表题,听对话回答问题和口头作文题。其中第二部分图标描述题和第三部分的听对话回答问题属于综合型任务,分别结合了‘读和说’以及‘听和说’等技能,考察了学生的英语综合运用能力。

表2 水平测试口试题型

序号	考题类型	考试时间	考试任务和要求	考察技能
Part 1	热身 (warm-up)	准备时间 15秒 答题时间 45秒	1. 提问考生个人信息例如自我介绍。 2. 问1个疑问句,涉及考生最熟悉的事物或与考生相关信息,考生简要陈述。	基本表达能力
Part 2	图表描述 (describing)	准备时间 50秒	为考生提供一张图表,图表一般为柱状图,线形图,饼状图及表格,	把图表信息转化成

	and explaining a graph)	答题时间 70 秒	考生有 50 秒的准备时间然后做连续陈述。	口语描述
Part 3	听录音回答问题 (Answer the questions after listen to a conversation)	准备时间 每题 30 秒 答题时间 每题 1 分钟	为考生提供一段 40 seconds 对话音频, 考生听完录音后会听到两个问题, 一个问题是考生需要总结归纳听力材料中的观点 (1 分钟), 另一个扩展问题是考生将就录音材料同样话题发表自己的看法提出建议 (1 分钟)。听力播放过程中可以做笔记。	归纳总结接收信息, 将听到的语言信息用自己口语表达出来
Part 4	口头作文 (Oral presentation)	准备时间 30 秒 答题时间 1 分半钟	口头作文题要求考生从给定的两个相反对立的观点中做出选择, 充分深入的陈述自己的看法, 并就自己的观点展开说明, 要求论据充分, 最好有例子或亲身经验。	连贯表达和语言组织能力

四、研究结果和分析

1 考试难度和考生成绩分布情况分析

由表 3 可知, 整个考试的难度值为 0.401 各, 接近理想难度值 0.5。每部分的难度系数在 0.397 到 0.459 之间, 在规定的 0.3 到 0.7 的难度范围之间, 说明所设计题型符合考试大纲要求。考试大纲评分标准规定 050 为最高等级, 010 为最低等级, 表 4 显示大多数考生成绩集中在中间等级 030 至 040, 63.93% 的考生成绩为 030, 19.67% 的考生成绩为 040, 高分段和低分段人数较少, 由于样本数量有限, 最高分人数为 0。

表 3 考试难度系数分析

类别	第一部分	第二部分	第三部分	第四部分	总体
平均数	30.164	27.066	28.410	29.328	29.67
总分	50.00	50.00	50.00	50.00	50.00
难度系数	0.397	0.459	0.431	0.413	0.401

表 4 各级别人数和百分比

Final level	Population	Percentage
-------------	------------	------------

010	4	6.56%
020	6	9.84%
030	39	63.93%
040	12	19.67%
050	0	0%
Total	61	1

2 口试题型效度分析

相关分析

本研究对考生的机考口试成绩进行了相关分析。从表 5 可以看出题型与题型间的相关系数在 0.502~0.775 之间,这说明题型间的相关程度是不同的。一般来说,各题型之间相关系数越高,则考试的内容结构就越合理。但如果相关过高时则说明各部分并不是考察相对独立的语言能力因素,因此各部分对整个考试的作用不大。每部分与整体相关系数在 0.614 到 0.868 之间,实验研究分析表明考试每部分与整体以及部分与部分间的相关系数都较高,因此相关分析说明了试题考察的是考生的综合口语能力。

表 5 内部相关性分析

		第一部分	第二部分	第三部分	第四部分	总体
第一部 分	Pearson Correlation	1	.528**	.775**	.729**	.747**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	61	61	61	61	61
第二部 分	Pearson Correlation	.528**	1	.502**	.665**	.614**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	61	61	61	61	61
第三部 分	Pearson Correlation	.775**	.502**	1	.736**	.868**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	61	61	61	61	61
第四部 分	Pearson Correlation	.729**	.665**	.736**	1	.855**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	61	61	61	61	61
总体	Pearson Correlation	.747**	.614**	.868**	.855**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	61	61	61	61	61

** . Correlation is significant at the 0.01 level (2-tailed).

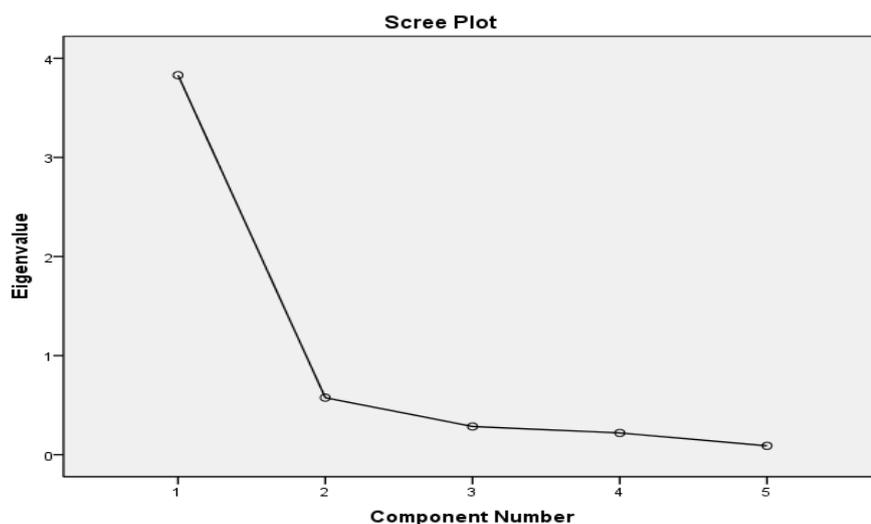
因子分析

由表6和表7我们可以看到通过主成份分析法和碎石法进行的验证性因子分析得到一个共同因子,这一结果进一步证实了考试的结构效度在合理的范围之内。通过相关性分析和因子分析共同证明了本口语题型具有较好结构效度。

表6 总的解释方差

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total
1	3.829	76.584	76.584	3.829
2	.576	11.516	88.100	
3	.285	5.702	93.802	
4	.220	4.399	98.201	
5	.090	1.799	100.000	

表7 碎石图



3 调查问卷分析

我们对本次口语测试的61位考生进行了问卷调查,了解考生对不同题型的看法和意见,为方便研究,共收回61份有效问卷。针对题型的问题简化为最喜欢的题型,最不喜欢的题型,最能测出口语水平题型,最难的题型,最容易的题型。并且在后续的访谈中,不少考生表示增加听力的输入,能够刺激他们的语言输出速度和数量。口语机考增加图片和音频双重视觉刺激,多种视听器官并用更接近真实的口语交际情景,使他们能更好地接受信息从而发挥自己的真实水平。

表8 部分实测问卷调查结果

问题 \ 反应	同意	比较同意	不太同意	不同意
考试反映出了英语口语水平	21 (34.4%)	28 (45.9%)	7 (11.5%)	5 (8.2%)
口语话题贴近生活,有话可说	19 (31.1%)	32 (52.5%)	10 (16.4%)	0 (0%)

从表 8 可见, 34.4%和 45.9%的考生选择“同意”和“比较同意”而持否定意见的考生只占 11.5%和 8.2%, 表明大多数考生赞同该考试反映出了他们的英语口语水平。对于选项“口语话题贴近生活,有话可说”, 31.1%和 52.5%考生表示赞同, 16.4%的少数考生选择“不太同意”。

下表整理了对于口试中四种题型分别的喜欢程度, 认可度和难易程度, 由表 9 可知多数考生 (37.7%) 最喜欢的口试题型是第一部分 (热身问答题), 最不喜欢的是第三部分 (听对话回答问题)。42.6%的大多数考生认为第四部分口头作文能够真实反映他们的口语能力, 而第三部分 (听对话回答问题) 最不能达到这一点。大多数 49.2%考生认为第三部分 (听对话回答问题) 难度最大, 普遍反映准备时间不充分, 且大多数学生认为口试时周围同学回答问题的声音以及音量调节问题对他们听对话的影响很大。67.2%的考生认为第一部分 (热身问答题) 相对容易, 最适合他们语言水平、难度适中、答题时感觉最舒适。

表 9 针对题型问卷调查结果

问题 \ 题型	第一部分		第二部分		第三部分		第四部分	
	(%)	人数	(%)	人数	(%)	人数	(%)	人数
最喜欢的题型	37.7%	23	13.1%	8	16.4%	10	32.8%	20
最不喜欢的题型	16.4%	10	31.1%	19	42.6%	26	9.8%	6
最能测出口语水平的题型	19.7%	12	19.7%	12	18%	11	42.6%	26
无法测出口语水平的题型	16.4%	10	31.1%	19	44.3%	27	8.2%	5
最难的题型	3.3%	2	32.8%	20	49.2%	30	14.8%	9
最容易的题型	67.2%	41	6.6%	4	6.6%	4	19.7%	12

五、讨论：对题型设计的反思

分析结果显示大多数考生对本次考试题型基本认同, 但针对调查发现考生普遍觉得综合型口试题较难, 不能反映他们的口语能力。本研究认为导致这一现象的原因可能有以下几点: 1) 准备时间不够, 新题型很少接触所以会有一些的不适应性, 建议适当增加答题准备时间, 题目之间留有合理的间隔时间让考生得以缓冲, 能够放松答题 2) 考生对第二部分图表描述

题意见不一致, 调查访问中部分抽到线性图的考生普遍表示难于柱状图, 饼状图等其他类型图表。此外不少考生表示经济类的话题明显难于校园生活, 社会现象等其他话题, 因此建议出题时要严格评估不同图表类型的难度, 较难理解的话题可以排除。3) 为考生提供一个尽量少受外界干扰, 能够轻松, 舒适答题的考试环境, 解决好考试过程中出现的音量小不方便调节等客观因素, 使该考试更能最大程度地反映考生真实英语口语水平。

六, 结论

本研究为我校英语水平考试设计了新的机考口试题型组合和命题规范, 在此基础上进行了一次预测试。通过对考生结果分析, 我们认识到本研究里题型设计中的不足并提出相应改进意见, 考试题型设计还需要在实践中不断完善更好的为今后的口语测试服务。另外由于本研究的实验范围较小, 我们将在今后的更大样本的口试操作中, 对本研究所得出的结论做进一步的检验。

参考文献

- [1] Alderson, J., Clapham, C. & D. Wall. 2000. Language test construction and evaluation [M]. Beijing: Foreign Language Teaching and Research Press. 183-185.
- [2] Bachman, L.F. 1990. Fundamental considerations in language testing [M]. Oxford: Oxford University Press.
- [3] Bachman, L.F. & A.S. Palmer. 1996. Language testing in practice [M]. Oxford: Oxford University Press.
- [4] Bonk, W.J. & Ockey. 2003. A many-facet Rasch analysis of the second language group oral discussion task [J]. Language Testing, (20): 89-110.
- [5] Brown, A., Washita, N & T. Mcnamara. 2005. An examination of rater orientations and test taker performance on English-for-academic-purposes speaking tasks (TOEFL Monograph 29) [R]. Princeton, NJ ETS.
- [6] Fred, D. & K.L. Brain. 2001. Testcraft: A teacher's guide to writing and using language test [M]. Yale University Press.
- [7] Hsu, L-Y. 2009. The effects of family background, academic achievement, and gender on the written narratives of taiwanese second graders [D]. Master's thesis, national Cheng Kung University, Taiwan.
- [8] Munby, J. 1978. Communicative syllabus design [M]. Cambridge: Cambridge University Press.
- [9] Popham, W. J. 1990. Modern educational measurement: A practitioner's perspective [M]. Needham Heights, MA: Allyn and Bacon.
- [10] Shohamy, E. 1994. The validity of direct versus semi-direct oral tests [J]. Language Testing, (11): 99-124.

- [11] Skehan, P. & P. Foster. 1997. Task type and task processing conditions as influence on foreign language performance [J]. *Language Teaching Research*, (1): 185-211.
- [12] Takanori, S. 2011. The contribution of test-takers' speech content to scores on an English oral proficiency test [J]. *Language Testing*, (29):223-241.
- [13] Underhill. 1987. [M]. *A handbook of oral testing technique* [M]. Cambridge: Cambridge University Press.
- [14] Van Lier, L. 1989. Reeling, Writhing, Drawling, Stretching, and Fainting in Coils: Oral Proficiency Interviews as Conversation [J]. *TESOL Quarterly*, (23): 489-508.
- [15] Weir, C.J.1990. *Communicative Language Testing* [M]. London: Prentice Hall International(UK) Ltd.
- [16] Wesche,M.B.1983.*Communicative Testing in a Second Language* [J].*The Modern Language Journal*, 67:41-55.
- [17] Widdowson 1978. *Teaching Language as Communication* [M]. Oxford: Oxford University Press.
- [18] Xi, X. 2005. Do visual chunks and planning impact the overall quality of oral descriptions of graphs [J]? *Language Testing*, (22): 463-508.
- [19] Xiaoming Xi. 2010. Aspects of Performance on line graph description task: influenced by graph familiarity and different task features [J]. *Language Testing*, (27):73-100.
- [20] Yan, J. & Jing, S. F. 2011. Test for English majors(TEM) in China [J]. *Language Testing*, 28(4): 589-596.
- [21] 高丙梁. 2008. 大学英语口语测试研究热点评述 [J]. *外语教学理论与实践*, (1): 92-96.
- [22] 何莲珍,王敏. 2003. 任务复杂度、任务难度及语言水平对中国学生语言表达准确度的影响 [J]. *现代外语*, (2):172 — 179.
- [23] 金艳, 郭杰克. 2002. “大学英语四, 六级考试非面试型口语考试效度研究” [J]. *外语界*, (5): 72-79.
- [24] 金艳. 2005. 大学英语四六级考试改革思路与未来展望[J]. *中国大学教学*, (5) 49-53.
- [25] 吕长竑. 2008. 口语测试评分标准比较研究[J]. *外语教学与研究*, 11(6): 440-445
- [26] 邱东林. 2005. 大学英语听说机考尝试[J]. *外语界*, (4): 76-79
- [27] 秦晓晴. 2003. 外语教学研究中的定量数据分析[M]. 湖北: 华中科技大学出版社.
- [28] 杨惠中. 1999. 大学英语口语考试设计原则[J]. *外语界*, (3): 48-57.

On the Design of Task Type for College Computerized Oral English Proficiency Test

Feng Linglin

(College of foreign language and international education, Hunan University, Changsha / Hunan Province,

410000)

Abstract: On the basis of studying common task types of oral test, four types of test tasks have been designed for Computerized Oral English Proficiency Test of Hunan University according to the testing objectives and the characteristics of the college students. Empirical studies have showed that these four types of tasks are appropriate for oral proficiency tests and can reflect students' oral proficiency.

Keywords: English Proficiency Test; Task Type; Computerized Oral Test

作者简介 (可选): 冯陵琳为 12 级湖南大学外国语及国际教育学院硕士研究生, 外国语言学及应用语言学专业, 语言测试方向。

Tel: 18373177146, Email: 391072709@qq.com