

**ON MEASURING COMMUNICATIVE COMPETENCE:
A CASE STUDY ON THE CONSTRUCT VALIDITY OF
THE ASSESSMENT PRACTICE OF AN EFL PROGRAM**

Theresa Jiin-ling Tseng

ABSTRACT

The construct of communicative competence has been greatly emphasized in second language teaching at different levels of education. The purpose of this study is to examine the construct validity of the assessment practice of a Freshman EFL for Non-Majors (FENM) program that stresses communicative competence. The methods of this study include an inspection of both the assessment practice recommended in the *Teachers' Handbook, Freshman English for Non-Majors 2002-2003* ("FENM Teachers' Handook," 2002) and samples of the FENM program-wide exams, and interviews with three experienced teachers on their assessment practice. The findings indicate that the test format (multiple-choice questions) used in the program-wide exams of reading and listening does not match the characteristics of communicative language testing. Consequently, students' reading and listening skills that are measured do not permit teachers to predict the students' ability to use the skills to negotiate with others successfully in a naturalistic situation. The oral assessments often times do not allow authentic interaction to take place since most teachers let students write and memorize the scripts before the oral assessment. As a result, it is difficult to measure an important attribute of communicative competence—students' ability to process unpredictable data in real time. Overall, the construct validity of the assessment program appears to be low. The results of this study may spur teachers of CC-oriented programs to reflect upon whether their test practice measures what they intend to measure, and what can be done to increase construct validity of their tests. Recommendations on using authentic tasks and guidelines to increase construct validity of communicative language testing are provided at the end of this article.

Key Words: communicative competence, communicative language testing, construct validity, language assessment, authentic tasks, measuring communicative competence

INTRODUCTION

In language testing, Messick (1975) has defined *construct* as the nature of what the examinee displays or possesses. Bachman (1990) has treated *construct* as *a theory of abilities* that permits us to state specific hypotheses about the relationship between how the abilities we observed are or are not related to other abilities. *Construct validity* “concerns the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs” (Bachman, 1990, pp. 254-255). Construct validity has been recognized as central to the appropriate interpretation of test results, and it also has been considered to be the most important aspect in the evaluation of tests themselves in recent years (Bachman, 1990; Cumming, 1996; Young, 2002).

The construct of *communicative competence* has been applied to language teaching since the 1970s, and specifications about the testing of communicative performance have been available since the 1980s (e.g., Bachman, 1990; Carroll, 1980; Norris, Brown, Hudson, & Yoshioka, 1998). The major argument for this communication-oriented approach to language proficiency assessment has to do with test construct validity: Whether the nature of the attributes we measure permits us to predict other abilities—the knowledge (competence) of the language and the capacity for implementing the competence in specific contexts of use (Bachman, 1990; Canale, 1984). Samuel Messick (1989) has argued that testing is a procedure for drawing inferences about the unobservable, and it is inevitably indirect and uncertain. Thus, it is crucial to select testing methods that can best help test givers collect evidence on which they could rely to draw inferences about the underlying communicative competence the test taker has.

This case study examines the language assessment methods of a university Freshman English for Non-Majors (FENM) program in order to find out if the assessment practice measures what it intends to measure—students’ communicative competence. Like the FENM program under investigation, many English as a foreign language (EFL) programs at different levels of education have claimed that their goal is to build up students’ communicative competence. Thus, the findings of this study are by no means unique to the program being examined but significant to all the EFL programs that aim at building up and measuring students’ communicative competence. The results of this study may spur teachers or test designers to rethink if their assessment instruments can truly measure the learner’s communicative competence they intend to

measure. This study may also inform teachers and test designers who intend to test communicative competence what can be done to increase construct validity of their tests.

In what follows, I will first review the literature pertaining to the construct of communicative competence and the characteristics of communicative language testing. Then I will use the specified characteristics of communicative language testing to examine the assessment instruments used by the FENM program. The examination is followed by a discussion and a conclusion with implications.

LITERATURE REVIEW

Communicative Competence

The term *communicative competence* was coined by Hymes (1972a) to expand Chomsky's notion of *competence*, or the internalized linguistic knowledge that native speakers have. This competence enables native speakers to produce grammatically correct sentences in their language. For Hymes, Chomsky's notion of competence was too limited. Hymes (1972b) has asserted that there are *rules of use* without which the *rules of grammar* (i.e., linguistic knowledge) will be useless. He believes that for learners to be communicatively competent, they need both the knowledge about the language *and* the ability to *use* the language in context. They need to know when to speak, when not to, and what to talk about, with whom, where and in what manner (Hymes, 1972a). Hymes' view of communicative competence was complemented by Halliday's (1970) theory of the function of language, which has influenced many writers on communicative language teaching (e.g., Brumfit & Johnson, 1979; Savignon, 1983, as indicated by Richards & Rogers, 2001).

A pedagogically influential analysis of communicative competence is found in Canale and Swain (1980). They identified several components of communicative competence: (1) *grammatical competence*, which includes morphological, syntactical, and lexical capacity; (2) *sociolinguistic competence*, which refers to an understanding of social context in which communication takes place, including sociocultural appropriateness (e.g., role relationships) and discourse rules (e.g., shared information, purpose of communication); (3) *strategic competence*, which comprises coping strategies that communicators employ to initiate, terminate, maintain, repair, and redirect communication; (4) *discourse competence*, which

refers to the interpretation of individual message elements in terms of their interconnectedness and of how meaning is represented in relationship to the entire discourse or text (Richards & Rogers, 1988). From the perspective of language testing, Bachman (1990) has specified that communicative language ability consists of both knowledge and the capacity for implementing that competence (knowledge) in specific contexts. Among the four components, the first one, grammatical competence, is related to *usage* or knowledge of grammar, and the remaining three (sociolinguistic, strategic, and discourse) are related to *use*. While the evaluation of usage focuses on accuracy, the evaluation of use focuses on appropriateness. In short, writers on communicative competence have agreed that (1) communicative competence includes *both* knowledge about the language *and* ability to *use* the language, and (2) the appropriateness of use, or how to implement the competence in specific contexts, is to be stressed in language teaching.

Characteristics of Communicative Language Testing

The development of tests for assessing communicative performance has been greatly influenced by ideas about the role of language in communication: the *use* of language is the objective (Carroll, 1980). The mastery of formal patterns, or *usage*, is only a means to achieve that objective. The ultimate criterion of language mastery is therefore the learners' effectiveness in communication for the settings they find themselves in (Carroll, 1980).

Wesche (1983) has argued that meaning does not exist ready-made in the linguistic code, but is rather a function of the relationship between language forms, functions, and context, including the intentions of the speaker and expectations of the hearer. Therefore, in a communicative second-language test, as Wesche (1983) has pointed out, the objectives should be expressed in terms of what the examinee will be able to "do" in the target language in a naturalistic situation; i.e., whether he or she will be able to use the language effectively for a given communicative purpose. Wesche (1983) has specified some characteristics of communicative language tests:

1. They are intended to tap communicative competence in addition to grammatical competence. For tests to be valid, they must activate the internalized rule system by which discourse is meaningfully processed. Such tests should be *integrative*,

pragmatic tests, involving the use of naturalistic language in a verbal and situational context.

2. They are intended to ascertain whether the learner can “do” something in the second language with an acceptable degree of efficiency (e.g., speed, correctness, appropriateness to the situation). The rationale is that competence, the internalized grammar of a language, cannot be directly observed. Therefore, we can only rely on “performance,” or imperfect realization in concrete situations, to reveal the hidden competence.
3. They are intended to test a range of situations which reflect course objectives, i.e., to test the examinee’s manipulation of a variety of language functions.
4. Criterion referencing is more appropriate than norm referencing in such tests. That is, the performance of each examinee is compared with a definition of adequate performance on the task but not with the performance of other examinees.
5. The tests are intended to be reliable. This test characteristic is particularly problematic with respect to scoring criteria and procedures. High levels of inter- and intra-rater reliability can be achieved through the careful training of raters and long experience with a particular test format and scoring grid.

Wesche (1983) has indicated that it is difficult to design a test that has all of the above desirable characteristics all of the time; however, by taking into consideration these characteristics, we can at least strive to make our tests better measures of communicative competence.

Similarly, Morrow (1979) has contended that the purpose of communicative language testing is to test students’ ability to actually use the language. Morrow (1979) has provided a list of features that characterize language use in communicative language testing that do not seem to be measured in conventional tests:

1. *Interaction-based*: Language in use is based on an interaction. The interaction process not only involves the modification of expression and content but also a combination of receptive and productive skills. What is said *by* a speaker depends crucially on what is said *to* him.

Theresa Jiin-ling Tseng

2. *Unpredictability*: Processing of unpredictable data in real time is a vital aspect of using language.
3. *Context*: Language forms which are appropriate will vary in accordance with its context in terms of context of situation (e.g., physical environment, role, formality) and linguistic context (e.g., textual cohesion).
4. *Purpose*: Every utterance is made for a purpose.
5. *Authenticity*: All tasks undertaken should be real-life, interactive communicative operation, or day-to-day discourse, not edited in the interest of simplification.
6. *Behavior-based*: A test of communication should measure what the candidate can actually achieve through language (e.g., to be able to use the target language to talk with a postal clerk and get a package sent).

In summary, what characterizes communicative language testing is that it is integrative in its testing of language skills, and it is behavior-based, interaction-based, unpredictable in nature, context-dependent, purposeful, involving completing authentic tasks, and criterion referencing. These are also the criteria I use in this study to examine the practice of language assessment.

METHODS

Two methods were used to examine the construct validity of the assessment instrument of the FENM Program: document inspection and interviews. The documents included the descriptions and samples of the content, formats, and criteria of the program-wide exams and program-wide assignments as provided in the FENM Teachers' Handbook (2002). Qualitative interviews were conducted to find out teachers' practice in assessing students' communicative abilities.

Classroom observation of assessments was not applied in this study because the appearance of a stranger during assessments might cause students to feel uncomfortable or nervous, which would not help reveal students' authentic competence.

The participants/interviewees of this study were three full-time experienced FENM teachers Z, C, and J at a private university in Central

Taiwan. All of them had obtained their advanced degrees in the U.S.: C and J have M.A. degrees in TESOL and Z has a Ph.D. degree in foreign language education. At the time of the interviews, Z was also the department chair, and all three had been teaching FENM for over ten years. They all had been exposed to the concept of communicative competence during their studies, and they were aware of the course requirements. In the interviews, each of them answered two main questions and some follow-up questions. The two main questions were:

1. In the teacher-designed parts (mid-term and final communicative components and in-class oral assignments), what are the content and format of your assessment instruments?
2. What are the criteria you use to judge students' communicative ability?

To examine the construct validity of the assessment instruments, I asked if the assessment instruments corresponded to the following characteristics of communicative language testing: (1) interaction-based (negotiation or continual modification in response to change) (2) involving unpredictability (spontaneous, unrehearsed response), (3) context-sensitive (e.g., to environment, role, text, sociocultural knowledge, real-world knowledge), (4) purpose-oriented in tasks, (5) authentic (day-to-day discourse), (6) behavior-based (whether the learner can "do" something in the target language), (7) integrative in language skills, (8) pragmatic in language use, and (9) appropriate (based on effectiveness and adequacy in performance).

RESULTS

In this section, I present the findings from an examination of program-wide (mandatory) assessments and each teacher's in-class practice to assess students' communicative competence.

The assessment of FENM consists of two parts: The program-wide common exams (mid-term and final, together 50%) and in-class assessment (50%). The passing grade is 56 on a 100-point scale for the exam part (average of mid-term and final) and 60 for the in-class part. Students have to have a combined (exam and in-class) average of 60 to pass the course. The average of the midterm and final for all classes is

Theresa Jiin-ling Tseng

adjusted to a norm of 75.

According to the FENM Teachers' Handbook (2002), the goal of the program is "to build up students' communicative competence." However, what constitutes "communicative competence" is not specified in the Handbook. In other words, it is assumed that the construct of communicative competence is clear to all FENM instructors. One of the participants, Z, who was also the department chair, stated, "We trust our teachers' judgment on students' proficiency levels." Z expressed that "experienced teachers have internalized what to watch for and how well the students should perform," although he did not mention what criteria those less experienced teachers should follow.

I now turn to report how the four skills are evaluated in the FENM program.

Program-Wide Common Reading and Listening Exams

Every semester all students take program-wide common mid-term and final exams. The questions are designed by teachers who serve on the Test Committee, which consists of FENM full-time teachers who teach at different levels. The exams include two parts: reading and listening.

Part I of the program-wide exam is reading (see Appendix A). Students are to read two passages (usually one in humanities and the other one from science) and answer multiple-choice questions on the main idea of the articles and of the paragraphs in the articles. The questions also ask about details, vocabulary, pronoun references, and inferences. The FENM Teachers' Handbook (2002) has the following guidelines for Test Committee members:

The passages should have (1) a clear, straightforward, factual introduction and explicit thesis at the end of the introductory paragraph(s), (2) a body with unified coherent paragraphs headed by clear topic sentences, and (3) a clear conclusion in the last paragraph. (p. 9)

The department chair, Z, expressed that the level of the difficulty is usually controlled to the mid-level of the university freshmen English proficiency. A major task of the Reading Test Committee members is to modify (often to simplify) the reading passages so that they will meet the guidelines stated above.

According to the FENM Teachers' Handbook (2002), the main goal in

teaching reading is to maximize students' opportunities to use basic reading skills such as skimming, scanning, reading for the main idea, using inferences to draw conclusions, and guessing vocabulary in context. Students are expected to be able to apply these skills in the reading of clearly written academic articles that have clear thesis statements, body paragraphs with topic sentences and supporting details, and conclusions (p. 11), by the time that the students leave the program.

Part II of the program-wide exam is listening, which includes three sections: Stories, Dialogues, and Appropriate Responses. Students listen to each section with its questions twice (i.e., story/questions/story/questions) without reading the script, and they are to choose the best answer from the four choices provided (printed on the exam paper) for each question (See Appendix B). The content of all three sections is usually about young people's lives and/or culture with which students are familiar. The language in the listening section is easier than that of the reading passages. The questions for the Stories and Dialogues are basically factual, WH-questions. The Appropriate Response section requires the examinee to imagine himself or herself to be one of the interlocutors and select the best answer from among four choices upon hearing "What should I say?"

The emphasis of the FENM listening component is on communication. The exit criteria state that students are expected to be able to listen to and comprehend short (5 minutes or less) simple oral passages and conversations on familiar topics. They are also expected to be able to answer multiple-choice questions about main ideas and details of the aural passages in the exams.

The Assessment of Speaking

The speaking component is the component which is emphasized most in the FENM assessment, and every student is evaluated orally up to six times (two exams and four oral assignments) per semester. The objective of speaking is specified as "students will feel more comfortable using English outside the classroom upon their completion of the FENM." The recommended formats include "speeches, radio shows, advertisements/commercials, dialogues, reports (book, country, topic etc.), interviews, show-and-tells, songs, and news reports" ("FENM Teachers' Handbook," 2002, p. 3). The FENM teachers have the freedom to choose their own formats to evaluate students. However, most teachers in this study adopted a combination of some of the recommended formats to evaluate students in the two exams and four oral assignments.

Theresa Jiin-ling Tseng

There are two program-wide (mandated) oral assignments during a school year: a dialogue and a play. The FENM Teachers' Handbook (2002) explains that for the dialogue assignment, students need to work in pairs to write and perform a dialogue about a real-life situation. For the play assignment, students need to work in groups of 4 to 6 to write and perform a 7-10 minute play, focusing on a specific theme (p. 3).

In addition to the mandated dialogues and plays, each teacher also carries out oral assignments that are given only in his or her own classes. Table 1 below summarizes the interview results on the content and format of the oral assessments given by the teachers Z, C, and J.

As can be seen from Table 1, all three teachers used interviews as a speaking-test format. However, the manner in which the three teachers conducted the interviews was different. Z asked follow-up questions, J interrupted her students' talking when she found them reciting answers from memory, and C simplified her questions when she found her students did not know what to say and asked questions for clarification when she could not understand them.

The criterion that the three teachers adopted for interview evaluation is knowledge of the content of the textbooks and lab materials. J adopted extra criteria in her own oral assessment, which were ability to comprehend the teacher's questions, the intelligibility of the student's speech, and effectiveness in using conversation strategies (e.g., to initiate and close the conversation). J stated that if students failed to answer her questions, she would ask them why they had not read the required text. If the students were able to give good excuses for not knowing the answers, then "the students will have some points." J mentioned that she taught conversation strategies (such as how to initiate a topic, how to express a different opinion, and how to end a conversation) in her classes; therefore, strategic competence was part of her evaluation criteria.

Table 1. The Oral Assessments Given by Teachers Z, C, and J

Teacher	Oral Exams (mid-term and final)	Oral Assignments in Addition to the Mandated Dialogues and Plays
Z	<p>Interviews (5-7 minutes)</p> <p>-Questions: 20 questions given to students two days before the interview. Students write and memorize the answers.</p> <p>-Content: textbook and lab materials, and topics about student life</p> <p>-How: The teacher asks some randomly picked questions from the 20 questions. He also asks follow-up questions.</p>	<p>Individual: Show-and-tell</p> <p>Paired activities:</p> <ol style="list-style-type: none"> a. Introducing a conversation partner b. Dialogues (any topic) c. Dialogues about New Year activities and resolutions
C	<p>Interviews</p> <p>-Questions: given to students two days before the interview.</p> <p>-Content: textbook, lab materials and teacher-selected Studio Classroom magazines</p> <p>-How: The teacher asks the assigned questions. She simplifies her questions or asks further clarifying questions when students have trouble in answering questions.</p>	<p>Individual: 3-minute News reports (1 person)</p> <p>Paired: Dialogues</p> <p>Group: Skits (4-5 in a group, 15 minutes, different from the mandated plays)</p>
J	<p>Interviews</p> <p>-Questions: given to students two days before the interview.</p> <p>-Content: textbook, lab materials, and topics about student life</p> <p>-How: The teacher asks the assigned questions and makes interjections when students recite answers.</p>	<p>Individual: Speeches (about personal life)</p> <p>Paired: Dialogues (e.g., role-playing interview with a celebrity)</p> <p>Group: Panel discussion on an assigned reader</p> <p>Group: Teaching a lesson on a student-selected topic to the class</p>

Theresa Jiin-ling Tseng

The Assessment of Writing

Writing is probably the least stressed skill in the program. The FENM Teachers' Handbook (2002) stated that the goal is "to enable students to use the vocabulary and grammar they learned in high school and write a 3-5 paragraph expository essay that consists of an introduction, one to three main body paragraph(s) preceded by a topic sentence and followed by a conclusion" (p. 6). Yet no text is recommended, and writing is not tested in the exams. The only required graded writing assignments are the dialogue scripts and play scripts, which the students write themselves. These are not evaluated in terms of academic writing but in terms of content, creativity, and grammatical accuracy. Although the dialogue and play are not individual performances, the FENM Teachers' Handbook specifies that "the first draft of either the dialogue or play should be written by each student and graded as a composition assignment and there should be an individual grade for each student before students work together and revise their work to produce the final draft." Thus, the writing component is integrated with the speaking component.

Teachers' Perceptions of the Overall Course Goal and Language Testing

In my interviews with the three teachers, two teachers, J and C, did not feel that the goal of the course is communicative competence, although they also recognized that communicative competence is the objective specified in the FENM Teachers' Handbook (2002). J felt the course was intended to cover the four skills (listening, speaking, reading, and writing). C said, "Although the FENM Teachers' Handbook states the goal of the program is communicative competence, the more I think about it, the more I feel our goal isn't communicative competence. Look, in our program-wide exams, what we test is traditional academic reading and listening skills. We are not testing communicative competence." C gave students a lot of listening and speaking assignments, but her listening tests were basically listening to pre-recorded passages and answering multiple-choice questions about details in the texts, and her speaking assignments such as presenting TV newscasts and reciting memorized dialogues did not involve interaction.

Nevertheless, Z expressed that the goal of the course is communicative competence. He said, "Communicative competence is integrated into our instruction. . . . It is true that reading skills are emphasized, and our exams reflect this emphasis, but the pedagogy in

classroom is a different thing You don't just explain, but you have lots of question-and-answer's." To Z, the development of communicative competence was realized in the question-and-answer activities.

So far I have presented the findings, and I will now turn to a discussion of the findings.

DISCUSSION

As mentioned in the literature review, what characterizes communicative language testing is that the target language use is integrative (in language skills), interactive, behavior-based, unpredictable, context-sensitive, purposeful, and authentic. Using these characteristics as criteria, in what follows I will discuss the findings and evaluate whether the methods of testing students' four language skills are able to elicit valid evidence of students' communicative competence. I will then discuss the relationship among course objective, teachers' perceptions, and their practices.

The Assessment of Speaking

According to the FENM Teachers' Handbook (2002), there are two kinds of program-wide mandated speaking assessments: oral exams and oral assignments. In terms of oral exams, the findings indicate that all three teachers adopted the format of a prepared interview. An underlying assumption seems to be that students cannot speak spontaneously unless they have written out and memorized what they are going to say. However, in the prepared interview, a vital element of communicative language use—unpredictability—is missing. Since the questions are given to students two days before the interview, most students write out and memorize their answers. Because of this lack of unpredictability in the interview, students do not have to negotiate meaning through interaction or adjust their language according to sociocultural purposes or contexts. Furthermore, an interview by nature is more a one-sided pattern rather than two-sided interaction. This is especially so when the teacher is the interviewer and the student is the interviewee. At the interview, the interviewee is passively waiting to be questioned so that he or she can recite his or her answers rather than actively interacting with the interviewer. In addition, the interviews given by all three teachers are not context or purpose specific; thus the language use is less likely to be

authentic. Since the questions that all three teachers asked students were about the content of the required texts, it seems that the attributes being measured can reveal students' effort in studying and their memory of what they have studied rather than how well they can communicate.

Nevertheless, some additions initiated by the teachers in the oral interviews may validly test students' communicative ability. One is the follow-up questions that Teacher Z usually gives, in which students have to listen to what the teacher says in order to answer the questions properly. Similarly, Teacher C asks questions for clarification when she cannot understand her students. These follow-up or clarifying questions are not given out to the students in advance; thus, the students need to modify their language and content in the interaction. Another addition is the interventions from the Teacher J when she notices that her students are reciting answers. Teacher J believes that recitation of answers is not real communication. Thus, in her interviews with students, her students need to activate their knowledge of the language and modify their responses according to their roles, textual, sociocultural and world knowledge strategically. By checking if students can effectively respond to questions spontaneously, initiate a conversation topic, express different opinions and end the conversation, Teacher J is able to assess strategic competence, in Canale and Swain's (1980) terms.

It is clear from the findings of this study that whether the interviews given by teachers are communicative or not depends on whether unpredictability is involved. If teachers only require students to recite answers, the interviews do not help reveal students' communicative competence. However, if teachers ask follow-up/clarifying questions or intervene in students' recitation of answers, the teachers are able to measure how students process unpredictable data in real time, which is a crucial aspect of communicative language use.

In addition to the oral exams, oral assignments are also used as means to assess students' oral competence. Two program-wide mandated assignments are dialogues and plays, which are written and memorized before performance by the students. The grading criteria for the dialogues and plays are specified as: "content, creativity, grammar, and delivery (projection, pronunciation, pace, eye contact, gestures, and fluency)" ("FENM Teachers' Handbook," 2002, p. 11). Although grammatical competence is part of communicative competence, the criteria such as creativity and delivery are appropriate for evaluating a rehearsed performance (e.g., a play) rather than spontaneous, unrehearsed

communication. As indicated in the literature review, in evaluating communicative competence, the ultimate criterion of language mastery is the learners' effectiveness (including adequacy) in communication for the settings they find themselves in (Carroll, 1980). Student performance in the memorized dialogues and plays, however, cannot serve as valid evidence of students' ability to use language spontaneously and effectively in different settings.

In addition to the mandated assignments, there are teachers' self-designed oral assignments. All three teachers used individual, paired, and group oral assignments to measure students' communicative competence. However, the individual oral assignments adopted by teachers, no matter whether they were show-and-tells, news reports, or speeches, do not involve interaction. Therefore, successful performance of the aforementioned individual oral tasks does not necessarily predict the ability to use the target language to communicate effectively in interaction.

The paired and group oral assignments such as dialogues and skits do require students to take turns to speak, but just like the prepared interviews, these assignments also lack the unpredictability that exists in any natural interaction since all the teachers allow students to write scripts beforehand. Savignon (2002) has argued that in an EFL context, communicative competence characterizes learners' ability to interact with other speakers and to make meaning, as distinguished from the ability to recite dialogues. In this FENM context, what is missing is the dynamic quality of interaction: the negotiation of meaning between the interlocutors. As a result, the type of ability that is being assessed is different from what the teachers aim to assess—communicative competence.

Teacher J's panel discussion on books is probably the most communicative oral assignment among all alternatives because her students need to pay attention to what their group members say in order to speak. That is, what is said *by* a speaker depends crucially on what is said *to* him or her. Being able to respond spontaneously in real time is one important characteristic of communicative competence (Morrow, 1979).

A commonly accepted criterion in communicative language assessment is the outcomes of behavior, but the findings indicate that outcomes of behavior have not been adopted by the three teachers as criteria for assessment. One may ask if it is feasible to use authentic tasks (i.e., tasks of real-life situations) to measure students' communicative ability and outcomes of their behavior in an EFL classroom-setting. The answer is yes.

Theresa Jiin-ling Tseng

Recommendations on how to use authentic tasks to assess EFL students will be provided in the last part of the Discussion section. I will now turn to the discussion of the assessment of listening competence.

The Assessment of Listening

The findings show that two thirds of the listening sections of the exam are not quite communicative. In the first two listening sections, students listen to pre-recorded stories and dialogues and answer multiple-choice questions. The questions test students' listening comprehension of facts presented (e.g., "According to the passage, why was Cathy's father disappointed?"), logical reasoning ability (e.g., "What can be inferred about Great Britain before the Industrial Revolution?"), and short-term memory (e.g., "What is the woman worried about?"). While skills in listening to the prerecorded messages to identify details, make inferences, and retain information in short-term memory are essential for academic success, the testing method, which requires students to answer multiple-choice questions, does not reflect the characteristics of communicative language testing. In other words, we cannot predict whether students can use these listening skills to interact with someone purposefully in a specific context in order to accomplish an authentic task. Since authentic communication is essentially interactive, an appropriate criterion for assessing listening competence can be that whether students can respond in speech, in writing or by carrying out some actions in the communicative chain after they hear something meaningful.

The third section of the listening, the appropriate-response section, is closer to a communicative orientation than the stories and dialogues sections because the examinees need to listen carefully what is said to them in order to select adequate answers. The students need to take role relationships and discourse rules into consideration when they hear "What should I say?" The ability to do so demonstrates the examinees' sociolinguistic competence. However, the element of uncertainty of this section is compromised since the examinees can read the choices provided and pick the best answer. If the choices were not provided, the examinees would have to face more uncertainty, which would require them to modify their speech and negotiate the meaning in order to provide appropriate answers. By doing so, their responses would serve as more valid evidence to predict what they can do in natural, authentic communication.

The Assessment of Reading

The reading section of the exams requires students to read passages and demonstrate what Carroll (1978) and Morrow (1979) called *enabling skills*, such as distinguishing the main point from supporting details, understanding text relations through grammatical cohesion devices, deducing meaning of unfamiliar lexis, figuring out pronoun references, and making inferences. These skills need to be mobilized in order to complete communicative global tasks, for instance, searching text for specific information (Carroll, 1978; Morrow, 1979). Morrow (1979) noted that the candidate's ability to use enabling skills should be *deduced from an analysis of task performance*. Directly testing these enabling skills equals testing only the parts but not the whole. As Morrow (1979) cautioned us, "it is conceivable that a candidate may prove quite capable of handling individual enabling skills, and yet prove quite incapable of mobilizing them in a use situation or developing appropriate strategies to communicate effectively" (p. 153). Recommendations for the testing of reading in a communicative and holistic manner are provided later in this article.

In addition to the parts/whole representation problem, the reading passages lack the authenticity that characterizes communicative language testing. According to the guidelines for designing the reading sections of the exam, the passages are to be simplified in order to control the level of difficulty. Although the intention of this control is justified, the guidelines render the reading passages less authentic than those that students usually encounter in an academic setting. That is, one cannot expect that all the readings in college textbooks or daily life to have (1) a clear, straightforward, factual introduction and explicit thesis at the end of the introductory paragraph(s), (2) a body with unified coherent paragraphs headed by clear topic sentences, and (3) a clear conclusion in the last paragraph ("FENM Teachers' Handbook," 2002, p. 9). Morrow (1979) argued that measuring the ability of the candidate to read a simplified text tells us nothing about his or her actual communicative ability, since an important feature of such ability is precisely the capacity to come to terms with what is unknown.

Furthermore, the use of a multiple-choice format is less likely to measure students' knowledge of the language and capacity in using the language to communicate effectively and appropriately for a purpose. The reason why multiple-choice is selected as a format for evaluation may lie in what Williamson (1994) called "the worship of efficiency": Multiple-choice format tests are considered to be objective, fair, and cost

Theresa Jiin-ling Tseng

efficient. Yet the efficiency is at the cost of validity in evaluating communicative competence in this case. The format of multiple-choice questions by nature allows test-takers to guess answers, so it is difficult to deduce the communicative skills or strategies the learner has employed to answer a particular question.

The Assessment of Writing

As mentioned in the Results, writing is the least emphasized skill and is not tested in the exams. Although the FENM Teachers' Handbook (2002) states that the goal of writing is "to enable students to use the vocabulary and grammar they learned in high school and write a 3-5 paragraph expository essay that consists of an introduction, one to three paragraph(s) preceded by a topic sentence, and a conclusion" (p. 6), the only required graded writing assignments are the first drafts of the dialogue and play scripts written by students themselves. Obviously, there is a mismatch between the stated goal of writing instruction and the writing assessment. As specified in the FENM Teachers' Handbook, students are to be told to write dialogues and plays *about* real life situations, yet script-writing is different from writing something *in* real life situations or to communicate for real purposes. The grading criteria listed in the FENM Teachers' Handbook are "content, creativity, and grammar accuracy," rather than the appropriateness of language use. Content and creativity may be valid criteria for a good dialogue or play script, but they are not necessary characteristics that mark successful communication. As Morrow (1979) and Carroll (1980) have argued, communicative language use should be assessed in terms of a behavioral outcome. In evaluating writing, teachers may want to evaluate whether students can use their writing to do something successfully. Thus, it is questionable as whether the dialogue and play scripts can be treated as valid evidence of students' communicative competence.

So far I have discussed how the four language skills are assessed. In the section below, I take a broader view to look at the relationship among the course objective, teachers' perceptions, and their practice of assessment.

The Course Objective, Teachers' Perceptions, and Assessment Practice

Although the course objective of the FENM is to build up students' communicative competence, what constitutes communicative competence is not specified in the FENM Teachers' Handbook (2002). In other words,

it seems to be assumed that the construct of communicative competence is clear to all FENM teachers. In spite of this assumption, however, teachers' perceptions of the course objective differed.

As mentioned in the Results section, two participants felt that objective of the program was not communicative competence, yet the department chair Z felt that it was. Z's comments that "communicative competence is integrated in our instruction. . . [because] you have lots of question-and-answer's" and "It is true that reading skills are emphasized, and our exams reflect this emphasis, but the pedagogy in the classroom is a different thing" reveal his beliefs that (1) we can claim a course to be communicative as long as it involves interactions in pedagogy, (2) reading skills are not communicative in nature and cannot be tested communicatively, and (3) pedagogy and testing can be two separate things—while the pedagogy may reflect the course objective, the assessment does not have to reflect the course objective.

Z appeared to value question-and-answer (Q-and-A) activities. Undeniably, Q-and-A may be important in teacher-student interaction, but such activities cannot help the teacher assess students' communicative competence if the teacher asks close-ended factual questions only. Close-ended factual questions call for fixed answers; thus they offer few opportunities for students to demonstrate their capacity in negotiating meaning in the target language.

In addition, the objective for the most emphasized skill, speaking, is fairly vague. It is difficult to define what constitutes "feeling more comfortable using English outside the classroom." A student might feel more comfortable using the target language outside the classroom upon finishing the course, but this does not necessarily mean that the student has built up or made progress in his or her communicative competence. In practice, however, all three teachers interviewed students, and a major criterion shared by them was whether students were familiar with the content of their texts and lab materials. The inconsistency between course objective and evaluation criterion of the speaking component reveals that the course objective needs to be modified so that it is clearer and more feasible for language assessment.

In brief, there seems to be a gap between theory and practice. The gap may have been caused by the course designers' lack of understanding of the characteristics of communicative language testing although they had intended to design a communicative-competence-oriented program.

RECOMMENDATIONS

Recommendations in Using Authentic Tasks to Assess EFL Students

From the Discussion we have learned that there seems to be a gap between theory and practice in testing students' communicative competence. This study reveals that although all three teachers are familiar with the theory of communicative competence, many of the tests or assignments used by the program and teachers to measure students' communicative ability are not authentic tasks. When I asked the department chair, Z, whether he used authentic tasks in assessing students, his answer was no, as he explained that in an EFL context it is difficult to have authentic tasks or authentic materials. Whether it is feasible to use authentic tasks to evaluate EFL students has long been a question in numerous EFL teachers' minds. In fact, authentic tasks can be designed for use in an EFL context.

Below are some practical suggestions adapted from Canale (1984) for a communicative approach to language proficiency assessment that comprises authentic tasks in a non-English context:

1. Listening comprehension: Students listen to or watch recorded mini-dialogues of various situational factors¹ and respond to open-ended questions that demand understanding of factual information as well as making judgments as to purposes, participants' roles, attitudes, probable outcomes, etc.
2. Reading comprehension: Students read short passages to complete tasks which require them to understand instructions, labels, advertisements, school and job application forms, tourist information (directions, transportation schedules), etc. Students can also read long passages, and their tasks are to work with partner(s) to figure out the meaning of new information and use contextual cues to understand the meanings of new words. For both sets of texts, students will respond to open-ended questions requiring understanding of not only factual but also sociolinguistic and attitudinal information.
3. Oral interaction: Have a discussion of the student's background, current interests and aspirations, or role-play in different situations for different purposes.

4. Written expression: Students discuss viewpoints of a contemporary issue in order to generate ideas for writing. When the writing is done, students work in pairs or groups to read aloud, clarify meanings that are vague, examine the cohesion of the text, and edit errors in both usage (e.g., sentence structure, vocabulary, punctuation, and spelling) and use (e.g., sociolinguistic appropriateness and discourse cohesion).

Other writers of communicative language testing have also provided ample ideas about authentic, task-based language testing (see Bachman, 1990; Carroll, 1980; Norris et al., 1998; Nunan, 1989 for more methods). With the fast development of the Internet, designing authentic tasks for EFL assessment has become easier than before. Communicative tasks usually call for the integration of a number of language skills (speaking, listening, reading, and writing), such as looking up information about certain topics on the World Wide Web and discussing information presented, taking notes in order to ask or answer questions orally, or asking for clarification in order to complete a written task. What test designers need to keep in mind is that any test of communicative ability needs to measure not only the knowledge of grammatical rules, but also the knowledge of how language is *used* to achieve particular authentic communicative goals. Most importantly, test designers need to recognize that any authentic language use is a *dynamic interactive process*, which involves continual modification in response to change (Bachman, 1990). Thus, for example, for interviews or role-plays to be interactive, students need to listen to what their interviewers or interlocutors say before they respond rather than reciting the scripts written.

In addition to learning about designing authentic tasks for language assessment, it is also important for EFL teachers to know what may threaten or enhance the construct validity of communicative language testing, to which subjects I now turn.

The Threat to and the Enhancement of the Construct Validity of Communicative Language Testing

Messick (1994) has cautioned test designers that “two major threats to validity in performance assessments” are “construct under-representation (which jeopardizes authenticity) and construct-irrelevant variance (which jeopardizes directness)” (p. 14). Construct under-representation means the problem of generalizing from one or a few observations of language

behavior to other real-life instances, and construct-irrelevant variance refers to a problem that performance attributes have nothing to do with language ability per se (Norris et al., 1998). In the FENM assessment program, students' ability to communicate in real-life instances is inferred from a few observations of their recitation of speeches, written dialogues or plays and two reading and listening exams that involve answering multiple-choice questions; thus, the construct of communicative competence has been under-represented. Moreover, writing and reciting scripts of dialogues or plays and answering multiple-choice questions are not quite the same as *using* the language appropriately in context. Thus, the results of assessing the writing and recitation of plays or dialogues appear to lack generalizability.

For teachers who are interested in enhancing the construct validity of communicative language testing, writers on communicative language testing (e.g., Norris et al., 1998, pp. 9-10) suggest that the tasks should

- (1) be based on communication needs analysis,
- (2) be as authentic as possible with the goal of measuring real-world activities,
- (3) have collaborative elements that stimulate communicative interactions,
- (4) be contextualized and complex,
- (5) integrate skills with content,
- (6) be appropriate in terms of number, timing, and frequency of assessment, and
- (7) be aligned with the daily actions in the language classroom.

In short, teachers who wish to test students' communicative competence may effectively enhance construct validity by adhering to the guidelines stated above.

CONCLUSION

This study has examined the construct validity of the assessment practice of a Freshman English for Non-Majors program. Since the objective of the program is to build up students' communicative competence, the evaluation methods adopted naturally need to be communicative in order to validly measure communicative competence. The literature has shown that what distinguishes communicative from

conventional language tests is that the former emphasize certain features of *genuine* language use. That is, in communicative language tests, language use is interactive, unpredictable, purposive, authentic, contextualized, performance-based and assessed in terms of behavioral outcomes (Carroll, 1980; Morrow, 1979). The more the learner's performance is assessed with regard to the production of genuine language use, the more representative the test results are in reflecting the communicative competence that underlies the behavior performed. Based on the features of genuine language use, I examined the samples of the program-wide exams and the assessment methods teachers adopted.

The findings indicated that many of the assessment methods did not match the features of *genuine language use*, which characterizes communicative language testing. Judging from the formats of assessment recommended in the FENM Teachers' Handbook (2002) and those adopted by the three teachers in this study, there seemed to be a misconception that as long as a language test involved the target language speaking, listening, reading, writing, or even answering multiple-choice questions, it could serve the purpose of measuring communicative competence. However, for instance, assessing students' ability to recite speeches, dialogues or play scripts about life-related topics/themes is quite different from assessing what they are able to do in the target language in naturalistic, real-life situations, for recitation does not involve interaction and negotiation of meaning, and it lacks the unpredictability that characterizes natural interaction. In short, any practice in language assessment that does not consider the features of genuine language use will not yield results from which we can infer students' knowledge of or competence in the language and their capacity for implementing this competence.

The findings also showed that most of the criteria used in department-wide exams and oral assignments were not representative samples of knowledge and strategies from real-world use. For instance, the criteria for assessing speaking were "content, creativity, grammar and delivery." Except for grammatical competence, the rest of the criteria bore little relevance to the construct of communicative competence. In judging communicative competence, *grammaticality* is not enough. Test designers also need to take *acceptability* into consideration because the most important criterion in communicative language assessment is the outcomes of behavior—whether the student is able to use the language to complete a task effectively and appropriately, i.e., in an acceptable manner.

Acceptability concerns sociolinguistic, strategic, and discourse competence. Yet, in most of the assessment methods used by teachers, acceptability was not adopted as a criterion because the formats or methods of assessments were not communicative. Consequently, the attributes that were measured did not permit the evaluator to predict students' knowledge (competence) of the language and the capacity for implementing the competence in specific contexts of use.

Overall, the findings revealed that the nature of what makes a test "communicative" might not be clear to the teachers/test designers. According to the FENM Teachers' Handbook (2002), the course objective was to build up students' communicative competence, but the specifications in the FENM Teachers' Handbook and teachers' actual practice often times did not match the claimed objective. The methods that were adopted in FENM assessment were not different from traditional listening and speaking assessment methods. Thus, an implication of this study is that test designers need to have a clear understanding of the characteristics of communicative language assessment. If the goal is to measure communicative competence, then the construct should be reflected not only in the teaching techniques but also in the assessment methods. It appears that professional development in designing communicative language tests is needed, and the program-wide exams and mandated communicative components need to be revised in order to reflect the construct of communicative competence.

In recent years, communicative language teaching has been one of the most popular language teaching approaches not only in Taiwan but also many other English as a foreign language (EFL) contexts. Needless to say, it is important to know how to measure students' otherwise unobservable communicative competence via students' performance. The term communicative competence may be widely known, yet from this study we learn that the characteristics of communicative language testing may not be obvious to test designers or classroom teachers. The testing methods that were used to measure students' communicative competence in the FENM program might also be the methods that are being used by many elementary school, high school, or university EFL programs that stress building up students' communicative competence. Thus, it is important for test designers of the communicative language programs to rethink if their testing methods validly measure the competence they intend to measure.

A future study may examine elementary or high school EFL teachers'

assessment methods in order to find out the construct validity of communicative language tests in such settings. A limitation of the current study is that the generalizability is compromised since only three teachers were interviewed and only one testing practice in one program was examined. Despite the limitations, the findings of this study may raise test designers' as well as classroom teachers' awareness of what needs to be taken into consideration when they design tests that measure students' communicative competence.

ACKNOWLEDGEMENTS

This article is an elaborated version of a paper "Construct Validity of an EFL Assessment Program: A Case Study," presented at The 16th International Symposium and Book Fair on English Teaching, Taipei, Taiwan, November 11, 2007. I thank Muchun Yin and two anonymous reviewers for their useful comments on earlier drafts of this article.

NOTES

1. Situational factors include, for example, topics, number and roles of participants, purposes, settings, attitudes, accents, rate of speech and complexity of and familiarity with ideas.

REFERENCES

- Bachman, L. F. (1990). *Fundamental considerations in language testing* (pp. 81-110). New York: Oxford University Press.
- Brumfit, C. J., & Johnson, K. (Eds.) (1979). *The communicative approach to language teaching*. Oxford: Oxford University Press.
- Canale, M. (1984). A communicative approach to language proficiency assessment in a minority setting. In C. Rivera (Ed.), *Approaches to language proficiency assessment: Research and application* (pp. 107-122). Clevedon, Avon, England: Multilingual Matters.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Carroll, B. J. (1978). *An English language testing service: Specifications*. London: The British Council.
- Carroll, B. J. (1980). *Testing communicative performance*. New York: Pergamon Press.
- Cumming, A. (1996) Introduction: The concept of validation in language testing. In A. H. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 1-14). Philadelphia: Multilingual Matters.
- Halliday, M. A. K. (1970). Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics* (pp.140-165). Harmondsworth: Penguin.
- Hymes, D. (1972a). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth, UK: Penguin.
- Hymes, D. (1972b). Towards ethnographies of communication: The analysis of communicative events. In P. Giglioli (Ed.), *Language and social context* (pp.21-47). Harmondsworth, UK: Penguin.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30, 955-966.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C. J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 143-157). New York: Oxford University Press.
- Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments*. Honolulu, HI: University of Hawaii at Manoa, Second

Assessment of Communicative Competence

- Language Teaching and Curriculum Center.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. New York: Cambridge University Press.
- Richards, J., & Rogers, T. (1988). *Approaches and methods in language teaching*. New York: Cambridge University Press.
- Richards, J., & Rogers, T. (2001). *Approaches and methods in language teaching* (2nd ed.). New York: Cambridge University Press.
- Savignon, S. (1983). *Communicative competence: Theory and classroom practice: Texts and contexts in second language learning*. Reading, MA: Addison-Wesley.
- Savignon, S. (2002). Communicative English teaching in Asian contexts: The challenge for teacher education. In J. Katchen, Y. Leung, W. Dai, & P. Chen (Eds.), *Selected Papers from the Eleventh International Symposium on English Teaching* (pp. 162-174). Taipei, Taiwan: Crane.
- Teachers' Handbook, Freshman English for Non-Majors (2002-2003)*. (2002). Taichung, Taiwan: Tunghai University, Department of Foreign Languages and Literatures.
- Wesche, M. (1983). Communicative testing in a second language. *The Modern Language Journal*, 67, 41-55.
- Williamson, M. (1994). The worship of efficiency: Untangling theoretical and practical considerations in writing assessment. *Assessing Writing*, 1, 147-173.
- Young, R. (2002). Discourse approaches to oral language assessment. *Annual Review of Applied Linguistics*, 22, 243-262.

CORRESPONDENCE

Theresa Jiin-ling Tseng, English Language Center and Department of Foreign Languages and Literature, Tunghai University, Taichung, Taiwan
E-mail address: ttseng3@gmail.com

APPENDIX

Appendix A. Sample Reading Passage and Questions

1. It's not just technology that's changed in the last couple of hundred years. Most of us expect that we'll finish our educations and get a job, and that we may change jobs several times throughout our lives. We expect that we'll find a mate, get married, and perhaps have children someday, and that those children will grow up and have their own lives, with their own families. Our lives are filled with change—new places to live, new jobs, new friends. Although it's very ordinary to expect these things, it's also true that someone who lived before the Industrial Revolution might think we were crazy for having such ideas. The Industrial Revolution played a big part in changing lives from predictable ones in rural settings to the more diverse existences we now enjoy.
2. We're used to the idea of constant change, but throughout most of human history, this has not been the **norm**. For thousands of years, people's lives were much like their parents' had been. A vast majority of the world's population lived in the same village or on the same farm their whole lives. Boys grew up learning their father's work so that they could continue it. Girls' fates were decided by their parents' choice of a husband for them, and there was mostly only one career: wife and mother—raising the next generation who would again live in the same place, doing the same things.
3. The Industrial Revolution changed many of those patterns. One of the greatest changes was **urbanization**, the move to cities by large numbers of people who lived in rural areas. Beginning in the 19th century in Great Britain, factories needed huge numbers of workers; these factories were built in cities because **they** needed electricity, gas, water and roads—things that were not available in rural areas.
4. People came to take these jobs perhaps not knowing that doing so would change whole societies. No longer would women be just another piece of property to be **auctioned** to the man with the most money. Even though most of the work in the factories required little training, it was still training that had to be standardized, so that it could be repeated to lots of workers. This kind of training was one of the foundations of mass education, and it was available—for the first time in history—equally to men and women.
5. Industrialization was no picnic, though. Often the work was dangerous and hard, and there were many factory-owners who cheated their workers. The work was often boring and repetitive, and workers were replaceable, so they

Assessment of Communicative Competence

felt disconnected from their work, partly because factory-work sometimes meant that a worker assembling part of a machine would never see the finished product. Perhaps the most tragic effect of industrialization, though, is pollution. For many years after the beginning of the Industrial Revolution, pollution from factories filled many rivers; more pollution came from the many people who came to work in them. The air became clouded not only from the factories, but also from the products they produced—machines that created more pollution.

6. If we can dream of deciding to live in almost any city in the world, doing work that our parents cannot even imagine, we get a sense of historical perspective from recognizing that these are ideas that were unheard of only a couple of hundred years ago, before the Industrial Revolution. From modern-day urban culture, to free public education, to equal rights for women, the effects of the growth of 19th-century factories truly changed the world in ways the first industrialists could not have foreseen.

Sample Questions

1. What is the main idea of this passage?
 - A) There are many changes in life; people get jobs and have families.
 - B) The Industrial Revolution changed the way a lot of people live.
 - C) The major effect of the Industrial Revolution was the change in women's status.
 - D) 19th century industrialists could not have foreseen the changes they would cause.
2. What is the main idea of Paragraph 2?
 - A) We lead lives that are different from those lived by people before the Industrial Revolution.
 - B) A couple of hundred years ago, most people were farmers who never went to cities.
 - C) Women have much more power in society now than they did before the Industrial Revolution.
 - D) The Industrial Revolution was the greatest change in all of history.
3. What is the main idea of Paragraph 5?
 - A) Industrialization did not mean that people could eat in parks.
 - B) Factory-workers during the Industrial Revolution had to work hard.
 - C) Air and water-pollution are effects of the Industrial Revolution.
 - D) The Industrial Revolution had some bad effects.
4. What does **norm** mean in Paragraph 2?
 - A) new way
 - B) different way
 - C) best way
 - D) ordinary way

Theresa Jiin-ling Tseng

5. What does **urbanization** mean in Paragraph 3?
A) moving to cities B) working in factories
C) getting a new job D) making revolution
6. What does **auctioned** mean in Paragraph 4?
A) married B) engaged C) sold D) given
7. What does **they** refer to in Paragraph 3?
A) factories B) workers C) cities D) rural areas
8. What or who does **them** refer to in Paragraph 5?
A) factories B) rivers C) years D) people
9. What can be inferred about Great Britain before the Industrial Revolution?
A) Public education was not available to both men and women.
B) Before the Industrial Revolution, there were no cities in Great Britain.
C) London, the largest city in Great Britain, is very polluted.
D) Factory-workers are usually unhappy people.

Appendix B. Sample Listening Scripts and Questions

The listening component of the midterm and final exam is to test general comprehension of situation-oriented listening texts and communicative competence. The purpose of the appropriate response section is to test students' immediate listening skills through the use of an appropriate response within the context of what the students heard.

Sample Short Passage and Question

Students will hear: "You will hear each passage and question twice—story/question/story/question, and you choose the best answer."

Passage 1 (Note: Students can only hear the passage read by a native speaker but cannot see the script)

[Carter's father was a famous architect and he wanted Carter to be an architect as well. He thought that it would be wonderful if his son would take over his business some day. Carter, however, was never interested in designing houses; he was interested in photography. He had a great talent for photography, especially for taking pictures of people. Carter's father was disappointed, but he respected Carter's choice.]

- Question 1: Why was Carter's father disappointed?
- A) because Carter was a poor architect
 - B) because Carter was a poor photographer
 - C) because Carter wouldn't continue his business
 - D) because Carter wouldn't become famous

Assessment of Communicative Competence

Sample Dialogue and Question

The students will hear each dialogue twice but cannot see the script. At the end of the second time, they will hear one question and choose the best answer.

Dialogue 1

[Mother: Where are you going, Alan?
Son: To the movies with my friends, Mom.
Mother: To the movies? Are you done with your midterm exams?
Son: Yes, I am. I took the last exam this morning.
Mother: How did you do on the exams? Did you pass all the exams?
Son: Give me a break, Mom! I don't know my grades yet, but I'll pass them all.
Mother: Don't be so sure.
Son: Mom, I've been studying hard for the past two weeks. Now that the exams are over, can I go to the movies? Please?
Mother: OK, but don't stay out too late.
Son: I won't. Bye.]

Question 1: What is the woman worried about?

- A) that her son might fail his midterm exams
- B) that her son might not tell his exam grades to her
- C) that her son might not know which movie to see
- D) that her son might be late for the movie

Dialogue 2: Appropriate Response

Students will hear a dialogue between a man and his wife. Sometimes the wife doesn't know what to say. You help her by choosing the most appropriate response from the choices given.

[Husband: Look, I got this big watermelon from my co-worker.
Wife: Why did he give you the watermelon?
Husband: I spent 3 hours on Sunday and got his computer fixed.

What should I (the wife) say?]

- A) Oh, my! Where is his computer now?
- B) Oh, my! A mystery present!
- C) Well, you deserve it.
- D) Well, it's really boring.