A Model of Formative Task-Based Language Assessment for Chinese as a Foreign Language

Chuanren Ke
University of Iowa

This article presents a model of formative task-based language assessment implemented in the Chinese language program at the University of Iowa. The model was conceptualized from both empirical evidence, namely previous research done by other team members and me, and assessment theory, namely task-based learning and testing. Our formative performance assessment tests tend to be criterion-referenced and are designed to assess how much of the content of the course or program is being learned, thereby providing links to the course objectives and program goals. Other features of our formative task-based language testing include (a) componential and analytic scoring, (b) skills integration, and (c) derivation from classroom activities. Numerous examples covering various skills are provided.

Second-language acquisition (SLA) researchers and second-language pedagogy practitioners have been paying more attention to task-based instruction. The move from structured approaches to language learning and teaching to the use of more communicative language teaching cannot be attributed to just one or a few particular individuals. The development of task-based instructional approaches, both in theory and in pedagogical practice, has been influenced by a number of concepts and approaches that have come to prominence over the past two decades: the concept of communicative competence (see, for example, Halliday, 1973; Hymes, 1972; Swain 1985), communicative language teaching (see, for example, Savignon, 1983), negotiation of meaning and interactive language teaching approaches (see, for example, Pica, 1994; Long, 1983, 1985), focus on form (see, for example, Long, 1991; Long & Robinson, 1998), cognitive approaches (see, for example, Skehan, 1998), and sociocultural approaches (see, for example, Lantolf,
2000). Recently, there has also been a call for unifying the cognitive and the social approaches in educational theory and practice (see, for example, Schoenfeld, 1999). For a comprehensive review of the history of the theoretical developments pertaining to task-based instruction see Willis (1996), Skehan (1998, 2003), and Ellis (2003).

Bygate (2001) offered a core definition of the term *task*: “A task is an activity which requires learners to use language, with emphasis on meaning, to attain an objective” (p. 11). Similarly, Ellis (2003) defined a task as one that “requires the participants to function primarily as ‘language users’ in the sense that they must employ the same kind of communicative process as those involved in real-world activities” (p. 3). In addition, as a task can be divided into subtasks and as achievement of a task or its subtasks is definable and measurable, implementation of task-based instruction is conducive to addressing the issues of accessibility and accountability (David, 1997; Lee, 1995; Skehan, 1998; Walz, 1996).

Language testers have increasingly recognized the value of tasks for assessing learners’ ability to communicate in a second language (Brindley, 1989; McNamara, 1996; Skehan, 1998). Interest in task-based language assessment can be attributed to such factors as alignment of task-based assessment with task-based instruction in the interest of achieving systemic validity (Frederiksen & Colin, 1989), positive washback effects of assessment on instruction, the limitations of discrete skills assessments (Long & Norris, 2000), the need for both structural items and communicative (functional) approaches (Widdowson, 1990), and the need for multiple sources of information for a more complete picture of students’ ability, effort, and progress (Gipps, 1994; McNamara & Deane, 1995). In addition, task-based language assessment is considered to have a favorable advantage in face and content validity and enables outcomes of the assessment to be reported in a way that is intelligible to stakeholders (Ellis, 2003).

**A BRIEF DESCRIPTION OF THE UNIVERSITY OF IOWA CHINESE LANGUAGE PROGRAM CURRICULUM**

Before we discuss our formative assessment system, a brief description on a number of major components in our program is in order. The development of our curriculum objectives and program goals are based on the American Council on the Teaching of Foreign Languages (ACTFL; 1986) model, on the Common European Framework (2001), on data from a five-year proficiency baseline obtained from a domestic summer intensive program (Ke & Reed, 1995), and on data from a 3-year baseline obtained from the regular academic year Chinese program at the University of Iowa (Ke, 2000). Task-based instructional approaches are implemented at the program (Ke & Peterka, 2002), and the curriculum aims at maximizing a bal-
anced learning program of fluency, accuracy, and complexity and at integrating listen-
ing comprehension, speaking, reading comprehension, and writing skills in a well-articulated curriculum. (For the curriculum goals of the University of Iowa Chinese language program please see Appendix A.)

Both formative and summative assessment procedures are used at the University of Iowa Chinese language program. These two types of tests are used to supplement each other for the purpose of enhancing learning and teaching and ensuring that our instructional practice is moving in the right direction. Three summative tests that are modeled on the ACTFL Proficiency Guidelines are administered at the end of each course: the Chinese Speaking Test (a simulated oral proficiency interview; Center for Applied Linguistics, 1992)\(^1\), the Computer-Adaptive Test for Reading Chinese (Yao, 1994), and the Chinese Computer-Adaptive Listening Comprehension Test (Ke & Zhang, 2002).

Formative and Summative Task-Based Language Assessment

In addition to summative task-based assessment, our program also employs formative testing. Each type has its specific purposes in our curriculum, and it is important to maintain a high awareness of these specific purposes in testing (Henning, 1990). Harlen and James (1997) characterized the purpose of these two testing procedures as follows:

Formative assessment is essentially feedback, both to the teacher and to the pupil about present understanding and skill development in order to determine the way forward. Assessment for this purpose is part of teaching. To use information about present achievements in this way means that the progression in ideas and skills must be in the teacher’s mind—and as far as possible in the pupils’—so that the next appropriate steps can be considered.

Summative assessment has a quite different purpose, which is to describe learning achieved at a certain time for the purpose of reporting to parents, other teachers, the pupils themselves and, in summative form, to other interested parties such as school governors or school boards. It has an important role in the overall educational progress of pupils but not in day-to-day teaching as does formative assessment. (p. 370)

There has been a tendency to discuss task-based assessment in its summative role, such as the ACTFL Oral Proficiency Interview (OPI) and Test of English as a Foreign Language Test of Spoken English; as a result, formative task-based assess-\(^1\)Chalhoub-Deville (2001) argued that the Oral Proficiency Interview and the Simulated Oral Proficiency Interview are based on the same view of language use and language acquisition as underlies task-based instruction.
ment that concerns curriculum matters has often been overshadowed. Weir (2001) observed

In evaluating student progress, achievement of objectives, curriculum success or school performance, the focus now appears to be summative, on the product(s), rather than formative, on the process(es) …. The central area of interest for funding agents and authorities often appears to be in the summative results and not the learning process or the learners per se. (p. 121)

The role of formative assessment in a task-based curriculum is apparent. Formative assessment can be closely linked to a task-based instructional syllabus. As a task-based syllabus is organized in stages to reflect the achievement of communicative ability, teachers need information on how well the students are mastering the instructional materials and developing their skills in relation to each stage of a course. Such formative assessment data can help teachers identify areas of strength or weakness of the students and help them make informed decisions for future teaching and learning process (Ellis, 2003; Ke, 1999; Weir, 2001). This article reports a formative task-based assessment model implemented at the University of Iowa Chinese program.

Model Conceptualization

Our formative task-based assessment model was derived from a combination of empirical evidence and assessment theory. In a study investigating the nature of progress of 222 adult learners of Chinese as a foreign language in an American university setting (Ke & Reed, 1995), the Chinese language skills of speaking, listening comprehension, grammar, and reading comprehension were measured by the ACTFL OPI and the Center for Applied Linguistics’ Chinese Proficiency Test, respectively, at the beginning and end of a 9-week period of intensive study. Only a moderate correlation between OPI scores and Chinese Proficiency Test scores was found. In other words, the oral proficiency of most students cannot be accurately determined based on results from the Chinese Proficiency Test, a test measuring the skills of listening comprehension, grammar, and reading comprehension. Further analysis of the data showed that there were different kinds of profiles within and between different modalities among the learners. A subsequent follow-up study was conducted to compare OPI scores and the scores on the teacher-developed monthly speaking achievement tests of 122 Chinese as a foreign language students (Ke, 2000). The correlations between the OPI ratings and the scores from the achievement tests were only moderate. This result once again indicated that knowing a person’s scores on a large-scale comprehensive proficiency test such as the OPI would not enable a tester or administrator to make an
accurate prediction on how well that individual would perform during the course of the learning.

Data from these two Chinese as a foreign language studies are consistent with recent interlanguage studies on a number of other languages. In her most recent comprehensive review of major interlanguage studies conducted over the past two decades on a number of languages, Han (2004) supported the modularity nature of SLA in that “it is highly probable for a given learner to achieve native-like proficiency in one or more linguistic subsystems while still falling short of such attainment in others” (p. 111).

In addition to the aforementioned empirical evidence, the arguments for our formative assessment model have also been prompted from theoretical plausibility. For instance, in his review on test taker characteristics and test performance, Kunnan (1998) pointed out that research so far has not been able to identify the specifics of the second-language construct and how those components interact. Furthermore, as the field so far cannot define the construct of individual second-language skills and as the field is yet to generate knowledge on the proficiency models available such as the ACTFL model (1986) and the Common European Framework (2001), it would seem appropriate that we adopt a system that employs multiple measures involving a variety of assessment techniques. A system that involves accumulative information obtained on a number of successive occasions over a period of time would be more advantageous than one that employs the one-time use of the large-scale standardized proficiency test. Although formative assessment may not be grounded in many of the psychometrically oriented validating and standardizing procedures and measure given the dynamic and contingent nature of classroom interaction, one may claim that the many desirable features of formative teacher assessment do contain high construct validity. Gipps (1994), for instance, contended that

The information which the teacher … obtains may be partial or fragmentary…. But repeated assessment of this sort, over a period of time, and in a range of contexts will allow the teacher to build up a solid and broad-based understanding of the pupil’s attainment. Because of these characteristics formative teacher assessment may be seen as having a high validity in relation to content and construct…. If the assessment has sampled broadly across the domain and in depth within it then the assessment is likely to be generalizable (within that domain). (pp. 123–124)

Such a multiple-measure formative assessment system when connected systematically to curriculum objectives and program goals through the establishment of a grid with horizontal and vertical dimensions covering the band of proficiency at the targeted levels to deal with the nonlinear nature of the proficiency models and for the purpose of documenting lateral progress can be argued to possess a high construct validity.
CHARACTERISTICS OF THE FORMATIVE ASSESSMENT MODEL AT THE UNIVERSITY OF IOWA CHINESE PROGRAM

Criterion-Referenced Testing

One of the major features of our formative task-based language assessment is criterion-referenced. Brown and Hudson (2002) defined criterion-referenced test as “any test that is primarily designed to describe the performances of examinees in terms of the amount that they know of a specific domain of knowledge or set of objectives” (p. 5). In other words, criterion-referenced tests are devised to measure specific sets of objectives. For this reason, criterion-referenced tests are sometimes called objectives-referenced or objective-based tests (Brown & Hudson, 2002). Formative task-based tests devised in reference to the objectives of different stages of a course or program can operationalize the objectives of a particular stage of the program.

Our formative task-based tests are grounded in course objectives tailored to the types and ability range of students. They provide feedback to the students in terms of what they accomplished in the course; and they provide feedback to the teachers on the success of their syllabus. For sample activities for our criterion-referenced formative speaking test from first- to fifth-year courses see Appendix B. To test the abilities of our first-year students to successfully perform tasks on concrete and familiar personal topics and activities, we design role-play activities for each of our first-year lessons with an eye toward the objectives. Notice that, in general, two or three situations are created based on the content of the lesson. Pairs of students, on being told about the situation, have a few minutes on their own to plan their performance before being asked to perform in front of their instructor.2

For the second-year Chinese course, in which the objectives are being able to deal with familiar topics at a discourse level, our speaking test aims to encourage the students to use various discourse connectors, sentence patterns, and vocabulary items from the lessons to perform the various speaking tasks. From the example activity for our second-year Chinese in Appendix B, one can see that to present an opinion on the topic of social inequality between men and women, students are required to use such discourse connectors in the lesson as 就拿 …来说 (taking something as an example), 实际上 (in fact), 随着… (as … [used as a conjunction]), 同样 (similarly), and 反过来 (on the other hand). By encouraging the use of the linguistic material covered in the course to show that the functional objectives have been reached encourages accuracy, promotes the development of complexity, and ensures close attention to communicative functions.

2A number of studies (Foster & Skehan, 1996, 1999; Skehan & Foster, 1997, 1999) show that planning time can affect a learner’s performance on complexity and fluency.
It also promotes positive washback effects. Knowing that linguistic expressions of the lessons covered will be vital components in the task-based achievement tests following each unit, students are more likely to exert more efforts in internalizing them.

For the third-year Chinese course, in which one of the objectives is to perform daily situations with complications, we try to devise speaking test activities that enhance students’ ability to handle such tasks. Appendix B is an example of a task for third-year Chinese in which the students were asked to summarize and evaluate essays distributed to them during their lessons. So, for example, students are asked to summarize for their instructors the contents of the two essays on population increase and housing problems. Students are also encouraged to include their own opinions. Students are told that they will be scored on fluency, accuracy, and completeness and that the more new vocabulary and sentence patterns they make use of, the better they will score. Note that for Situation 1, the instructor lets the student start talking first and then asks questions about why the meeting was so important, what kind of meeting it was, how the students would suggest that the driver—the teacher—drive faster in all the traffic, why they didn’t leave earlier if it was so important, what would happen if they were late, and so on. For Situation 2, the instructor is a concerned friend who asks questions such as why they weren’t worried about overpopulation, how they could afford to raise a lot of children, where they would live, what the chances of a happy marriage would be with such high divorce rates, and so on.

For fourth- and fifth-year Chinese courses, one of the objectives is for students to be able to deal with abstract topics, so students are frequently asked to give appropriate presentations. The example for the fourth-year course given in Appendix B deals with topics such as management and education; that for the fifth year deals with cultural and literary topics. It should be mentioned that formative tests of this nature can be made to complement the summative tests (OPI/simulated oral proficiency interview) administered at the end of course. Summative tests, such as final oral proficiency tests, involve interactive skills rather than presentation skills, so formative tests, which do include the latter, can play an important supplementary role.

**Skills Integration**

Another important characteristic of our formative task-based assessment is skills integration. In the real world, one does not usually deal with a single skill when performing a task. To make our formative task-based tests more “authentic,” we not only take the texts of our task-based activities from authentic sources, but we also make sure that they integrate the relevant modalities. Appendix C illustrates several examples of such activities. Example A involves three different skills: listening, reading, and writing. Students listen to a phone conversation regarding a
hotel reservation. Then after the students listen to the conversation, they are required to fill out a reservation form based on the content of the phone conversation. Filling out the form requires reading. For the activities in Example B, after the students listen to a conversation about sports they to respond to a number of questions orally. Example 3c involves reading and writing. Students read a personal letter and then respond to several short-answer questions to verify their reading comprehension. Arguably, short-answer questions are quite naturalistic in that they are the sort of questions that readers might really ask themselves after reading such a letter. Short-answer questions are a much better choice than multiple-choice questions in a task-based curriculum.

Componential Scoring

It is well known that in second-language learning, different domains of interlanguage may be acquired at quite different rates (see, for example, Han, 2004; Ke & Reed, 1995). Given the differential nature of SLA and the holistic nature of the summative assessment administered in our program, we adopt a componential and analytic scoring system for our formative assessment procedure.

Holistic assessment (of the sort characteristic of the ACTFL OPI and the ACTFL Writing Test) tends to assign a single score to a task performance based on an overall impression. However, such a single score does not provide useful diagnostic information about particular dimensions of a learner’s achievement, such as fluency, accuracy, complexity, and sociocultural performance. Depending on the nature of the formative assessment activity and their place in their Chinese study at the University, componential (analytic) scoring is used to assess key criteria such as pronunciation (both segmental and tones), vocabulary, grammar, fluency, ability to comprehend at a normal rate of speech, ability to comprehend without need for repetition, accuracy of response, and sociolinguistics. In addition, we vary the weight of each of these components at each instructional stage to reflect the needs of a particular student body and the focus of a particular curriculum stage. For example, at the initial stage of students’ Chinese learning, we assign more weight to pronunciation to highlight the importance of laying a solid foundation for the learning of a tonal language. For the grammatical component, students are encouraged at all instruction levels to use new vocabulary and new structures, and distinctions are made between recurring errors and restructuring ones. Attention to the sociolinguistic component increases with instructional level. Appendix D is a Chinese composition profile adapted from Jacobs, Zinkgraf, Wormuth, Hartfiel,
and Hughey’s (1981) scoring profile to reflect the nonalphabetic nature of the Chinese script and the characteristics of Chinese grammar.

Derivation From Classroom Activities

One of the advantages of formative testing is its close relation to instructional activities. Formative testing gives an immediate indication as to how well the students are mastering the instructional materials and developing their skills in relation to the curriculum goals. According to the system approach (Frederiksen & Colin, 1989), to be considered valid, the contents and formats of the tests should closely align with the instructional activities. Our task-based classroom activities—filling information gaps, doing interviews, taking roles, reporting information, and having group discussions—are, in accordance with instructional levels and curriculum objectives, frequently adopted for formative testing situations. This kind of testing procedure can be considered an extension or repetition of the relevant tasks, or posttask activities in the classroom. The alignment of formative assessment with classroom activities can be used to counter objections that students might make as to the value of such assessment.

TEACHER DEVELOPMENT IN TASK-BASED ASSESSMENT

Given the difficulty of developing and scoring task-based assessment activities, along with the fact that the bulk of our instruction and the administration of both summative and formative assessment procedures in our program are carried out by instructors and teaching assistants, it is essential to provide help to our instructional staff in the form of theory, information on research findings, and guidance and practice with these two assessment procedures so that they can make informed decisions in their efforts to enhance learning and teaching. In our program, workshops are conducted at the beginning of each academic year to train our instructional staff on various aspects of the curriculum, including task-based assessment.

Teachers need to be aware that “the kind of information that is gathered by teachers in the course of teaching is not tidy, complete and self-consistent, but fragmentary and often contradictory” (Harlen & Mames, 1997, p. 376). The field

---

4Recent research on performance testing shows that repetition of the task may give rise to changes in performance, in terms of fluency, accuracy, complexity, or a combination of these (Bygate, 2001) and that posttask activity leads to significantly greater accuracy (Skehan, 2003).

5In a recent study surveying teachers’ and students’ perceptions of formative assessment (Weir, 2001), one perceived negative value from the students’ perspectives was that such assessments “Take time away from teaching” and that it “does not contribute to learning.”
is not ready to define the construct of communicative language ability. There have been attempts to list attributes of this construct, but we have not been able to generate any knowledge on how such attributes interact in a principled manner (Kunnan, 1998). Assessment is not a science, even though it needs to be informed by relevant theory and research findings. There are many variables that may affect the learners’ performance and the teacher’s rating. Those variables include interaction between performer and situation, the linguistic and nonlinguistic behavior of the tester, and the way assessment activities are sampled. In our in-service training to our instructional staff, efforts are made to provide opportunities for our teachers to develop their confidence and expertise in making sound and responsible judgment about learners’ performance. In our in-service teacher development workshops, samples of students’ performance from both proficiency (summative) testing and task-based achievement (formative) testing, representing a wide range of performance, are used to sensitize our teachers’ awareness of the many facets in administering and scoring communicative language performance. Efforts have also been made in our teacher development workshops for the teachers to design their formative tests in a way that is conducive to collecting data on the interaction between learners’ ability and context.

CONCLUSION

This article explores the role of formative task-based language assessment in a task-based curriculum. It explains how such testing is integrated in our instructional efforts and how different features of the model contribute to the learning and teaching goals of our program, particularly in the areas of diagnosis, progress, and achievement. Such formative testing allows our teachers to tailor their teaching energies toward continuing instruction and toward providing timely feedback for developmental purposes. For future research, empirical evidence needs to be collected to see the interaction effect between this model and the other curriculum components vis-à-vis the accomplishment of the curriculum objectives and program goals as well as the washback effect of the model on the teachers.

ACKNOWLEDGMENTS

This study was supported by a Career Developmental Award from the University of Iowa for 2004–2005 and the University of Iowa Obermann Center for Advanced Studies. I am grateful to four anonymous reviewers for their constructive comments.
REFERENCES


---

**APPENDIX A**

**University of Iowa Chinese Language Program Objectives and Descriptors for the Proficiency Levels**

<table>
<thead>
<tr>
<th>Curriculum Objectives</th>
<th>Listening</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Year Chinese</td>
<td>Intermediate High</td>
<td>Intermediate Mid</td>
<td>Intermediate Mid</td>
<td>Intermediate Low</td>
</tr>
<tr>
<td>First-Year Chinese for Heritage Learners</td>
<td>Intermediate High</td>
<td>Intermediate High</td>
<td>Intermediate Mid</td>
<td>Intermediate Low</td>
</tr>
<tr>
<td>Second-Year Chinese</td>
<td>Advanced High</td>
<td>Intermediate High</td>
<td>Intermediate High</td>
<td>Intermediate Mid</td>
</tr>
<tr>
<td>Third-Year Chinese</td>
<td>Advanced High</td>
<td>Advanced Advanced</td>
<td>Advanced Intermediate</td>
<td>Intermediate High</td>
</tr>
<tr>
<td>Fourth-Year Chinese</td>
<td>Advanced High</td>
<td>Advanced High</td>
<td>Advanced High</td>
<td>Advanced High</td>
</tr>
</tbody>
</table>

**Descriptors for Each of the Proficiency Levels**

**Listening**
- **Intermediate**: Able to understand sentence-length utterances on a variety of topics such as basic personal background and needs, social conventions, lodging, transportation, shopping, personal interests and activities, and a greater diversity of instructions and directions.
- **Advanced**: Able to understand main ideas and most details of connected discourse on a variety of topics beyond the immediacy of the situation. These texts frequently involve description, narration interviews, short lectures on familiar topics, and news items and reports primarily dealing with factual information.
- **Superior**: Able to understand the main ideas of all speech in standard Chinese, including technical discussion in a field of specialization. Can follow the essentials of extended discourse in academic/professional settings, in lectures, speeches, and reports. Able to make inferences within the cultural framework of the target language.
APPENDIX A Continued)

**Speaking**

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>Able to handle successfully a variety of uncomplicated, basic, and communicative tasks and social situations. Can talk about self and family members. Can ask and answer questions and participate in simple conversations on topics beyond the most immediate needs (e.g., personal history and leisure time activities).</td>
</tr>
<tr>
<td>Advanced</td>
<td>Able to satisfy the requirements of everyday situations and routine school and work requirements. Can handle complicated tasks and social situations, such as elaborating, complaining, and apologizing. Can narrate, describe, communicate facts, and talk casually about topics of current public and personal interest.</td>
</tr>
<tr>
<td>Superior</td>
<td>Able to speak Chinese with sufficient accuracy to participate effectively in most formal and informal conversations on practical, social, professional, and abstract topics. Can discuss special fields of competence and interest with ease. Can support opinions and hypothesize and command a wide variety of interactive and discourse strategies.</td>
</tr>
</tbody>
</table>

**Reading**

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>Able to read simple, connected texts dealing with a variety of basic and social needs. Texts impart basic information about which the reader has to make minimal suppositions and to which the reader brings personal interest and/or knowledge. Examples may include short, straightforward descriptions of persons, places, and things written for a wide audience.</td>
</tr>
<tr>
<td>Advanced</td>
<td>Able to read prose of several paragraphs in length. The prose is predominantly in familiar sentence patterns. Texts at this level include descriptions and narrations such as simple short stories, news items, bibliographical information, social notices, personal correspondence, routinized business letters, and simple technical material written for the general reader.</td>
</tr>
<tr>
<td>Superior</td>
<td>Able to read expository prose on unfamiliar subjects and a variety of literary texts. Superior-level texts feature hypotheses, argumentation, and supported opinions and include grammatical patterns and vocabulary ordinarily encountered in academic/professional reading.</td>
</tr>
</tbody>
</table>

**Writing**

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>Able to meet a number of practical writing needs. Can write short, simple letters. Content involves personal preferences, daily routine, everyday events, and other topics grounded in personal experience.</td>
</tr>
<tr>
<td>Advanced</td>
<td>Able to write routine social correspondence and join sentences in simple discourse of at least several paragraphs in length on familiar topics. Can write simple social correspondence, take notes, write cohesive summaries and resumes, as well as narratives and descriptions of a factual nature.</td>
</tr>
<tr>
<td>Superior</td>
<td>Able to express self-effectively in most formal and informal writing on practical, social, and professional topics. Can write most types of correspondence, such as memos as well as social and business letters, and short research papers and statements of position in areas of special interest or in special fields.</td>
</tr>
</tbody>
</table>
APPENDIX B
Sample Criterion-Referenced Speaking Test Activities
for First-to-Fifth Year Courses in the University of Iowa
Chinese Program

First-Year Chinese

Lesson 5 Visiting Friends

Situation 1
Your friend comes to visit you on campus and you meet at IMU. Introduce him/her to your classmates.
He/she tries to find out where you and your classmates study, work, and do other activities (listen to music,
play ball, dance, watch movies, etc.). Offer your friend something to drink.

Situation 2
You are talking with a classmate’s brother/sister for the first time. Find out if he/she is a student, where
he/she studies/works, and what his/her hobbies are. You also talk about each other’s daily routines, such as
going to school or work, and returning home.

Situation 3
You are visiting a friend in his or her room. Compliment him or her on the room. Your friend offers you
something to drink, but you just want a glass of water. He/she introduces you to his/her roommates (two). You
find out where they work/study and their hobbies.

Second-year Chinese

Lesson 15: Inequality between man and woman

Please fully expound on the following topic (At least 12 sentences are required).
You have 3 minutes for preparation, and your presentation time will be 3 minutes.

Topic:
It has been widely recognized that there still exists serious inequality between men and women around the
world. Please find sufficient details to support this argument, and seek feasible ways to solve the problems
you’ve raised.

Suggested lists (At least 10 of them must be used)
就拿 ... 来说 实际上 随着 ... 反过来 往往 主要 大概 不管 明显
社会 现象 担子 家务 同样 重男轻女 甚至 体现 不断 自由
出现 办法 解决 乐观 职位 得到提高

Third-year Chinese

Lesson 1: Population increase and housing problems

A situation to be performed with your instructor; you will be asked to assume a character in a situation in
which there is a desired goal. The best way to prepare for such a situation is to be extremely familiar with
vocabulary and sentence patterns (not just the new, but also the most appropriate). You will be scored on
how well you manage the situation (in terms of cultural appropriateness) as well as fluency and accuracy.

Situation 1
You are in Shanghai and in a hurry to get to your next important meeting. It is too far to walk and you just
missed the bus, so you jump into a cab. Try to convince the cab driver (your instructor) that s/he should
find a way to get you to your meeting on time. (Bribery will not work.)

Situation #2
You want to have many children and when you bring this up to your friend (your instructor) s/he is not
crazy about the idea. Try to convince your friend that having a big family is a great idea.
Fourth-year Chinese

Monthly Speaking Test 2
考试包括两部分：
1) 学生就一个题目，做一个演讲。使用 Power Point，作为辅助工具。
2) 根据演讲内容，提问，演讲者回答。

从下列题目中选择一个，准备演讲：
1) 谈谈美国大学的高等教育方针和方式，并加以评论。
2) 比较中美教育。
3) 假如你是爱荷华大学校长，你会怎样管理学校，改革教育体制？
4) 婴儿潮父母的成长经历怎样影响他们对孩子的教育。
5) 孩子成长过程中，应允许他们有实验时期，自己去尝试，去选择，而不是父母帮他们选择信息，向他们灌输人生观。你同意吗？
6) 谈谈美国高中教育。
7) “活着”第十五—二十五幕有关吃饭问题的描述，怎样反映老百姓和干部对政治的理解和对生活的要求？

这些题目都比较宽泛，你可以就一个问题目的的某一方面加以深入、透彻的阐述。

Fifth-year Chinese

Monthly Speaking Test 1
从下列题目中选一个，做篇演讲：
1) “婚姻制度本是社会一个寒暑表，一个文化的最反映的镜子。”谈谈你对这一观点的看法。
2) 谈谈乡土社会的风俗在现代社会中的流变。
3) 老舍的“黑白李”描述了各种各样的爱：男女之爱、兄弟之爱、对民众之爱……在这些关系之中，个人感情会与道德修养或政治追求产生矛盾，就个人感情、理想与社会政治准则、道德习惯之间的冲突，谈谈你的见解及感受。

方法及步骤：
1) 这些题目仅仅泛泛地规定了演讲的范围，你可以决定自己的副标题，就这一范围中的某一方面有理有据地加以深入、透彻地阐述。
2) 每组演讲大约十五分钟，另有十分钟回答观众有关演讲内容的问题。
3) Power Point 的磁盘星期一交给老师。Power Point 要有助于讲解，而不能喧宾夺主。

222
APPENDIX C
Samples of Skills-Integration Formative Task-Based Tests
in the University of Iowa Chinese Program

Example 3a: Integration of listening, reading, and writing skills.
Listen to the following phone conversation and fill out the reservation form accordingly.

Phone conversation:

女：您好，这里是五洲宾馆。
男：您好，我想订两个房间。
女：您稍等，好，请问您什么时候入住？
男：我们五月二十八号到，六月五号离开。
女：您是乘火车还是坐飞机？需要安排接送吗？
男：我们坐飞机往返。不用接送了，谢谢。不过请你告诉我宾馆的地址好吗？
女：噢，我们在深圳市福田区深南大道 6001 号。
男：好，谢谢。
女：请问您想要什么样的房间？我们有标准房，高级房，普通套房，和豪华套房。
男：标准房和高级房有什么区别？
女：标准房里有两张单人床，高级房间里有一张宽大的双人床。其他配备差不多。
男：那好，我要一间高级房，一间标准房。
女：高级房一天 1389 元，标准房一天 1260 元。
男：好。
女：请问您的姓名？
男：我叫高杰，杰出的杰。
女：请问有几位客人？
男：一共三位。
女：您的信用卡账号？
男：中国恒通，1348-4926-5701。
女：截止日期？
男：2004 年 12 月 31 号。
女：好，谢谢。您的联系电话？
男：我的手机号是 138-7653-9011。
女：如果需要取消订房，请在五月二十八号六点以前通知我们。
男：好。
女：谢谢您惠顾五洲宾馆。
男：不客气，再见。

五洲宾馆预订单(The reservation form):

<table>
<thead>
<tr>
<th>客户基本信息</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>入住日期</td>
<td>2003 年 月 日</td>
</tr>
<tr>
<td>离店日期</td>
<td>2003 年 月 日</td>
</tr>
<tr>
<td>入住人数</td>
<td></td>
</tr>
<tr>
<td>到达方式</td>
<td></td>
</tr>
<tr>
<td>房型</td>
<td>间 数</td>
</tr>
<tr>
<td>标准房</td>
<td>价格</td>
</tr>
<tr>
<td>高级房</td>
<td></td>
</tr>
<tr>
<td>普通套房</td>
<td></td>
</tr>
<tr>
<td>豪华套房</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B (Continued)

<table>
<thead>
<tr>
<th>联系人姓名</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>信用卡名称</td>
<td>账号</td>
<td>截止日期</td>
</tr>
<tr>
<td>电话</td>
<td>手机</td>
<td></td>
</tr>
<tr>
<td>备注</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example 3b: Example from a listening test (integration of listening and speaking skills): First year Chinese –Lesson Twenty-one 运动**

Listen to the following conversations and answer questions.

**Conversation one:**

男：你来美国做什么？
女：我来美国上大学。
男：为什么你不在台湾上大学呢？
女：台湾的大学很难上，二十一 个人里只有一个人考得上大学。

问题:
1) 女的是什么地方人？
2) 她为什么来美国上大学？

**Conversation two:**

男：运动对身体真好，每天运动，你就不容易生病。
女：有时候运动对身体也不好。
男：为什么？
女：你要是跑得太多，你很容易累，你的脚也容易受伤。
男：可是运动总比不运动好。

问题  : 
1) 女的为什么觉得运动好？
2) 男的为什么觉得运动太多不好？

**Example 3c: Example from a fourth-year reading test (integration of reading and writing skills with a short-answer test component):**

杨元：
最近好吗？
你上次说要邀请我去北京玩一次，这个邀请让我高兴了好几天，我甚至工作的时候都觉得精神百倍，经常一边唱歌，一边做那些平常让我既疲倦又无聊的工作。我的同事和学生都觉得我有一些“不正常”，可是当我告诉他们我为什么这么高兴时，他们都能理解我了。

可是令人遗憾的是，我不可能今年暑假就去北京，尽管你在信里一再表示希望我能尽快去，可以参加你新书（已经是你职业生涯的第六部长篇小说了！）出版的发布会，而我也很想到时能陪陪你，并且分享你的喜悦，但是我实在没有时间。我这里的科研和教学任务都很重。目前，美国学生对中国文化的兴趣日益浓厚，越来越多的人选修中文课程与中国文化有关的课程，包括中国诗歌、现代文学和中国民间民俗等。

所以，我这个学期除了要教三个年级（每个年级三个班，每个班十个学生）的中文课
之外，还要担任中国文化课的主讲，教学任务非常繁重。同时，国际中国文化研讨会
和中文教学会议都要在下半年举行，我必须参加这两个会议。所以必须在七月份之前完成
一篇有见论文，这样我有资格申请参加两次会议的“入场券”，为此。我必须每天从
时间来阅读大量的学术文章，分析教学研究中得到的一些数据，桌子上的资料堆积如山，
非得经过一场苦战才能完成这些工作。我实在是身乏力复啊！

所以，尽管我对你的邀请很高兴，但也不得不在今年年底才能考虑去北京，我也许明
年秋天（10 月份左右），红叶已经红遍了香山的时候，我会到北京去，到时候，我们一起
在枫树下畅饮几杯，此致

祝一切顺利

思远

2002 年五月七日于美国爱荷华大学家中

问题：
1. 写信的人应该是从事什么工作的？
2. 收信人和写信人的关系是什么？
3. 收信人现在在什么地方？
4. 根据信的内容，收信人可能是从事什么工作的？
5. 写信的人的同事为什么说他不正常？
6. 收信的人为什么说自己很忙（你回答的不能超过 15 个字）
7. 收信的人说他要参加两个会议，根据信的内容，你觉得他们时间
   是什么？


8. 信中有几个很特别的词语，请写出它们在这封信里的意思。（注意要简短回答）
   分身乏术     苦战     堆积如山     入场券

9. 根据这封信提供的信息，你觉得写信的人哪年哪月可能会到北京？
Appendix D
Chinese Composition Profile
(Adapted from Jacobs et al.’s
(1981) scoring profile)

<table>
<thead>
<tr>
<th>Categories</th>
<th>Score for each category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>30</td>
</tr>
<tr>
<td>Organization</td>
<td>20</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>20</td>
</tr>
<tr>
<td>Language use</td>
<td>25</td>
</tr>
<tr>
<td>Mechanics</td>
<td>5</td>
</tr>
</tbody>
</table>

Score level criteria – Content

30–27 EXCELLENT TO VERY GOOD: knowledgeable substantive thorough development of thesis relevant to assigned topic

26–22 GOOD TO AVERAGE: some knowledge of subject adequate range limited development of thesis mostly relevant to topic, but lacks detail

21–17 FAIR TO POOR: limited knowledge of subject little substance inadequate development of topic

16–13 VERY POOR: does not show knowledge of subject non-substantive not pertinent OR not enough to evaluate

Score level criteria – Organization

20–18 EXCELLENT TO VERY GOOD: fluent expression ideas clearly stated/supported succinct well-organized logical sequencing cohesive

17–14 GOOD TO AVERAGE: somewhat choppy loosely organized but main ideas stand out limited support logical but incomplete sequencing

13–10 FAIR TO POOR: non-fluent ideas confused or disconnected lacks logical sequencing and development

9–7 VERY POOR: does not communicate no organization OR not enough to evaluate

Score level criteria – Vocabulary

20–18 EXCELLENT TO VERY GOOD: sophisticated range effective word/idiom choice and usage appropriate choice of words

17–14 GOOD TO AVERAGE: adequate range occasional errors of word/idiom choice, usage but meaning not obscured
13–10 FAIR TO POOR: limited range frequent errors of word/idiom form, choice, usage meaning confused or obscured
9–7 VERY POOR: essentially translation little knowledge of Chinese vocabulary, idioms, word form OR not enough to evaluate

Score level criteria – Language use

20–18 EXCELLENT TO VERY GOOD: effective complex constructions few errors of cohesive devices, aspect markers, word order
21–18 GOOD TO AVERAGE: effective but simple constructions minor problems in complex constructions several errors of cohesive devises, aspect markers, word order but meaning seldom obscured
17–11 FAIR TO POOR: major problems in simple/complex constructions frequent errors of cohesive devises, aspect markers, word order and/or fragments, deletions meaning confused or obscured
10–5 VERY POOR: virtually no mastery of sentence construction rules dominated by errors does not communicate OR not enough to evaluate

Score level criteria – Mechanics

5 EXCELLENT TO VERY GOOD: demonstrates mastery of conventions few errors of spelling, punctuation, paragraphing
4 GOOD TO AVERAGE: occasional errors of spelling, punctuation, paragraphing but meaning not obscured
3 FAIR TO POOR: frequent errors of spelling, punctuation, paragraphing poor handwriting meaning confused or obscured
2 VERY POOR: no mastery of conventions dominated by errors of spelling, punctuation, paragraphing handwriting illegible OR not enough to evaluate