

**TESTING ACHEIVEMENT AND PROFICIENCY
IN CHINESE LANGUAGE LEARNERS :
THE ELEMENTARY CHINESE LEARNERS AND
JOB APPLICANTS FOR CHINESE COMPANY**

Chanyaporn Parinyavottichai, Ph.D.¹

Abstract

This paper discusses two types of test utilizing in two testing situations. The first type is criterion-referenced testing. The writer has used it to evaluate Chinese language learning achievements and progress of a group of foreign students enrolled in an elementary Chinese course. The second type is norm-referenced testing. In particular, two proficiency tests are designed for assessing the listening and speaking abilities of applicants who are applying for jobs in a foreign company in China. The writer has also employed the American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines as the

¹ Lecturer, Chinese Section, Department of Oriental Languages,
Faculty of Arts, Chulalongkorn University

basis for choosing test formats and for creating test questions

1. Introduction

Foreign language testing is a kind of assessment in which individuals are asked to demonstrate their knowledge or skills in a particular language. It is a process of gathering accurate information needed for making decisions (Bai, 1998). Basically, testing in a foreign language can be classified into two families based on their functions. The first family, called norm-referenced testing, includes proficiency and placement tests, which play a significant role in assisting instructors in making program-level decisions; the other family, called criterion-referenced testing, includes diagnostic and achievement tests, which assists instructors in making classroom-level decisions (Brown, 2005).

In the first part of this paper, I will describe a set of achievement tests that I have designed for a group of foreign students enrolled in an elementary Chinese course. I will discuss how to use those tests effectively to

assess students' achievements or progress, or whether the teaching has satisfied the objectives of the instructional course. The focus of this section is on when those tests should be given, how to test students' progress in the four language skills (listening, speaking, reading and writing), and which type of test formats should be considered suitable for this type of test. In the second section of the paper I will report on two proficiency tests that I have designed on the basis of the advanced-low scale ² of the ACTFL proficiency guidelines³ and discuss how to use the two tests for evaluating the listening and speaking abilities of the

² In 2004 the Board of Education approved the proficiency level of "advanced-low" for the American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview as alternate tests to the Modern Language Association (MLA) Proficiency Test for Teachers and Advanced Students (DeMary, 2004).

³ ACTFL Proficiency guidelines provide scoring rubrics for summative performance assessment in four major language skills of second language (L2) learner: speaking, listening, reading, and writing. The descriptions of the stages of proficiency are based on four broad levels: novice, intermediate, advanced, and superior. The proficiency of examinees is rated according to their demonstrated abilities in communication in variety of topics and settings, their range of vocabulary used, their language strategies, and so forth.

applicants who are applying for jobs in a foreign company in China.

1. Testing for the elementary Chinese course

The set of tests that I have designed for a group of foreign students in an elementary Chinese course belongs to achievement tests, which are concerned with assessing how students are able to perform in a foreign language and whether they are able to demonstrate essential language abilities in relation to the pre-determined meaningful objectives in a formal language learning setting (Hughes, 2003). Each student's performance is measured according to the preset standard, without comparing the scores to those of other students (McDonald, 2002).

In a foreign language classroom, testing plays a crucial role in indicating the progress and success of L2 learners. In addition to such achievement tests as mid-term exams, administered in the middle of the course, and final exams, given at the end of a term, there are also some progressive achievement tests, for example, unit tests and pop quizzes, which help teachers in measuring

the degree of students' progress over time towards the predefined objectives. Progressive achievement tests are normally administered throughout the duration of a course, and they usually contain certain specific course contents. Unit tests should be given right after a lesson is learned, while a pop quiz, an informal style of testing, should be given whenever an instructor wants to trigger or improve students' motivation for foreign language learning, or to direct students' attention to some important issues that students may not have sufficiently practiced.

As learning outcomes are derived from the objectives of an instructional course, before designing any assessment activities, teachers should first identify the course objectives, which should include all of the knowledge and skills they expect their students to master. The instructional criteria not only serve as a compass in leading students onto the right path of achievements in a foreign language, but they can also be identified and adopted as useful measurement techniques in the classroom (McDonald, 2002)

Since the students who attend the elementary Chinese course have no or only a little prior experience in the Chinese language, the broad goal of this course should be to help them acquire basic skills in speaking, listening, reading comprehension, and writing. In this study, I decided that these four skills should be set approximately at the novice-high level on the ACTFL proficiency scale. A detailed description of the level is given below:

Listening	Students are able to understand short and certain sentence-length utterances they learned, especially where context heavily supports understanding. Students are also able to comprehend limited vocabulary involving those necessary for survival in the target language society such as kinship terms, immediate need expressions, addresses and directions, activities, time. Simple questions and statements, high-frequency commands, and courtesy formulae are also anticipated.
Speaking	Students are able to make statements involving learned materials by relying on memorized utterances. Conversation is limited in predictable daily situations relating to topics areas covering basic personal information, basic objects, like and dislike, and limited number of activities. Vocabulary centers on areas such as basic objects, places, and most common kinship terms.
Reading	Students are able to use learned vocabulary in reading for instructional and directional purposes, standardized messages, phrases, or expressions, such as some items on menus, schedules, maps, and signs.
Writing	Students are able to write simple, fixed expressions, limited work requirements and some recombination thereof. Students are also capable of providing information on simple forms and documents, and writing their simple autobiographical information, numbers, dates, some short phrases, and simple lists.

Using the criteria above, I can decide on appropriate formats for the test items that I shall construct

for assessing the students' four language skills. One thing that should be kept in mind is that each assessment method has its own limitations; therefore, good tests should employ a variety of formats in accordance with the contents of the test and the objectives of the course, and as a result can bring the most beneficial backwash effects to students. For example, in testing speaking or writing we should not use multiple choice tests. We should test them directly instead, and that can present a truer picture of students' abilities (Hughes, 2003; McDonald, 2002). Otherwise, it can create a situation where the backwash effects are harmful to learners (Hughes, 2003). Apart from selecting suitable methods of testing, teachers should also make sure that students are familiar with the test format and testing techniques. Any unfamiliar format is likely to make them perform less well, thus bringing unfairness to the assessment (Hughes, 2003).

In testing listening comprehension of the elementary Chinese learners, we should be concerned with how much of the Chinese language the learners at this level are exposed to. If they are at the initial stage of

learning, they may have more difficulties in learning Chinese pronunciations. The testing at this early stage of their Chinese acquisition should therefore be focused on how to distinguish the initials, finals, tones, and *pinyin* of Chinese. Examples of how I would test the initials and tones are presented below:

(1) Listen to the tape and select the initial you hear.

Tape: bān

Answer sheet: A. bān B. pān

(2) Listen to the tape and write down 'S' if the pair of words you hear have the same tone and D if they are different.

Tape: 1. děng lěng 2. tǐ mǐ

Answer sheet: 1. _____ 2. _____

After the students have become familiar with Chinese pronunciations, the teacher can move on to test their abilities in understanding sentences such as (3-5).

(3) Listen to the tape and select the activities you hear.

Tape: 哥哥正在听音乐。

Answer sheet: A.



(4) Listen to a short talk and select the best answer

Tape: 女: 明天晚上你教我学电脑, 可以吗?

男: 明天晚上我得帮王朋练习打球。后天早上怎么样?

What does the man mean?

- A. He is planning to teach the woman to play a ball tomorrow.
- B. He is going to practice playing a ball the day after tomorrow.
- C. He is hesitating over whether he should really help the woman.
- D. He is willing to teach the woman how to use the computer the day after tomorrow.

(5) Listen to the tape and write your answer to each question in English.

Tape: 男: 那是谁的照片? 你可以给我看一下吗?

女: 可以啊。这是我表妹的照片。

男: 她长得很好看。她结婚了没有?

女: 结婚了。

男: 有几个孩子?

女: 只有一个男孩儿。

男: 她儿子今年几岁了?

女: 已经七岁了。

男: 是在美国生的吗?

女: 不是。是在中国生的。

Question: 1. Who did these two speakers talk about?

2. Where was the boy born? _____

Multiple-choice and short-answer question formats are chosen here because they are appropriate techniques for inspecting students' learning behaviors and their language skills in listening comprehension. Multiple-choice test questions can not only work well for testing lower-level skills such as phoneme discrimination (Hughes, 2003), but their scoring system is also perfectly reliable since it is a kind of objective testing. To

reduce the guessing or cheating factor that may be involved in the multiple-choice technique, the short-answer method may be used as an alternative, as is shown in example (5).

One thing worth noting here concerns examples (4) and (5). All of the multiple choices in (4) are in English instead of Chinese, and in (5) the instructions ask students to write the answers in English instead of Chinese. The question arising here is whether it is a good idea to do so. According to Bai (1990), using English in the directions and in the responses of such listening comprehension tests can reduce the possibility of test bias resulting from differences in students' reading and writing abilities. Moreover, the aim of testing listening progress is to let students demonstrate their current listening ability, so only listening skills need to be tested.

Tests measuring one point in language ability at a time like the examples here are known as discrete-point tests. There is a controversy in the L2 testing field over whether a discrete-point method (Lado, 1961; Harris, 1969; Valette, 1967) or the method of using many skills

of a language at the same time, so-called integrative testing (Oller, 1976), is the best choice in foreign language testing. I tend to use a compromised solution by scholars such as Rivers (1981) and Bai (1998), who explain that it is necessary to combine both theories in assessment of learners' language capacities. As Rivers points out, teachers should be aware of the specific skill they want to test and only test for that skill; nonetheless, they can later consider testing overall performance by employing more integrative testing. Examples of the integrative formats that I would use for assessing the students are presented in examples (6) and (7).

(6) Listen to the tape and write down the Chinese character you hear.

Tape: Nǐ shì shénme shíhou kāishi xué Yīngyǔ de?

Answer sheet: _____

(7) Listen to Xiao Wang's schedule for next week. Complete all of the activities she is planning to do, including the places and times you hear on the calendar by using Chinese characters you learned.

Tape: 明天早上七点半，我要到公园去散步。星期二晚上六点一刻，我要到老高家去玩儿。星期三是王朋的生日，所以晚上六点我要到他家给他过生日去。星期四晚上四点四十分，我要到图书馆去借一本书。星期五晚上七点，我和弟弟要去看电影。星期六中午十二点，我要到中国餐厅去吃饺子。

Answer sheet:

星期日	星期一	星期二	星期三	星期四	星期五	星期六
今天						

Dictations, such as (6), and graphic fill-ins, such as (7), are useful test formats for measuring listening and writing abilities. Dictation also offers a way for assessing of other general language abilities such as grammar and spelling skills. Similarly, graphic fill-ins resemble note-taking activities that students may do outside the classroom (Omaggio, 2001).

To test the students' speaking abilities, the most basic test contents should cover *pinyin*, as in (8) and (9),

and simple expressions, as in (10).

(8) Pronounce the pinyin: *wǒmen*

(9) Read the pinyin: *Wǒ māma shì lǎoshī.*

(10) Read the sentence: 我周末常常看电影。

After students are capable of forming simple sentences, we should use other forms of testing, especially limited response techniques such as interviews and series of pictures. Examples (11) and (12) demonstrate how I would make use of these two techniques. Both techniques work well with beginning-level students whose oral abilities are rather limited in that these testing activities can be controlled by teachers.

(11) Teacher: 小高, 明天是周末, 你要做什么?

Student: _____

Teacher: 你喜欢不喜欢跳舞?

Student: _____

(12) Tell me about the pictures.



For assessing beginning-level foreign students' initial Chinese reading achievements, tests are normally limited to differentiating initials, finals, and tones. A typical test format is as follows:

(13) Read the following Chinese word and select its pinyin.

Question: 英国

A. Yīngguō B. Yìngguò C. Yīngguó D. Yīngguá

After students are able to deal well with Chinese characters and their pronunciations, we can start to test students' sentence and passage comprehension respectively. True-false questions such as (14) and multiple-choice test items such as (15) are the two formats I would choose to use. These two test methods are suitable for beginning-level students in that both tests are relatively easy to follow and students can feel relaxed about them.

(14) Read the following sentence and write down „T“ if you think that the sentence is true and write down „F“ if you think that it is false.

小王：八点开会，你七点就来了。

Answer sheet: Xiao Wang thought that his friend came to the meeting too early. _____

(15) Read the following passage and select the best answer.

我们从上星期一起，就已经开始学习了。从一点半到四点四十五分，我们有四节课。上课的时候，王老师说中文说得很慢，很清楚，我们都能听懂王老师说话。复习的时候，从北京来的亲戚还来帮我们。

1. According to the passage, how many hours does the speaker spend in class?
 - A. four hours
 - B. one and a half hour
 - C. three hours and fifteen minutes
 - D. four hours and forty- five minutes

2. What does this passage imply about the speaker's feelings towards his Chinese teacher?

A. disappointed B. dislike C. bored D. impressed

Although the true-false method is useful for testing beginning students, many scholars (e.g., Gronlund, 1993; Madsen, 1983) think that it is not a good testing technique since students can easily get the right answer from guessing. However, Heaton (1990) argues that in most cases students can apply their knowledge and skills to get the right answer and guessing just happens when they feel that the questions are too difficult. In assessment of foreign students' writing in Chinese, most tests stress the accuracy of writing Chinese characters. Dictation might be the most appropriate method as all students are familiar with this format, for example (16).

(16) Please write the following sentences into Chinese characters.

Běijīng de tiānqì hěn lěng. _____

After the students have achieved a basic accuracy in writing Chinese characters, I would employ two guided writing methods to assess the students' process: sentence-writing and essay-writing. These two forms of testing are helpful not only in assessing students' specific knowledge, and application of their learned knowledge, but also in checking whether students have met the instructional objectives such as being able to organize facts and ideas by adopting the vocabulary and grammatical structures they have learned (Kubiszyn & Borich, 1996). Examples of the guided sentence-and essay-writing tests that I would use are provided in (17) and (18) respectively.

(17) Make a sentence by using the set of words given below.

我 / 出门 / 爸爸 / 提醒 / 钱

(18) Write a paragraph describing your plan to travel to China during the winter break. Base your paragraph on answers to the following questions:

1. Where are you planning to visit?
2. How many people you are planning to go with?
3. What would you do?
4. How would you travel?

One thing that we should consider in integrative testing is that the subjectivity of scoring may influence the accuracy of those tests. As I mentioned above, I have chosen to use a more integrative method to evaluate students' speaking and writing abilities; thus, for the sake of reliability, those test items which require subjective decisions need a set of scoring rubrics. One should consider what components of those abilities the rubrics will cover. For example, content, organization, and structure are commonly used as major components in scoring criteria for assessing writing; and pronunciation, intelligibility, fluency, accuracy, and appropriateness are the common components in rubrics for judging speaking performances (Davies et al, 1999).

2. Assessing the Chinese language proficiency of job applicants

For assessing the Chinese listening and speaking proficiency of the job applicants for a foreign company in China, the major concern is how to rank the applicants effectively according to their levels of proficiency in

listening and speaking and select the applicants whose proficiency level is as high as, or higher than, that described for the advanced-low level on the ACTFL scale for listening and speaking. To rank the applicants well, the test should be able to spread out the applicants. For this purpose, I shall design a norm-referenced proficiency test. Proficiency tests are often used for certifying language learners' competence or for evaluating programs. The particular proficiency tests that I have designed can help identify how well the applicants are able to communicate in Chinese.

The proficiency tests I have designed here for evaluating the job applicants' Chinese speaking and listening abilities are different from the achievement tests I described in the previous section of this paper. Whereas the achievement tests assess the achievements and progress of those students in learning Chinese as a foreign language in the classroom setting and are thus based on the pre-established criteria, the proficiency tests for the job applicants focus on discriminating the job applicants in accordance to their demonstrated listening

and speaking abilities or skills and therefore can provide information about which applicant is more proficient in listening and speaking in Chinese.

In the following, I will show how I have designed the two tests for measuring listening and speaking abilities, two of the four skills of L2 learners, these two skills are needed for actual communications in real-life situations. For the applicants who are applying for jobs in the foreign company in China, the company anticipates that their proficiency should reach at least the advanced-low level on the ACTFL scale in both listening and speaking skills.

To design the listening and speaking tests, we should bear in mind that we are using the ACTFL proficiency guidelines for evaluating both language skills. According to ACTFL listening guidelines, examinees reaching the advanced-low level should be able to:

1. Understand main ideas and most details of connected discourse on a variety of topics beyond the immediacy of the situation. Comprehension may be uneven due to a variety of linguistic and extralinguistic factors, among which topic familiarity is very prominent.

2. Able to understand listening materials covering description and narration in different time frames, such as present, non-past, habitual, or imperfective. The listening texts may include interviews, short lectures on familiar topics and news items and reports primarily dealing with factual information.
3. Aware of cohesive devices but may not be able to use them to follow the sequence of thought in an oral production.

To evaluate whether the job applicants have the abilities described in the above guidelines, the listening test I have designed covers three parts. The first part begins with listening to short exchanges and answering the corresponding questions. The second part requires the applicants to listen to short discussions or radio news items. The third part asks the applicants to listen to a variety of longer conversations or talks. I have chosen to use a multiple-choice test format for the first and third parts, while for the second part I have chosen a short-answer question format. I made this decision for several reasons. First, although multiple-choice questions are objective and easy for scoring purposes, there is a guessing factor involved. Short-

answer questions can eliminate the guessing problem but are subjective and introduce some problems for fair scoring. For this aspect, multiple-choice test items are more reliable and can be scored more consistently. Second, short-answer questions can provide more detailed information about the applicants listening comprehension such as how well they can interpret the message they hear. Third, the listening material in the second part is neither too short nor too long in comparison with the first and third parts and therefore would not be too much of a memory burden for the applicants. To better measure the applicants' proficiency, I shall try to reduce the influence of memory factors. For these three major reasons, I have included both formats in the listening test. Some samples of the specific test items are given below.

Listening comprehension (Part I)

Directions: Listen to some short exchanges and select the best answers. (1 - 4 题)

(1 - 2 题)

吴鹏: 明天的比赛蓝队肯定赢。

李明: 谁能赢还不一定呢! 比赛结果很难预测。

吴鹏: 反正我对他们很有信心。

李明: 通过上次的比赛, 红队看到自己的不

足, 做了很多改进。昨天在电视上看到红队教练胸有成竹的样子, 我觉得红队一定做好了准备。

(1) 按照李明的说法, 哪个队会赢?

- A. 蓝队很可能赢。
- B. 红队很可能赢。
- C. 两个队不可能交锋。
- D. 两个队有可能平手。

(2) ‘胸有成竹’ 是什么意思?

- A. 计划成熟
- B. 心理放松
- C. 成功在望
- D. 自吹自擂

(3 - 4 题)

小林: 我真怀念簋街。我念书时, 每周末必去。我们抽控一起去一次。

小张: 大卫的女朋友说上周她去的时候人很少。出摊数只有 6 成, 与附近的隆福寺相比差很多。

小林: 我万万没想到它会变成这样。

(3) 根据小张说的话, 我们可以知道什么?

- A. 隆福寺快要倒了。
- B. 现在簋街的状况特别冷清。
- C. 逛夜市是小林和小张的周末活动。
- D. 大卫的女朋友在隆福寺开了一个摊位。

(4) 根据他们俩的对话，我们可以做出什么推论？

- A. 最近小林很少到簋街去逛街。
- B. 隆福寺的开放时间也是每星期六晚上。
- C. 小林一直喜欢在休闲时间去逛街。
- D. 小林没想到簋街会搬到别的地方去。

The short dialogues in Part I serve to ascertain whether or not the applicants are able to grasp the main ideas and/or details of short discourses. If they fail to comprehend the dialogues (e.g., can not correctly answer main-idea questions such (1) and (4) and inference questions such as (3) in Part I), I can say that the applicants' listening ability does not reach the ACTFL advanced-low level, which requires listeners to be able to understand main ideas in informal activities related to school, home and leisure activities. Notice that Question (2) is testing whether or not a learner is able to handle cultural connotations of the listening text, which is required for the advanced-high learners. If the applicants can answer this question, we may be able to assume that their listening ability has probably reached the advanced level.

Listening comprehension (Part II)

Directions: Listen to some short news items and write your answer to each question. (5-7 题)

人的情感有很多种，有亲情，友情，爱情等等。今天想说的是自己比较在意的一种情感——那就是怀旧。虽然好象我还没到怀旧的年龄，虽然我还很年轻。每当听到一首儿时熟悉的歌曲，或回到成都时扑鼻而来的春天气息，都会让我想起妈妈，想起爸爸，想起外公外婆给我做的一顿顿好吃的饭。这两天不时听见一些小时候经常听见的歌，《站台》，《映山红》，《故乡的云》，《沂蒙颂》。都让我回忆起自己儿时的岁月。音乐的力量就是无穷啊，它可以带你穿越时空，回到从前，飞向未来。母亲节快要到了，祝我的妈妈节日快乐！也祝普天下所有的母亲节日快乐！我把前几天和霍克先生在博鳌一起演唱的一首老歌《ONE DAY WHEN WE WERE YOUNG》的歌词献给所有的母亲，祝她们永远年轻，快乐！(Zhang, 2006)

(5) 说话人是干什么的？ _____

中央社孟买十四日专电，印度中部一所大学最近被爆料，校方竟然雇用十岁到十二岁的儿童改大学生的考试卷，原因是他们的工资不多。校方以每张试卷两卢比（约新台币一元四角）的代价雇用这些孩子，而这些试卷来自大学企管及计算机科系。警方表示，当他们接获线报后突击改试卷的地点时

，发现将近一万份的试卷等着批改。根据这些评卷儿童表示，他们给分的标准有二，一是答案的长短，二是字体的美丑。想当然耳答案长、字迹美的考卷得分最高，至于答案对错，就不是这些评卷儿童关心的重点了。(Wanwei duzhe, 2006)

(6) 根据新闻内容“这些孩子”指的是什么？_____

(7) 孩子们根据什么来给那些作文打分？_____

Questions (5) to (7) will most likely solicit correct answers from applicants whose proficiency has reached the advanced-low level because these two questions require listeners to understand the details of the narration and description. Question (6) is designed to test the applicants' awareness of cohesive devices that applicants reaching this level should have although they may not be able to utilize them to follow the order of thought in an oral text.

Listening comprehension (Part III)

Directions: Listen to an interview and select the best answer to each question. (8-9 题)

主持人:各位听众好,欢迎收听,实说实话',我是主持

人芳芳。今天我们要谈的话题啊可以说非常地小,有多小呢?只有两寸多长,我们也可以这个话题非常地大,因为呢它牵扯到中国的工农兵学商各个行业,而且和中国的三亿人有切身的联系,那就是吸烟。我们现在有请今天的嘉宾,郭天赐,心理学的研究员和一位有名的电影演员陈国安来跟我们谈谈。那么我们先从心理学的研究员谈起。(对郭天赐)您平时都是研究别人的,今天能不能当着大家,研究研究自己。你吸烟吗?

郭天赐:我这个吸烟很怪,因为我这个吸烟呢从开始的时候,从那个开始抽烟,就跟人家不一样,我这个抽烟呢是饿的。

主持人:饿的?(笑声)因为饿抽烟。

郭天赐:60年那时候在大学读书呢,下乡劳动,吃不饱肚子,晚上睡不着觉,我们一个房东老大爷,特别好心,就端了个大笸箩,里面装的是烟叶,他说:“你抽一口啊,你就不饿啦!”(笑声)

陈国安:是,那时候对这个还有一个名称叫“云彩宝”。(笑声)

郭天赐:你抽了一口之后,他就果然不饿了,只恶心光想吐。(笑声)打那以后,每顿饭都吃不饱,所以每顿饭晚了以后,要饭后一支烟,就这样抽上了,到现在也没戒掉。

主持人:因为饿,你当了神仙。(笑声)那么陈国安,我想问您,您也是因为饿才抽上烟的吗?

陈国安: 我不是, 我不是, 我是人勾引我抽烟 (笑声)
主持人: 谁啊?

陈国安: 因为那会儿呢参加工作以后年级比我大的人, 他们就抽, 抽完了以后问我: “来一根? ”“不会, 不会”, 说“你瞧你, 你, 不抽你把他扔喽!”后来我就尝了。尝了以后也没什么感觉, 后来.....就跟人要啊, (笑声) 他就给我。

主持人: 跟人要的时候有什么样的感觉?

陈国安: 要的时候, 理直气壮的。 (笑声) 因为你教给我抽的, 你不给我行吗? 所以当时我就跟他要, 他就给, 后来他也负摊不起了, 我就自己买一盒烟, 大约能抽那么一个星期, 这一盒烟能抽一个星期, 后来慢慢改成四天, 慢慢改成三天, 现在基本上一天不到一盒烟。

主持人: 你希望不希望自己的后代抽烟呢?

陈国安: 看将来的社会环境了。从我们的现在的小孩子开始说起, 绝对不可以吸烟, 到那个时候我们中国没有种烟的, 没有做烟的, 没有卖烟的, 没有吸烟的。 (掌声)

主持人: 郭先生您希望您的后代吸烟吗? 允许他们抽烟吗?

郭天赐: 那时他们的人权问题了。我觉得在国际上最丢人的还不是抽烟, 是吐痰。随地吐痰, 太恶心了!

陈国安: 其实他们有人说, 吐痰哪, 那个被动吸痰比被动吸烟厉害。 (笑声)

主持人: 下次我们该重点讨论被动吸痰问题。真可

惜！时间已经到了。谢谢你们的参加。希望大家经常收听我们的节目。再见！(Adopted from Wang, 2005)

(8) 他们谈论的核心问题是什么？

- A. 吸烟的原因和被动吸痰问题
- B. 吸烟的特殊原因和对后代抽烟的看法
- C. 吸烟的经历以及它对社会所造成的危害
- D. 吸烟对社会所造成的危害的范围和吐痰的问题

(9) 主持人用什么样的口气来反应郭先生的饭后吸烟的习惯

- A. 建议
- B. 讽刺
- C. 幽默
- D. 厌烦

Question (8) is designed to assess the applicants' understanding of the main idea of the story, and question (9) requires the applicants to grasp the details of the interview discourse, as is required for this level in the ACTFL guidelines for listening. It should be noted that question (9) is designed to assess whether the applicants are able to grasp sociocultural nuances of the listening text, an ability required for advanced-high level listeners.

Questions varying in their level of difficulty like (9) serve to better spread out the applicants.

To measure the applicants' speaking proficiency, I would adopt the ACTFL oral proficiency interview (ACTFL OPI), a standardized method for measuring examinees' functional speaking skills in a foreign language. The reason I would use the OPI testing technique is that it is not only a well-established procedure, widely employed in assessment for academic purposes such as teacher certification and job qualification, but also accredited at quite a few U.S. undergraduate institutions (NAFEA, 2000). OPI uses a face-to-face interview between an ACTFL tester and an examinee, which lasts anywhere from ten to thirty minutes.⁴ The test consists of five stages: warm-up, level checks, probes, role play, and wind-down. Warm-up is the first stage of an OPI

⁴ The time spent on the whole OPI procedure may vary due to the speaker's proficiency, the use of the probe, and the level check of the interviewer. For example, the entire interview may take only ten to fifteen minutes in testing a novice, while it can last more than thirty minutes when the interviewer alternates several times between the probes and level checks in order to confirm the interviewee's level (Yoffe, 1997).

proficiency test, designed to make the interviewee feel relaxed (Standsfield, 1992) and to make the interviewee familiar with the interviewer's way of speaking and pronunciation (Yoffe, 1997). During the level-check stage, the interviewee is allowed to demonstrate his or her ability to handle the test tasks of a particular level with accuracy and confidence. During the probe stage, the interviewer attempts to find weaknesses of the interviewee by raising the conversation to a higher level of difficulty. The role-play stage further helps the interviewer to determine the interviewee's level. During the final stage, wind-down, the interviewer brings the interviewee to the level that he or she feels comfortable and then concludes the interview. Examinees' speech will be recorded on a tape and reassessed by another certified tester. The proficiency levels of examinees are determined by comparing their performances against the criteria of the ten proficiency levels defined by ACTFL Proficiency guidelines—speaking, revised in 1999.

Before describing how I would apply the OPI procedure to the testing of applicants, I will provide a

summary of the ACTFL proficiency guidelines for evaluating advanced-low level of speaking proficiency. To reach this level, examinees should be able to:

1. Handle a variety of communication tasks with confidence and accuracy.
2. Participate in conversations in most informal situations related to home, school, activities, and leisure pursuits.
3. Demonstrate limited ability in conversing on formal matters, particularly on topics related to events of works, current, public, and personal interest.
4. Narrate and describe events using time frames (past, present, future) in a smoothly paragraph-length discourse
5. Deal effectively with a complication or an unexpected turn of events through a variety of communicative devices

For Stage one, warm-up, I would interview the applicants by asking them some questions about themselves, their studies, and their interests, as shown in the sample questions (1-4):

- (1) 请你做个自我介绍一下。
- (2) 你有空儿的时候喜欢做什么？

- (3) 你学中文学了多长时间了?
- (4) 你觉得学中文的难点在哪儿?

The four questions ask about the personal backgrounds of the applicants. Speakers at the advanced-low level should be able to handle such simple questions with ease.

During the level-check stage I would ask the applicants to describe the procedures of making a cup of good coffee. This question allows the applicants to demonstrate their ability to narrate and describe sequences of events in major time frames in paragraph-length discourses, a requirement for the ACTFL advanced-low level. Question (5) provides a sample for level checks.

- (5) 请你描述怎样才能煮出一杯好咖啡?

If the applicants can deal with the above question with ease and confidence, I will go on to the third stage, probes. In this stage, I will try to identify the weaknesses of the applicants by raising the conversation to a higher level of difficulty. According to Lower, Sanders, Miles, and

Swender (1999), when speakers attempt to talk about topics of a higher level of difficulty, the quality as well as quantity of their speech will suffer.

(6) 你认为用‘不打不成材’‘棍棒出孝子’的方法来教育孩子可行与否？为什么？请你说出你的观点。

Applicants who can deal with complicated questions such as (5) and (6), which requires combining sentences into paragraphs and constructing arguments to support their opinions, have already fulfilled the requirement for the advanced-high level. However, if applicants are able to describe the procedures of making a cup of good coffee for (5) but cannot manage to answer Question (6), which requires making effective arguments, I would assume that the applicants' proficiency is below the advanced-high level in that describing the complicated events of making a cup of coffee may involve more routine activities and thus may be less difficult. Making arguments for formal topics related to current and public interests requires using communicative devices such as rephrasing and

circumlocution and may therefore be more suitable for advanced-high level applicants to handle. The proficiency level of the applicants can be double-checked during the following stage, role-play. For the role play, I would ask the applicants to play one of the roles in an interaction in real-life situations, and I would play the other role in the interaction. I would select a guided role play, in which I take the fixed role, and the participant responds. The reason that I would choose a guided role play is that the applicants can be familiarized with the requirements for the role-play assessment in advance, which ensures a fair testing. A sample role-play activity is given below.

(7) Directions: Act out the following roles:

Role A: Company A's boss (played by the interviewer)

Role B: Company B's secretary (played by the interviewee)

Situation: You are company B's secretary. You went with your boss to discuss business with another company (company A). Company A's invites your boss to a heavy drink session, but your boss is not good at drinking. Help her manage the situation.

The speaking test ends with the final stage, wind-down. During this stage I would select a question which is suitable for the proficiency level demonstrated by the applicants during the previous stages. For example, I would ask a more difficult question for those applicants who perform well during the previous stages but a less difficult one for those who perform less well. Questions (8) and (9) provide samples of difficult and easy questions respectively.

- (8) 在中国的酒文化中，如果你拒绝别人的敬酒，就是不给敬酒的人‘面子’。你如何评价中国人的‘面子’观？
- (9) 你认为中国人这种酒桌上谈生意有什么样的看法？

For scoring the applicants' speaking, I will develop a rubric which rates the applicants according to such components as content, organization, fluency, accuracy, and appropriateness of their responses. To ensure fair and consistent scoring, I would require a third rater when the interviewer and the second rater

differ in their grading.

4. Conclusion

This paper discussed two kinds of test that I would suggest utilizing in two testing situations. The first kind is criterion-referenced testing. I used it to assess Chinese language learning achievements and progress of a group of foreign students enrolled in an elementary Chinese course. This first set of test is progressive achievement tests, including unit tests and pop quizzes. In designing those tests, I took efforts to select appropriate test formats for different objectives and instructional criteria of this course. The second kind of test that I described in this paper is norm-referenced testing. In particular, these two proficiency tests are design for assessing the listening and speaking abilities of applicants who are applying for jobs in a foreign company in China. I employed the advanced-low scale of the ACTFL proficiency guidelines as the basis for selecting test formats, for creating test questions, and for creating the scoring rubrics.

References

Bai, J. (1990). Chinese proficiency test. *Journal of the Chinese language teachers association*. 25 (2), 101-103.

—. (1998). Constructing tests that reflect instructional goals. *Journal of the Chinese language teachers association*. 33(2), 71-80.

Brown, J. D. (2005). Testing in language programs: A comprehensive guide to English language assessment. New York: McGrawHill.

Daives, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge University Press

DeMary, J. L. (2004). *Proficiency levels approved by the board of education for the American council on the teaching of foreign languages (ACTFL) oral proficiency interview and the writing proficiency test. SUPTS*. Retrieved August 13, 2006, from <http://www.doe.virginia.gov/VDOE/suptsmemos/2004/infl96.html>

Gronlund, N.E. (1993). *How to make achievement tests and assessment*. (5th ed.). Boston: Allyn & acon.

Harris, D. (1969). *Testing English as a second language*. New York: McGraw-Hill.

Heaton, J. B. (1990). *Classroom testing*. Harlow: Longman

Hughes, A. (2003). *Testing for language teachers*. (2nd edition) Cambridge: Cambridge University Press.

Kubiszyn, T., & Borich, G. (1996). *Educational testing and measurement classroom application and practice*. HarperCollins College.

Lado, R. (1961). *Language testing*. New York: McGraw-

Hill

Lowe, Jr., Sanders, K.E, Miles., J & Swender, E. (1999). ACTFL proficiency guidelines-speaking revised 1999. *Foreign Language annals*. 33(1), 13- 17.

Madson, H.S.(1983). *Techniques in testing*. Oxford University Press.

Mc Donald, M.E. (2002) *Systematic assessment of learning outcomes: Developing multiple-choice exams*. Sudbury, MA: Jones and Bartlett Publishers.

NAFEA. (2000). *The American council on teaching of foreign language (ACTFL): The oral proficiency\ interview (OPI)*. Retrieved August 13, 2006, from <http://eden.clmer.csulb.edu/netshare/kclam/Apa/APA5/NAF2000/opi.htm>

Oller, J. W. (1976). *Language testing. A survey of Applied linguistics*.Ed. H.D. Brown. Ann Arbor: University of Michigan Press.

Omaggio, A. H. (2001). *Teaching language in context*. (2nd ed.).Boston, MA: Heinle & Rivers, W.

(1981). *Teaching foreign language skills*. Chicago: University of ChicagoPress.

Standsfield, C.W. 1992. *ACTFL speaking proficiency guidelines*. ERIC Digest. Retrieved August 12, 2006, from <http://www.ericdigest.org/1992-2/actfl.htm>

Valette, R. (1967). *Modern language testing*. New York: Harcourt, Brace and World.

Wang, J, Q. (2005). *Tell it like it is! Natural Chinese for advanced learners*. Yale University Press.

Wanwei duzhe.(2006). *University of Indonesia employs kids to mark exam paper* (印度大学雇用儿童改考卷).Retrieved August 13, 2006, from

[http://news.creaders.net/breaking
/newsViewer.php?id=674063&aid=13](http://news.creaders.net/breaking/newsViewer.php?id=674063&aid=13)

Yoffe, L. (1997). *An overview of ACTFL proficiency interviews:a test of speaking ability*. JALT Testing & Evaluation SIG Newsletter 1(2),3-9

Zhang, J.Y. (2006). *When we were young*. (当我年轻时). Retrieved August 13, 2006, from
<http://www.blog.sina.com.cn/u/45f2c7eb010003cc>