

印尼汉语学习者口语测评特点
**CHARACTERISTICS OF ORAL ASSESSMENT FOR CHINESE
LEARNERS AS A FOREIGN LANGUAGE IN INDONESIA**

Guan Nan¹⁾, Alvina Callista²⁾

Chinese Language Department, Bunda Mulia University
¹⁾guannan@bundamulia.ac.id, ²⁾alvinacalista97@gmail.com

Diterima tanggal 11 Februari 2020 / Disetujui tanggal 25 Februari 2020

摘要

口试是汉语口语课教学必不可少的环节，是检验教学效果最有效的方法之一。通过口试可以发现教学过程中的问题、对学生的学习过程中进行查漏补缺，有助于提高教师的教学水平。本文采用访谈法及观察法对慈育大学中文系一至三年级有关口语课程进行分析，探求口试中存在的问题。结果表明，教师大多使用直接式一对一的考核方式。一年级常用的题型为朗读、造句、问答以及主题讲说；二年级常用的题型为朗读、主题讲说以及角色扮演；三年级常用的题型为对话。一二年级常用分析评分法，而三年级常用任务分项评分法。不同年级教师的侧重点有所不同，一年级侧重于流利度及对时间的把控；二年级侧重于语音；三年级侧重于内容及词汇。慈育大学中文系口试中存在的问题是：口试环境及学校设备、考生因素、无明确的书面评分标准以及缺乏评分培训。

关键词：汉语教学；口语水平；测评

ABSTRACT

Oral assessment is an indispensable part of oral Chinese teaching and one of the most effective methods to assess the teaching effect. Through the oral assessment, problems in the teaching process can be found, and the gaps in the students' learning process can be checked and filled, which is helpful to improve the teaching level of teachers. This article adopts the method of interview and observation to analyze the oral classes of the first to third grades of the Chinese Department of Bunda Mulia University, and to explore the problems existing in the oral assessment. The results show that most teachers use direct one-to-one assessment methods. The commonly used question types in the first grade are reading aloud, sentence making, question and answer, and thematic speech; in the second grade are reading aloud, thematic speech and role-playing; and in the third grade are dialogue. The first and second grades often use the analytical scoring method, while the third grade commonly uses the task sub-item scoring method. Teachers of different grades have different focuses. The first grade focuses on fluency and time control; the second grade focuses on pronunciation; the third grade focuses on content and vocabulary. The problems in the oral assessment are: oral assessment environment and campus equipment, examinee factors, no clear written scoring standards, and lack of scoring training.

Keywords: Chinese teaching; oral proficiency; assessment

引言

口语是人类语言不可缺少的组成部分，语言教学也不能缺少口语教学。口语要求学生综合运用语言知识，在短时间内把想说的内容表达出来，逐渐养成用外语思维的习惯。而测试主要是考察人的知识和技能。通过检测可以测量目标是否达成。

吕必松（2005）认为狭义的口头表达包括学习、日常生活、社会、会议、会谈等口头交流；广义的口头表达包括讲课、讲演、辩论、解说等。汉语口语表达不仅把语音、语法、词汇这些要素知识融合为一体，而且还带有很浓厚的汉语语用和文化色彩。汉语口语测试主要考察学生能否运用正确的语音语调得体、流利地表达；所使用的词汇是否丰富、恰当，语气是否

自然；语法以及口语句式是否正确；能否根据交际的需要灵活选择和变换不同的表达方法，是否能达到交际目的。

吕长竑、宋冰、王焰、刘文丽、黎斌(2008)给出的评分标准定义是口语测试中一个重要的组成部分，如果评分方法不能做到科学、客观和具有可操作性，则很难保证像口试这样的主观性考试的信度和效度。另外，肖骆(2008)认为口试评分最大的问题是评分的主观性。而为了减少评分中的随意性，在评分时应明确评定的内容，预先制定一个科学的评分标准，以提高测试的信度。史天化、唐国平(2012)认为口语评分属主观性评分，往往通过让考生完成一项指定任务，展示其在问题解决中的推理、判断和表达等方面的技能，然后评分员依据既定的标准对考生的表现进行综合评定。

《对外汉语教学的成绩测试》将口试方式分为3类：直接口试、半直接口试及间接口试。而口试题型则分为7类：问答、看图说话、角色扮演、口头演讲、朗读、复述及小组讨论。

口语评分方法大体包括以下几种：分析法、整体法、融合法、分项任务法、记错减分法、加分法。

汉语口语测试相关的研究，在美国和中国对这方面的研究都不少，甚至有专门口语测试的一本书籍。该书籍专门讨论成绩测试，特别是口语测试的评分标准。如：FSI从语音、语法、词汇、流利度、可理解度五个方面来评价口语水平；ELTS从语法的恰当性、语音、词汇、流利程度和展开讨论的能力，四个方面来评价口语水平；CET-SET从语言准确性和语言范围、话语的长短和连贯性、语言灵活性和适切性六个方面来评价口语水平（杨翼；2010）；HSK(高等)口语考试规定从内容、语法、词汇、语音语调、流利程度五个方面评价汉语口语水平（王媛媛；2004）。尽管在国内对汉语口语相关的研究很多，但是在口语测试方面却寥寥无几。

在慈育大学中文系口语课（包括听说课以及商务口语课）进行过程中，我们发现随着年级的上升，教师慢慢不再关注学生的词汇或是语音方面的问题。因此，本研究选择口语测试情况为研究主题，以慈育大学中文系口语教师作为研究对象。探

究口语测试常用的方式、题型、评分方法、教师的侧重点及口试过程中存在的问题。

类似的前人研究包括王媛媛（2004）的《HSK(高等)口语考试两种评分方法的效度比较研究》，以高等汉语水平口语考试为研究对象，检验了HSK(高等)口语考试的两种评分方法（综合评分、分项评分）的效度，分别计算了两种评分方法与两种效标的相关系数。该研究通过问卷调查法对这两种评分方法进行分析。调查结果表明两种评分方法都具有良好的效度，但两者与效标的相关系数差异不显著。

刘若岚（2012）的《对长沙市中考英语口语测试的现状调研》，通过问卷调查、访谈以及观察法对长沙市中考英语口语测试的效度、评分信度、态度、看法以及各方面的评价进行分析。调查口语测试评分信度结果表明有6个因素导致口语测试评分过程中难以确保准确性和难以评价其公平与公正：1. 无明确的书面标准、2. 考官评分培训不够、3. 考官评分经验不足、4. 评分区分度不好、5. 评分员评分一致性不好、6. 对主观评分的监控不够。

本研究与前人研究的不同除了研究对象有明显区别之外，从评分方法角度来看，前人研究只针对两种方法（综合评分法和分项评分法）的效度进行研究，而笔者针对5种评分方法找出口语教师常用的评分方法，同时也找出常用的口试题型及方式。此外，本研究还考察了不同年级口语教师的侧重点是否相同。

研究方法

本研究的对象是慈育大学中文系六名教授口语课程的教师，包括：一年级口语课、二年级听说课、三年级商务口语课。研究采用访谈法以及观察法。访谈设计分成两个部分，第一部分是口语测试情况，包括口语测试方式及形式、题型及评分方法；第二部分主要围绕口语测试评分的情况。

在观察部分，笔者选择了三个年级口语课程的期中考试。其中由于二年级两个平行班考试时间重合，因此我们只选择其中一个班级进行观察。

研究结果与分析

针对收集到的数据和信息，我们从五个角度来进行分析：测评准备、时长及反馈情况；口语测评方法及形式；口语测评题型；评分方法；评分标准；测评中出现的问题。具体分析如下：

1) 测评准备、时长及反馈情况：

(1) 测评准备

教师在考前会给学生讲解测试范围，并提供 3-4 个取材自课本的主题，但没有针对性的复习或模拟练习。

(2) 测试时长

教师基本是按照学生的考试座位顺序或自愿排序来进行考试，考试时间一般控制在五分钟/人，入场前准备时间稍有区别，一、二年级五分钟即可，三年级因为难度提升较大，所以准备时间为十~十五分钟。

(3) 测评反馈

期中考试考完试的第一天上课，教师会向学生反馈口试中出现的问题。教师将会以分数形式向学生反馈，解释分数所包含的信息，以此鉴别学生的学习成果，诊断学生学习的不足。除此之外，有的教师在口试整个过程结束后，会立刻给学生反馈口试中出现的问题。

2) 口语测评方法及形式：

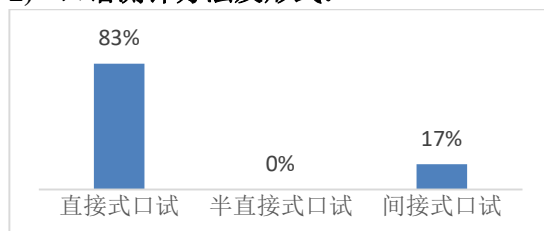


图 1 教师常用口试方式

直接式是慈育大学中文系口语教师最常使用的口试方式，占 83%。其次，根据观察与访谈所获得的信息，我们发现口试最常使用的形式分为两种，一是个人、二是小组。从口试题型或口试任务，我们可以看出教师所使用的口试方式及形式，比如：二年级的口试题型为朗读及主题讲说，那么在进行考试时，教师会使用一对一直接式来进行考试。另外，若考试题是角色扮演或者对话，就会采取小组形式来进行考试。使用直接式可以让教师直观考察学生语言表达能力，了解学生的口语水平。

虽然使用直接式费时费力，而且考官只能通过一次测试对考生进行有限的考核，但是在真实的交际情境中，考官身临其境观察考生的汉语表达、表情体态，考生也可以与考官互动交流。此外，口试的考官一般是任课教师本人，考生显得较为熟悉，产生信任感，因此可以在一定程度上降低考试紧张、焦虑等负面情绪，如：二年级二班的听说课，因为学生对任课教师都很熟悉，所以在笔者观察过程中很少遇到表现紧张的学生。

其实，上述口试题型也可以采用间接式口试来进行。间接式一般需要事先将考试内容录制成音频或视频等影像资料，通过多媒体播放。考试过程中学生面对电脑等多媒体设备，根据给出的口试题做出回答，考官在考试过程中不会当场评分或做记录，而是考试整个过程结束后才开始进行评分，例如汉语水平口语考试（汉语口语水平）就是典型的间接式口试。使用间接式既省时省力且评分也方便，但是对测试设备的要求较高，若出现设备损坏等情况会影响到考生成绩，如：二年级一班的听说课期中考试进行过程中，因为个别耳机的收音功能不完善，只能收录到考生一部分的语音信息，因此教师无法听清楚考生的完整口试内容。由此可见，直接式的比例比间接式高，主要原因是对学生的汉语水平提高实际效果更有帮助，能即时获知学生出现的问题并加以纠正，有助于学生获得最及时的反馈。

3) 口语测评题型：

为了保证测试的效度，题型的选择非常重要，试题的好坏直接关系到测试的效度高低。若试题的设计不合理，将无法达到测试目的，导致测试无效。

表 1 教师常用口试题型

题型	一年级	二年级	三年级
主题讲说	1	1	
朗读	1	1	
角色扮演		1	
对话			1
造句	1		
问答	1		
种类	4	3	1

慈育大学中文系口试题型都与学校考试大纲及课本内容保持密切的关系。较常

采用 6 种题型，分别是：主题讲说、朗读、角色扮演、对话、造句及问答。一年级采用的种类最多（4 种），其次是二年级（3 种），而三年级只采用对话一种题型。我们认为，这是由于一、二年级测试的是学生的整体口语水平，需要从多角度全方位来测试，而三年级主要侧重于对商务口语方面的考察，所以受特定场景所限，只通过对话形式呈现。

我们根据不同年级所采用的六种口语题型进行分析，具体如下：

(1) 主题讲说：

一、二年级常使用主题讲说作为期中或期末考试题型。主考教师提出预先准备好的话题，要求学生围绕所给的话题进行表达，主题讲说的内容与考生学过的内容关联紧密，所以主要目的是为了考察考生对所给定的主题是否已经熟练掌握。

通过进一步的分析，我们发现，同样是要求针对一个给定的话题进行陈述，不同汉语等级的主题内容会有一定差别：一年级的期中及期末考试的题目分别有：我的家人、我的爱好、学习汉语的理由、去菜市场买菜、圣诞节、我最喜欢的教师、我喜欢的运动、我的周末，等。从一年级期中及期末考试的题目可得知，一年级的话题都侧重于个人爱好、节日、喜欢的运动、我的周末等简单的话题。因为一年级的学生刚开始接触到汉语，且汉语水平能力有限，因此话题一般仅限于一些常见的简单话题。观察二年级期中考试题目，我们发现分别有：三天两夜、毕业以后的打算、我的理想，等等。可见，二年级的话题侧重于看法、理想等方面，原因是学生已经接触了一年多的汉语，汉语水平处于中级阶段，为了能更准确考察学生的汉语水平，在主题讲说的话题涉及到了看法、理想等方面。由此可以看出，不同的汉语水平主题内容上的难度也是有差别的。

(2) 朗读：

朗读是口语测试题第二种最常见的测试题型，通常出现在一、二年级的期中考试题目中，且两个年级的朗读句子有所不同，一年级朗读句子是日常口语表达，而二年级朗读句子大多是书面口语表达。对于初级阶段来说语音的准确性是学习的重点。而到了中级阶段，语音虽然不再是学习的重点，但是据我们的观察，有些二年级学

生的发音还存在印尼式腔调，不少学生也能明确意识到自身发音还不过关。教师通过朗读题型，可以重点考察学生的语音方面，根据他们对语音、语调、重音、停顿、语流、音变等项目的掌握情况，了解到学生的发音已达到什么程度。学习一门外语，要想达到发音标准、流畅地道，绝不可能一蹴而就，正音需要一个长期的过程。因此朗读是初级与中级口语测试中必不可少的题型。

(3) 角色扮演：

二年级的口试题型除了使用朗读及主题讲说以外，在期末考试还经常会采用角色扮演这种形式来进行。原因是因为听说课的教学目的要求，除了考察学生听的能力，也要考察说的能力。学生在角色扮演过程中除了要表演自己的角色，也要注意听对谈者在说什么。由此可见，通过角色扮演，学生能更好地发挥学习的积极性、主动性及创造性，充分参与学习的整个过程，有助于对课文内容有更深刻的理解。

(4) 对话：

三年级商务口语课常采用对话作为考试题型。虽然与二年级角色扮演有相似之处，即根据所给定的场景表演，但是因为三年级的口语课以商务方向为主要内容，而且一般在商务场合中经常出现的是两个人之间的对谈，如：面试、与某公司谈合作、买卖交易、签订合同等经济贸易领域方面的交流。通过对话这种题型可以有针对性的帮助学生操练，如何与客户沟通业务、如何确定展位等。因此三年级口试题型为对话形式，但是里面也包含一定的表演成分。

(5) 造句：

造句是一年级口试中常出现的一种考试题型，通常出现在期中考试，如一年级的期中考试的造句题：“太”、“别的”、“离”、“从”。另外，我们在观察中发现有些学生所造句子只有三个字，过于简单，或者造句时还是习惯性套用印尼语语法，例如：“太好了！”、“你好吗？”、“我走路从家到学校。”因为一年级的学生接触汉语的时间还很短，对很多词语的意思及用法都不甚理解，而造句题型可以直观考察学生对遣词造句知识掌握的准确性，若出现语法错误则一目了然。因此造

句题型较常出现在对初级阶段汉语学习者的口语考核中。

(6) 问答:

一年级期中考试除了常用朗读、造句、主题讲说口试题型之外,在期末考试也使用了问答题。在问答中较为常见的一种方法便是主考教师提出事先准备好的问题,要求学生回答该问题。另外一年级期末考试问答题的内容大多是关于个人喜好、节日、有关学习汉语的情况、有关天气等这些相对简单的问题,例如:问:“你喜欢吃中餐还是西餐?”答:“我喜欢吃中餐。”通过进一步的分析我们发现,问答题可以考察学生是否会使用汉语来回答问题,因为初级阶段汉语学习者词汇量有限,因此在回答过程中偶尔会使用母语来代替。由此可见,一年级口语教师较常采用问答题的目的,除了考察学生的汉语水平以外,也有助于鼓励学生尝试使用汉语来思考并回答问题。

4) 评分方法:

慈育大学中文系口语教师最常用的评分方法分为两种:一种是一二年级口语教师常用的分析评分法,另一种是三年级口语教师常用的任务分项评分法。

一二年级口语教师常用分析评分法来进行打分。原因是因为一二年级的学生属于初级和中级阶段汉语学习者,在这个阶段需要关注许多基本标准。虽然使用分析评分法费时费力,但是教师认为可以更全面、更细致地评估学生的口语水平,信度高,且有助于对考生提供较为丰富的评价信息,分数解释力较强。对于经验不足的教师也较为容易操作。

而三年级口语教师常用任务分项评分法,则是由三年级口语课的专业性决定的。由于是商务口语课,教师会要求学生在考试时使用相关商务术语来进行对话。在这个阶段,语音、语法等标准已不再是教师的教学和考核重点,而是对内容、词汇等更为关注。因此三年级的口语教师会根据考生对每项口试任务的完成情况分别打分,除此之外,为了能够更加方便记录,教师也会使用记错减分法。

5) 评分标准:

截至目前,慈育大学中文系尚未举办过专门针对口语测试评分员的相关培训,但据我们的观察,六名口语教师在评分过程中都不存在任何原则性的失误。其次,中文系没有统一固定的评分标准,一般都是将权责下放给任课教师,按照具体课程的要求设置口试的评分标准,有的教师在具体操作时参考上学期的评分标准,有的教师参照汉语水平口语考试的评分标准,而有的教师根据考试的要求设置评分标准。另外,任课教师所做的评分标准事先都会向系主任报备,由系主任判断批准是否符合考试的要求。此外,五位口语教师明确表示在评分过程中能够做到客观公正地对学生们的真实水平给出对应成绩,唯有一名教师表示不太确定,因为学生的成绩都是根据教师自己的标准来打分,没有客观统一的参照对象。

由此可见,学校虽然不曾举办过相关的评分培训,但这基本不会影响口语教师的实际操作。虽然没有固定的标准,但是为了保持口试的信度及效度,评分标准都会根据口试的要求来设置评分标准,且教师们也会参考前人的评分标准。至于那位对自己订立的标准是否客观公正表示不确定的口语教师,我们认为在口语评分过程中带有一些个人的主观性,这是不可避免也无可厚非的事实,只要在大方向的判断上不出现重大偏差,都还是在可以接受的范围之内的。

由于学生的年级不同,汉语水平有区别,有些教师在初级阶段较关注语音、语法等基本标准,而到了中高级阶段则不再成为关注的焦点。此外,不同教师想法各有不同,有的教师认为可理解度、有趣性也属于考核的范畴,而有的教师则认为不必列入到评分表里。

为了确认不同年级口语教师的侧重点,我们共设定了12个方面,分别是:语音、语法、词汇、内容、流利度、可理解度、话语的长短和连贯性、语言的灵活性和適切性、说话的逻辑性、时间把控、表情及有趣性。每项分成五个程度,最低的是一,为很不重要,最高的是五,为非常重要。具体结果如下:

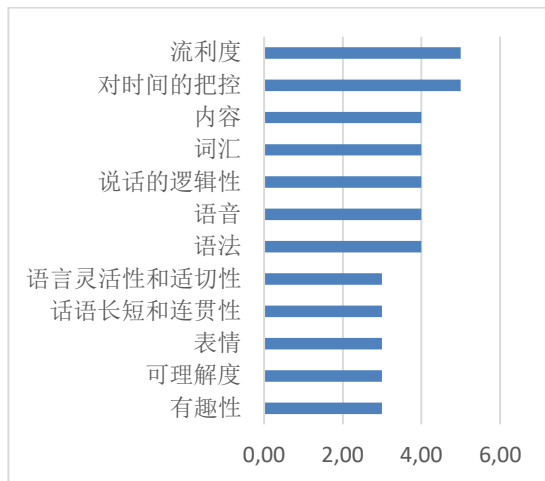


图2 一年级口语教师的评分侧重点

一年级口语教师较侧重于流利性及对时间的把控，因为一年级的口语题目语句都比较简短，如果学生说得太慢或出现多次停顿，会严重影响到听者对话语的理解度，所以教师侧重于根据流利度考察学生的语言交际能力。另外由于一年级学生的人数较多，再加上使用一对一直接式的考核方式，若学生超时会严重影响到后续考试的进度，因此教师尤其强调学生对时间的把控，在一年级口语考试过程中会严格规定学生不可超时。

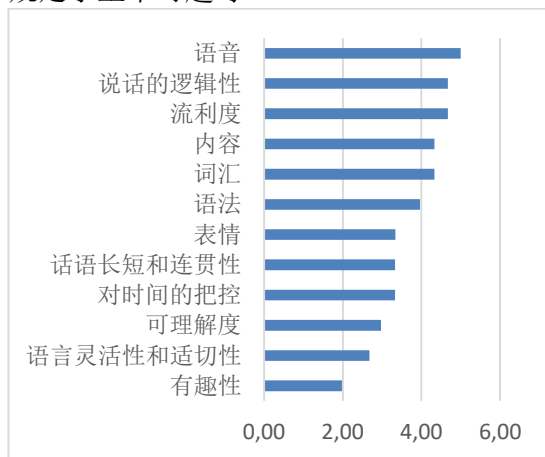


图3 二年级口语教师的评分侧重点

二年级教师对于12种标准的选择按比例最高的是语音。教师认为语音是语言技能教学的基础环节，从语音是否标准可以看出学生汉语水平的优劣，而且便于及时纠正。二年级的学生刚接触一年多的汉语，语音方面经常出现错误，如：某些汉语特有的发音不够准确、带有印尼语式固有腔调等，因此在进行评分时，二年级口语教师主要侧重于语音方面的评判。

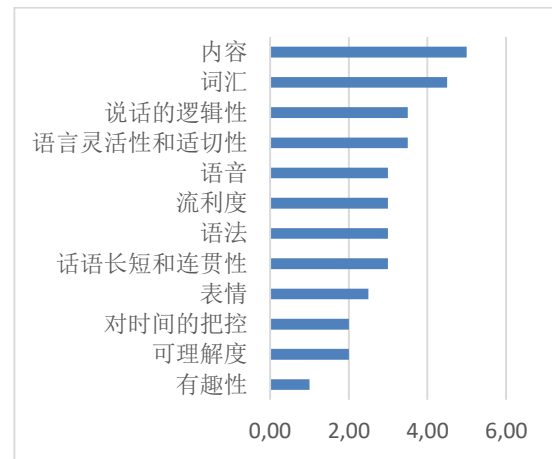


图4 三年级口语教师的评分侧重点

随着年级的增加，三年级口语教师对基本标准的要求不再像一二年级教师那么严格，因为三年级学生在语音、语法、流利度等基本标准这些方面出现的问题并不太多。这个阶段的关注焦点主要集中在陈述内容的深度与广度，及所使用的词汇尤其是商务方向相关术语是否恰当。因此，三年级口语教师更注重内容及词汇的原因是为了考察学生是否对课本上的知识点理解并掌握。

经过以上的分析可以得知，因为要考察学生的基础语言交际能力，且考试时间十分有限，因此一年级口语教师侧重于流利度及对时间的把控；二年级学生虽已接触汉语一年有余，但是在语音方面经常出现错误，因此为了提高学生发音的准确性，二年级听说课教师更侧重于语音；三年级学生在语音、语法等基本标准方面偶尔还是会出现一些错误，但是因为课程涉及到商务，因此三年级口语教师更侧重于内容及词汇。

6) 测评中出现的问题:

影响口语测试评分的因素很多。笔者对六名口语教师进行访谈，并对三个年级口语课的期中考试进行观察，发现有三方面因素会对口语测试产生影响，分别是：口试环境及口试设备、考生因素、评分员的因素，具体分析如下：

(1) 口试环境及测试设备:

① 口试环境

通过对三个年级的期中考试观察来看，考试当天三个年级的口试环境均十分嘈杂，其中尤以一年级的口试环境最为恶劣。因为先考完的学生并没有直接回家，而是仍

滞留在教室外和其他同学聊天，每位考生在离开考场教室的时候几乎都没有随手关门，因此外面的吵闹声在教室里都能听得清清楚楚。

采访中，一年级口语教师也表示，口试考场的周边环境会影响考官进行评分，而且从观察中笔者能看出该教师很难集中精力进行评分。然而，尽管该教师在整個考试过程中不胜其扰，但并未采取进一步行动警示学生们保持安静。这很可能是因为考试使用直接式，再加上时间很紧张，因此教师没有充足的时间，反复游走于考场内外来提醒学生维持考场周边肃静。我们认为，为了有效避免这一现象的发生，在开始考试之前，教师可以先跟学生宣布相关规定，如：进出考场随手关门，考试结束迅速离开，并指派同学互相监督，必要时可考虑与考试成绩挂钩。如此一来，确立了彼此认可的游戏规则，有助于共同营造一个舒适的口试环境，以便教师在考核评分时可以更加集中精神，保证客观公平。

② 测试设备：

作为监考教师，为保证考试过程顺利，需要事先确认相关环节，如考生的顺序、测试的时间、测试步骤、测试环境等。而采用间接式来进行考试，考官也要确认测试的设备，若准备不充分或者有特殊情况，会直接影响到测试的有效性。据我们的观察，二年级一班期中考试发生了设备故障的情况。该班采用间接式进行期中考试，尽管正式开始之前教师已为同学们进行了一次试听，大家也表示没有问题，但是直到教师准备评分时，才发现有三名学生佩戴的耳机收音功能出现了问题，断断续续，无法完整录制全部语音信息，迫不得已，只能请相关三名考生再次进行考试。所幸考生回答的结果与之前几乎一样，没有产生太大的偏差，事后也没有考生对公平性提出异议。

出现这种特殊情况确实超出了考官掌控的范围，属于口试客观因素。一般发生这种情况，除了或多或少会降低口试的信度，也会给考官和考生带来诸多不便，因为教师要再次安排时间对录音出现问题的学生进行补考。可见，考试若采用间接式，测试设备是很重要的一个环节。我们要事先反复确认，保证测试设备能够顺利地录

制考试的整个过程，尽量避免出现因器材故障问题导致考试无效的情况发生。

(2) 考生因素：

云雾（2014）认为考生的个体因素也是很重要的一个方面，具体还可细分为必然因素与偶然因素。必然因素包括考生的知识背景、性别、学习动机等等；偶然因素有考生的焦虑、紧张以及其他突发情况。在观察中笔者发现一至三年级的学生在考试时出现紧张、考试前不认真及迟到等现象。具体如下：

① 精神状态

通过观察我们发现，许多学生在考试过程中都出现紧张的状态，无论准备得充分与否，考试当天学生或多或少都会表现得比较紧张。我们认为这可能与口试方式有关联。直接式主要是以面试为主，学生与教师面对面进行谈话，考官由考生日常已经熟悉的任课教师来担任，原以为可以适当降低考试的紧张、焦虑等。但是从观察结果来看，情况却恰恰相反，很多学生特别是一年级及三年级在与考官面对面的谈话中无法以平常心面对，手脚肢体控制不住地颤抖，不敢与考官有眼神接触，交流中多次出现停顿，有时停顿长达几十秒，更有甚者还会导致发挥失常，比如三年级商务口语期中考试有一名考生，由于过分紧张，在和另一名考生进行现场对话过程中，无意间抢念了对方的脚本，打乱既定安排，对方考生也受该考生影响，思路出现混乱，导致语无伦次，整个过程让人啼笑皆非。

经进一步调查，我们发现，一年级的学生由于刚刚进入大学，对同学、教师及周边环境还不太熟悉，尤其是那些自身性格比较内向，平常本就不太善于与人沟通交流的学生，突然让其面对教师并使用尚不熟悉的汉语来交流，难免会噤若寒蝉；而三年级的课程基本都由中国教师讲授，每个学年都会更换新教师，学生与新外教之间还需要时间磨合，一切尚在适应当中，再加上商务汉语的专业性难度也远远高于普通交际，因此，学生产生紧张情绪在所难免。

② 考前准备

临考试之前，教师都会为学生提供考试范围或考试主题，方便大家做好相应准备。但是据我们的观察发现，有些学生没

有养成良好的考前复习习惯，不知从何入手；也有的是自大轻敌，认为自身水平足够好，即便考前没有认真准备，也足以应付。结果导致考试中经常出现停顿，陈述内容思路凌乱，语无伦次，逻辑性不强，等等，最终考试分数与之前的预期相去甚远。

在观察三年级商务口语期中考试时，我们发现很多学生有迟到的现象。而当教师在开考前统一解释考试规则时，不少学生都不认真聆听，而是各忙各的。这些都给监考教师带来一定的困扰，迟到的学生在进入教室时会影响到其他已经进入考试状态的考生。不认真聆听规则，也会导致在考试进行过程中出现这样那样的问题，造成教师必须反复解释强调，浪费时间和精力，导致效率低下。

我们认为，以上现象也是完全可以避免的，关键在于游戏规则的制定与严格执行。教师可以在考前最后一次课上跟所有同学达成协议，提前进行模拟考试，让大家熟悉规则和流程，考试当天只需要再重点提示相关规则，无需耗时反复强调。针对迟到的学生，若进入考场时还处于宣读考试规则阶段，则可以安排在靠门边特定座位，不需要大范围走动，基本不会干扰其他考生。而一旦考试开始则立即关闭考场，不得进出，直到考试结束。

(3) 评分员因素：

① 无明确标准

口语测试是一种主观性的测试，要保证测试的信度，其中很重要的一个方面便是要保证评分的客观和公正。据刘润清所说：“许多教师和测试专家都承认，口语测试的评分很难做到客观公正。根本原因是其中要注意的成分太多，而且无法进行定量分析。”因此 Underhill (1987) 建议评分标准的制定一开始就要融入到整个测试设计过程中（转引自刘若岚，2012）。但是从采访结果来看，慈育大学中文系没有明确统一的书面口语测试评分标准。各口语教师或参照之前同一课型的评分标准，或参考其他类似口语考试标准给予评分。另外，虽然有的教师一起讨论协商过评分标准，但是即便同样参考汉语水平口语考试的标准，也会有不同的看法，因为每个教师对评分标准有各自的看法，再加上系内没有详细的书面评分标准，而且口试时

要注意的因素也有很多，因此在短时间内教师无法做到完全准确地判断学生的口语水平，只能根据自己掌握的一些标准与现场的感受进行打分。即使评分要从不同的维度去衡量学生的口语水平，但这样也不能确保口试评分过程中评分的绝对准确性，很难评价其是否做到百分百的公平与公正。

② 缺乏相关培训

慈育大学中文系不曾举办过针对口语测试员评分标准相关的培训。虽然教师们都认为对评分不会有太大的影响，但是缺乏理论与实践相结合的培训，难免会导致在评分过程中产生误差，如：考试过程中学生 A 的语音语调特别好或特别差，又或者在某个观点与众不同或者特别庸俗，这时评分员很可能会受这些显著特征影响，而给出偏高或偏低的成绩。评分误差也可能来源于每个评分员对标准的理解不一致，比如在确定口语教师对不同年级的侧重点时，有的教师就表示可理解度及有趣性不必列入到评分标准里。

此外，曾用强（2007）在《论英语口语考试的评分误差》提到关于从管理手段控制评分误差，从评分培训来看，主要包含两方面的内容：第一，学习评分标准，对口语测试的目标能力达成共识，且必须明确各评分标准具体内容，了解哪些是评分所要考虑的因素，哪些是与评分标准无关的；第二，挑选具有典型示范性的口试样本进行试评，提高评分的信度。因此，即使六位教师有各自一套可操作的评分标准，但是对于其是否可以精准把握并熟练操作尚存在疑点。

结论

通过以上分析，我们进行总结归纳如下：

首先，为了能更加有效地提高学生的汉语口语水平，慈育大学中文系一至三年级最常使用直接式来进行口语考试。其次，根据考试大纲及不同年级的水平要求，一年级常用的口试题型为朗读、造句、问答及主题讲说；二年级常用的口试题型为朗读、主题讲说及角色扮演；三年级常用的口试题型为对话。最后，根据口试的要求，一二年级常用分析评分法来进行评分，而

三年级常用任务分项评分法进行评分，因为三年级对内容要求高，因此通过任务分项评分法可以更方便进行打分。此外，在实际操作中很多教师还会采用记错减分法辅助评分。

每个教师对口试评分标准的侧重点有各自的看法。从分析结果来看，一年级口语教师侧重于流利度以及对时间的把控；二年级口语教师侧重于语音；三年级口语教师侧重于内容及词汇。

根据观察及采访结果，我们发现慈育大学中文系一至三年级的口试中存在一定的问题，包括：口试环境（周边环境嘈杂）以及学校设备（耳机的收音功能不完整）、考生因素（学生紧张、考前准备不认真、迟到等）、无明确的书面评分标准、缺乏评分培训。这些问题都会影响教师对学生口试的评分。

根据上述分析结果，提出建议若干：

1) 对学校的建议

通过慈育大学中文系口语测试情况调查分析，我们建议定期为口语教师或者口试评分员举办专门针对口试评分标准的专业培训，这样对经验不足的教师来说，可以降低口试中出现的主观性因素，从而确保口试信度。

其次，依据不同年级、不同口试题型来制定统一明确的评分标准，并落实到书面文件，以方便后续教授相关课型的教师执行操作，并能较好地保障评分一致性。另外，每次口语测试前，相关技术人员一定要确保测试过程中所需要的设备处于完好状态，保障测试过程进展顺利。

2) 对教师的建议

我们建议教师在开展口语测试之前，可以为学生制定有效的复习清单，让学生更好地掌握所学的知识，做到心中有数。考前提前告知相关考试规则及评分标准。为了避免学生迟到的现象，可以让学生提前十到二十分钟入场准备。虽然中文系目前暂时没有明确固定的书面评分标准，但是教师们在制作评分标准时，可以多参考一下与测试题相关的评分标准。

在测试过程中，考试环境要安静、明亮、设备齐全。每位考生拿到考题后，留出适当的准备时间，考生顺序等其他具体操作步骤要注意客观公平，保证测试的效度和信度。教师们还可以借助计算机技术

和多媒体等有效设施来帮助实施口语测试。考生们在指定地点和规定的时间内考试，并用录音设备录制所有考试内容。当评分员评分有疑问时可反复收听考生的录音，以达到评分的一致。为了减少在评分过程中考官的主观因素，提高口试的可信度及有效度，我们建议口语教师尽量根据学生的真实水平去评分，不要以个人的主观因素（情绪、健康状态、专心程度等）来误导口试评分。

在测试结束后，测试的反馈作用应当得到重视和有效利用，对学生进行全面的评价，并对测试本身进行反思。最后，也希望教师们在进行评分时尽量做到科学与客观，避免随意性。

3) 对学生的建议

为了营造更好的口试环境，保障考试顺利进行，我们建议学生们在口语考试结束之后尽快离开，如果有必须停留的理由也一定要尽量保持安静。

其次，考试当天一定要调整好心态。对于容易紧张的学生，考前可以在面对镜子多多练习，这样可以适当舒缓心情。

此外，从态度及学习方法上要认真对待考试，尽量避免迟到、考前不认真等因素对测试成绩产生不良影响。

参考文献

- 云雾. (2014). 新疆中亚留学生汉语口语成绩测试现状调查研究. 乌鲁木齐: 新疆师范大学.
- 刘若岚. (2012). 对长沙市中考英语口语测试的现状调研. 长沙: 湖南大学.
- 史天化, 唐国平. (2012). 口语评分中评分员对评分标准的理解和使用——配对口试评分的报告分析. 广西大学学报.
- 吕必松. (2005). 对外汉语教学概. 北京: 北京大学出版社.
- 吕长竑, 宋冰, 王焰, 刘文丽, 黎斌. (2008). 口语测试评分标准比较研究. 外语教学与研究.
- 曾用强. (2007). 论英语口语考试的评分误差. 考试研究.
- 杨翼. (2010). 对外汉语教学的成绩测试. 北京: 北京大学出版社.

- 王媛媛. (2004). HSK(高等)口语考试两种评分方法的效度比较研究. 北京: 北京语言大学.
- 肖骆. (2008). 构建新课标下初中汉语口语测试模式的研究. 长沙: 湖南师范大学.