COMMUNICATIVE VALIDITY OF THE NEW CET-4 LISTENING COMPREHENSION TEST IN CHINA[1]

Chao Wang
Beijing Institute of Petrol-chemical Technology
email: wangchao@bipt.edu.cn

Ying Zuo
Beijing Institute of Petrol-chemical Technology
email: zuoying@bipt.edu.cn

Bao Liu
China Petrol University
email: liubao@cpu.edu.cn

Abstract: Based on the major dimensions of a communicative language test that Bachman proposed, this paper aims to have an investigation on the validity of the new CET-4 listening subtest in China from a communicative point of view. Both qualitative and quantitative methods are involved in the study. Material analysis falls into qualitative study, including analysis of the CET-4 testing syllabus and eight new CET-4 listening comprehension tests. Students' scores of two tests and the questionnaires are analyzed quantitatively. Through analysis, it is found that the new CET-4 listening subtest has a high validity and can measure test-takers' listening ability in real communication. First, the new CET-4 listening subtest has the quality of reliability. Second, the seven listening skills tested in this subtest can measure the communicative language ability required in the testing syllabus. The intra-correlation analysis shows that each part of the new CET-4 listening subtest focuses on different language abilities related to listening. Third, the authenticity of the new CET-4 listening subtest reaches a satisfactory level. The materials chosen in the test cover various topics and genres. Speakers' pronunciation, tone and speed are in accordance with the real situation. However, some shortcomings also exist in the test design and should be improved later. For example, its limited item types cannot represent the task types in real life, and the actual input is too ideal to be authentic.

Keywords: communicative language ability, communicative language testing, listening comprehension, test validity

VALIDITAS KOMUNIKATIF TES BARU CET-4 MENYIMAK PEMAHAMAN DI CHINA

Abstrak: Didasari oleh aspek-aspek utama pengujian bahasa komunikatif yang dirumuskan oleh Bachman, makalah ini bertujuan untuk meneliti validitas soal-soal ujian keterampilan menyimak CET-4 di Cina dari sudut pandang komunikatif. Penelitian ini menggunakan pendekatan kualitatif dan kuantitatif. Analisis material dilakukan secara kualitatif, yaitu analisis terhadap silabus CET-4 dan kedelapan jenis ujian keterampilan menyimak CET-4 yang baru. Nilai-nilai siswa dari dua jenis tes dan kuesioner dianalisis secara kuantitatif. Hasil analisis menunjukkan bahwa soal-soal ujian menyimak CET-4 memiliki validitas tinggi dan dapat mengukur keterampilan menyimak para pe-

serta tes dalam konteks komunikasi yang sebenarnya. Pertama, soal-soal ujian menyimak CET-4 yang baru memiliki reabilitas. Kedua, ketujuh keterampilan menyimak yang diujikan dapat mengukur kemampuan bahasa komunikatif seperti tercantum di dalam silabus ujian. Analisis intra-korelasi menunjukkan bahwa setiap bagian dari tes menyimak CET-4 fokus pada kemampuan menyimak yang berbeda-beda. Ketiga, keaslian tes menyimak CET-4 meraih tingkat memuaskan. Materi-materi tes mencakup beragam topik dan genre. Pengucapan kata, nada dan kecepatan penutur sesuai dengan konteks pembicaraan yang sebenarnya. Namun, terdapat beberapa kekurangan pada desain tes yang semestinya diperbaharui. Diantaranya, keragaman jenis soal yang terbatas sehingga tidak mewakili jenis kegiatan di kehidupan nyata dan input bahasanya yang terlalu asli.

Katakunci: Kemampuan bahasa komunikatif, tes bahasa komunikatif, menyimak pemahaman, uji validitas

The National College English Test Band 4 (CET-4) is a large-scale and socially acknowledged test. It has been carried out in Chinese colleges and universities since 1987, which aims to assess the fulfillment of The College English Teaching Syllabus, measure objectively and accurately the English proficiency of non-English majors, and provide feedback, exposing the weakness to improve teachers' teaching and students' learning.

With the process of China's modernization, more and more qualified personnel with English proficiency are required to meet the social needs. Therefore, improving learners' communicative competence has become a primary task for college English education. The college English Teaching Syllabus (1999) stated the goal as "…to cultivate a strong ability for reading and a fairly good ability for listening…so that students are able to exchange information in English…" However, the revised syllabus (2004) stated the goal as: "… to cultivate a comprehensive ability to use English, especially listening and speaking ability, so that the students are able to exchange information efficiently in both oral and written English..." It can be clearly seen from the two editions of syllabus that the emphasis of college English teaching has been shifted from "linguistic knowledge" to "communicative ability" (Xu, 2000). In order to increase its validity as a communicative test and achieve more beneficial washback, CET-4 is constantly under reform.

The new phase of innovation on CET-4 arises from the College English Curriculum Requirement (CECR) issued by China's Educa-tion Ministry (on trial) in January 2004. In order to reach the new curriculum requirements, the plan for the reform on the CET-4 & 6 (on trial) was issued by China's Education Ministry in 2005, followed by some radical changes to the CET-4 content and format, the scoring system and the score reporting method. The most eminent change is made in the listening comprehension test, whose weight becomes more and more significant, increasing to 35% of the total.

Major Dimensions of a communicative language test

A communicative test is designed to incorporate tasks close to those in real life with the purpose of assessing the effectiveness of communication instead of formal linguistic accuracy (Liu & Han, 2000). Bachman and Palmer (1996) identify six essential factors to evaluate the quality of a communicative test called "usefulness" (Figure 1).

Figure 1 Usefulness (as cited in Bachman and Palmer, 1996, p. 18)

$$Usefulness = Reliability + Construct\ validity +$$

$$Authenticity + Interactiveness + Impact + Practicality$$

Reliability is often defined as "consistency of measurement". In fact, it is about the degree to which the same test is likely to yield consistent scores. If they are not consistent, test scores can hardly offer valuable information about the ability we want to test. Many factors can affect reliability, such as "the extent of the

sample materials, the administration of the test, test instructions, scoring of the test and personal factors" of course.

Henning (1987) defines validity as "the appropriateness of a given test or any of its component parts as a measure of what it purported to measure". Hughes' (1989) definition is more concise and widely quoted — "a test is said to be valid if it measures accurately what it intends to measure". To put it more straight, the validity of a language test is the extent to which the test actually reflects the quality of "abstract concept e.g. language achievement, proficiency, etc."

According to Alderson, Clapham and Wall (1995), validity can be divided into three types: "face validity, content validity, and construct validity". Here we focus on construct validity.

Bachman & Palmer (1996) view construct validity as the "meaningfulness and appropriateness" of the interpretations, based on which we justify test scores. They clearly point out that we can interpret a given test score as an indicator of the ability we want to measure.

Calculating the intra-correlation between different test components is an efficient way to evaluate construct validity. We should expect these intra-correlations to be low — possibly within the range of +.3—+.5. Because different test components are designed to measure different aspects of ability and contribute to a general picture of language ability checked in the test. If the intra-correlation between two components is high, say +.9, the two components are testing essentially the same thing. Therefore, they are proved to be low efficient or meaningless.

Authenticity is the fundamental characteristic of communicative tests. In fact, it is also a major mark to distinguish communicative tests from conventional tests. Heaton (1988) noted that a communicative test is supposed to adopt authentic test materials as well as task types "related to real-life situations".

Interactiveness, another characteristic of communicative tests, means "candidates and tasks interact or interplay with each other" (Littlewood, 2000). Letter replying in conventional

tests, is actually a weak form of interaction, for test-takers are supposed to consider the expectations of the writer of the given letter, which will affect both the content and form of writing. Face-to-face interaction is suitable for communicative testing, for it reflects one's ability of "modification of expression and content mentioned above" (Brumfit, Johnson, 2000).

Practicality focuses on whether the resources required in "the design, development, and use of the test" (Bachman, Palmer, 1996) can match those that will be provided in these activities. If the resources available meet the demands for designing and implementing the test, we can say the test embodies practicality, or else it is impractical and will not be used.

Impact equals to washback effect in a narrow sense. There is a common phenomenon in education: "testing influences teaching and learning" (Alderson, Wall, 1993). Washback can be either beneficial or harmful to both individuals and the whole society. It is always the primary concern of test constructors and often demonstrated by "what we do in the classroom" (Prodromou, 1995). Therefore, in the present research, part of the two questionnaires would be used to examine students and teachers' response related to the new CET-4 listening test. It provides us another perspective to evaluate the test — whether it can encourage communicative teaching and learning (Gu & Guan, 2003).

In the present research, the validity of the CET-4 listening subtest would be examined from four qualities: reliability; construct validity; authenticity, and impact. Practicality is beyond the scope of the investigation here, since CET-4 has always been a nationally accepted mature test. The investigation of interactiveness is omitted here too, because it concerns with the interaction between the test taker's areas of language ability and the test tasks; and it can hardly be analyzed only through the material analysis.

METHOD
Considering that the communicative language competence has been attached with great im-

portance, and the assessment of listening comprehension has become more communication-oriented, the present author thinks it is of both theoretical and practical significance to carry out a validity study on the Listening subtest of the new CET-4 from the perspective of communicative language testing.

The purpose of this research is to find out whether the new CET-4 listening comprehension measures the communicative listening ability and to provide feedback to test constructors whether it is in accordance with their intentions of improving students' communicative competence. The research questions to be addressed are as follows: 1) Does the new CET-4 listening comprehension test cover the major qualities of communicative language testing? 2) Does the new CET-4 listening comprehension test reveal the communicative listening competence of test takers? 3) Does the test bring any positive washback effect to both test takers and teachers?

Subjects

Considering that the research is designed to find out whether the new CET-4 listening comprehension test measures the communicative listening ability and whether it can have positive washback effects on communicative teaching and learning, the subjects should involve both students as test takers and teachers who are influenced by the CET-4 as well.

Students: 120 sophomores (Batch 2011) come from Beijing Institute of Petrol-chemical Technology. They were randomly selected from 4 different majors, i.e. accounting (10), tourism management (35), business management (40), and thermal energy and power engineering (35). All of them were to respond to the student's questionnaire. Only the students majoring in tourism management were to attend the two CET-4 listening comprehension tests and the total number was 35.

Teachers: 30 college English teachers come from Beijing Institute of Petrol-chemical Technology. Teachers who were invited to complete the questionnaire were of different ages with different teaching experiences. However, they had one thing in common, that was,

all of them have had the experience of educating students to prepare for CET-4. Among those college English teachers, females took up 90% and males 10%, and 90% of the college English teachers investigated have taught college English for more than 5 years.

Instruments

Both qualitative and quantitative methods were adopted in order to make the present research effective and valid. In the present study, material analysis which included the CET-4 testing syllabus and eight CET-4 listening comprehension tests, belonged to qualitative method while students' scores of the two tests and student and teacher questionnaire fell into the quantitative one. Questionnaire was the main research instrument to increase the generalization of research findings. Students' scores and student and teacher questionnaire were used by the author to fulfill research questions 2 and 3; material analysis was exerted to explore research question 1.

Experimental Procedures

The research consists of three phases. Firstly, the eight test papers totally from 2006 to 2009 (two for each year) are analyzed in different aspects such as material choosing, actual input, task types, etc. to see whether these papers cover the major qualities of a communicative test.

Secondly, two tests were administered to students to get the scores which were important in verifying the reliability and construct validity of the subtest. The students were all from the same class and they were numbered from No. 1 to 35. The test takers took one test at the same time of each week (every Tuesday in their listening class). The time allocated for the two tests were exactly the same and the exams were conducted in the same class room. Two colleagues helped the researcher monitor the exam and the test papers were be rated by the researcher. The total scores and scores of each part for each test taker were collected and used in the discussion of the reliability and construct validity of the subtest.

At last, an empirical analysis was carried out to investigate with respect to test format, test construct, input materials, and the possible washback.

FINDINGS AND DISCUSSION
Reliability of the new CET-4 listening subtest

As mentioned above, the reliability of a test can be reported by the test scores. Therefore, besides the logical analysis, the researcher investigates the reliability by comparing the test scores of two test papers and calculating the reliability coefficient. These two test papers were chosen from the tests that have been conducted in July 2011 and December 2011. The 35 test takers were all from the same class and they were numbered from 01 to 35. The two scores for each test taker would be collected and the internal correlation of these two test scores would be calculated. According to the evaluation criteria stated by Hughes (1989), the reliability coefficient of an objective test of auditory comprehension should be in the 0.8 to 0.89 range. The results of the two tests for each test taker and the reliability can be seen from Table 1.

Table 1 The results of the reliability coefficient processed by SPSS

Reliability

[DataSet1]

Scale: ALL VARIABLES

Case Processing Summary

Cases		N	%
	Valid	34	97.1
	Excluded[a]	1	2.9
	Total	35	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.867	2

As it is shown in the table above, the correlation between the scores of two tests is 0.867. This result shows that, the listening subtest in CET-4 is comparatively reliable.

The construct validity of the new CET-4 listening subtest

Construct validity concerns whether the test actually tests the abilities we want to measure. During the listening procedure, the abilities to obtain the information required concern more on strategic competence and psycho-physiological mechanisms, rather than on the language proficiency itself. The test takers would use strategic competence to get background information from the written information in the test paper, and planning for their response. And of course the psycho-physiological mechanisms would be involved in such a process. However, they can not be tested separately, which makes the analysis of the construct validity less reliable. Hence, the author will focus on listening skills, which can be analyzed through the test paper and seem to be more controllable and objective.

According to Testing Syllabus (2006), there are seven skills required in the listening subtests of CET-4 as follows:

01 understanding gist
02 understanding specifics and important details
03 determining speaker's attitude/intentions toward listener/topic
04 making inference and deductions
05 recognizing the communicative function of utterances
06 understanding phonological features (stress, intonation, etc)
07 understanding relationship between sentences such as comparison, cause, result, degree, purpose

Among them, sub-skills 02, 06 and 07 are at the lower level, since they mainly concern the language proficiency. Sub-skills 01, 03, 04 and 05 involve topical knowledge and affective responses as well. Thus, they seem to be more communicative and difficult and occupy the higher level. Moreover, it is not easy to identify

listening skills in question items sometimes because two or more skills may be involved. In this case, the item will be analyzed on basis of the skill that is mainly used.

To tell the frequency each skill covers, we would like to refer to the data collected. The following tables (Table 2, Table 3 and Table 4) show the distribution of these skills in each part of CET-4 listening subtest, and this in turn list out the construct being analyzed in the listening test.

Table 2 Listening skill coverage of Short Conversations

Test paper	01	02	03	04	05	06	07
2006.06	/	/	3	5	/	/	/
2006.12	/	1	3	4	/	/	/
2007.06	/	1	1	5	1	/	/
2007.12	/	2	1	5	/	/	/
2008.06	/	/	4	3	1	/	/
2008.12	/	1	2	5	/	/	/
2009.06	/	4	1	3	/	/	/
2009.12	/	2	2	4	/	/	/
Total	0	11	17	34	2	0	0
Percentage	0	17.2%	26.6%	53.1%	3.1%	0	0

From Table 2, we can find that there are four listening skills involved in Short Conversations, among which the skill of making inference and deductions is the most checked one, occupying more than half of the question items. The skill of determining speaker's attitude or intentions is the second. It is in accordance with the features of conversations in daily communication. People involved in conversations act as not only a hearer or a speaker, but also a participant who play a certain role in the particular situation. They need to make inference and de-ductions to understand each other totally and definitely to ensure their proper response either in words or actions. The skill of understanding specifics and important details is the third, aiming at test listeners' language knowledge. Although only 3.1 per cent of question items tests the skill of recognizing the communicative function of utterances, such items add much communicative function to conversations. In general, skills on higher level are more checked in this part.

Table 3 Listening skill coverage of CET-4 Long Conversations

Test paper	01	02	03	04	05	06	07
2006.6	/	2	1	4	/	/	/
2006.12	/	6	/	1	/	/	/
2007.6	1	5	1	/	/	/	/
2007.12	/	5	1	1	/	/	/
2008.6	/	7	/	/	/	/	/
2008.12	/	4	1	2	/	/	/
2009.6	1	4	1	/	1	/	/
2009.12	/	6	/	1	/	/	/
Total	2	33	5	9	1	0	0
Percentage	4%	66%	10%	18%	2%	0	0

Different from Short Conversations, one more skill is covered in Long Conversations (see Table 3). That is the skill of understanding the gist. It is rational because the long conversations with around 5 to 7 turns in CET-4 are long enough to offer context to check test-takers' understanding of the gist. Contrast with Short Conversations, the skill of understanding specifics and important details is the one most checked, while the skill of making inference and deductions is the second one. In general, skills on lower level are more checked in this part.

Table 4. Skill coverage of CET-4 Short Passages

Test paper	01	02	03	04	05	06	07
2006.6	/	7	/	/	/	/	3
2006.12	/	6	1	2	/	/	1
2007.6	/	6	/	4	/	/	/
2007.12	/	7	/	3	/	/	/
2008.6	1	7	/	1	/	/	1
2008.12	/	8	1	/	/	/	1
2009.6	/	7	/	3	/	/	/
2009.12	1	8	/	1	/	/	/
Total	2	56	2	14	0	0	6
Percentage	2.5%	70%	2.5%	17.5%	0	0	7.5%

As is shown in Table 4, five listening skills are involved in Short Passages. Among them, the skill of understanding specifics and important details is the most checked one, while the skill of making inference and deductions is the second. The skill of understanding gist and the skill of determining speaker's attitude or intentions is tested in the same small proportion. Only in this part the skill of understanding relationship between sentences such as comparison, cause, result, degree, purpose is tested. As a whole, skills on lower level were more tested in this part.

The above logical analysis on test papers proves that the skills being checked cover all aspects of communicative language ability and the skills tested in each part has different focuses. Thus, the construct on which the test design is based is valid and communicative.

Besides the above logical analysis, correlation analysis was also employed here to assess the construct validity of the new CET-4 listening subtest. As mentioned in the previous parts, the intra-correlations between different components of a test are supposed to be fairly low — possibly within the range of +.3 — +.5, if these components are designed to test different aspects of one's language ability. Here we regard Short Conversations, Long Conversations, Short Passages and Compound Dictation as four different components of the listening subtest, the intra-correlation should therefore be in accord with the range. The following table shows the intra-item correlation coefficients.

Table 5. The internal correlation of the different parts in CET-4 listening subtest

Inter-Item Correlation Matrix

	short conversation	long conversation	short passages	compound dictation
short conversation	1.000	.500	.517	.359
long conversation	.500	1.000	.316	.386
short passages	.517	.316	1.000	.402
compound dictation	.359	.386	.402	1.000

From the table, we can find that the correlation between different parts within the listening subtest is basically within the range from 0.3 to 0.5. This indicates the different parts of the subtest examine different language abilities related to listening.

Authenticity of the new CET-4 listening subtest

Authenticity is the fundamental characteristic of communicative tests. It is a major mark to distinguish communicative tests from conventional tests.

To analyze the authenticity of the new CET-4 Listening subtest, we put our focus on two levels: the input level and the task level.

A listening test is said to have the characteristic of authenticity when its content covers the topic or genre of language that test-takers are likely to use in the real communication. According to the requirements of the listening materials in teaching and testing syllabuses of CET-4, the topics should be extensive, including the fields of humanity science, social science and natural science, etc. And in the present research, eleven specific topics are classified in the short and long conversations, seven in the passages and compound dictation (two for nonacademic topics and five for academic topics), based on the analysis of the eight CET-4 listening comprehension tests (2006-2009).

The topics of Short conversations and long conversations are analyzed based on the eleven topics, all of which are on nonacademic topics, such as study, job, housing, dining, personal budget, reservations, weather, entertainment, travelling, people and their stories. The results are presented in Table 6.

As is shown in the table, the topic of daily life occupies the overwhelming majority, which covers various aspects of our daily life, such as renting, buying or decorating a house, dining out, making personal budget, reserving a hotel room, talking about weather or friends' mood and changes, and telling one's own experiences or stories, etc. Generally speaking, they are all familiar topics to test-takers, to whom the listening materials do not cause cultural bias. Study topics and job topics are respectively the second and third largest categories, which are equally important.

Study topics are about preparing one's examination, listening to professors' lectures and making comments on them, electing courses, waiting for the professor at the office, borrowing books from library, etc. All these topics are not only familiar to test-takers but also vivid and practical to them. Most college students have experienced such situations in their real life.

Job topics mainly involve the feelings of and opinions on present jobs, the situation of preparing for job interview or having a job interview, working conditions, or promotions in work, etc. This type of topics is also practical for students, because after graduation, they will face job hunting and sometimes they will be interviewed in English. In addition, they may use English in their future work.

The topic of leisure life accounts for the fourth place. It contains two large parts named entertainment and travelling. The former concerns the activities like going to a party, talking about a TV program, a movie or some new digital television system, going to a theater, and so on. The latter concerns the activities, like asking for advice to travel abroad or preparing the package and passport for travelling. Actually, this type of topic is quite common in daily life, and seems attractive and interesting to test-takers.

Table 6. Topics of Short Conversations and Long Conversations

Test papers	Study	Job	Daily Life							Leisure life	
			Housing	Dining	Budget	Reserva-tions	Weather	Friends	Personal experience	Entertain-ment	Travelling
2006.6	1	2	2	1	1	/	1	1	1	/	/
2006.12	2	/	/	1	1	/	/	1	3	1	1
2007.6	3	2	/	/	/	1	/	/	3	1	/
2007.12	2	2	1	/	/	1	/	1	2	/	1
2008.6	2	1	/	/	/	/	/	/	4	3	/

Test Papers											
2008.12	2	1	2	/	/	/	/	1	3	1	/
2009.6	/	4	1	/	/	/	1	2	2	/	/
2009.12	3	2	/	/	/	/	/	/	2	3	/
Total	15	14	6	2	2	2	2	6	20	9	2
					40					11	
Percentage	18.8%	17.5%			50%					13.8%	

To sum up, the topics chosen for short conversations and long conversations bear the qualities of extensiveness, authenticity and familiarity. Compared with the non-academic topics of short conversations and long conversations, more academic topics appear in the materials for short passages (see Table 7), including culture, sports, medicine, language, science and technology, which helps students to understand some lectures or materials concerning their majors. The non-academic topics mostly concerns social life and people's stories, which are apparently more familiar topics to test-takers and easier to understand, such as Hollywood kids' life and the social problems caused by it (CET-4 200906). All these indicate that generally the Short passages are more difficult to understand than conversations.

Table 7. Topics of Short Passages

Test Papers	Academic topics					Non-academic topics	
	Culture	Sports	Medicine	Language	Science and technology	Social life	Personal Story
2006.6	1	1	/	/	/	1	/
2006.12	/	/	2	/	/	/	1
2007.6	1	/	/	/	/	1	1
2007.12	/	/	/	/	/	2	1
2008.6	/	/	/	/	1	/	2
2008.12	/	/	/	1	/	1	1
2009.6	/	/	/	/	1	1	1
2009.12	/	/	1	/	2	/	/
Total	2	1	3	1	4	6	7
Total			11			13	
Percentage			45.8%			54.2%	

The variety and richness of the topics of Compound Dictation can be seen from Table 8.

The number of academic topics exceeds the non-academic ones.

Table 8. Topics of Compound Dictation

Test Papers	Academic topics					Non-academic topics	
	Culture	Sports	Medicine	Language	Science and technology	Social life	Personal Story
2006.6	/	/	/	/	/	1	/
2006.12	/	/	/	1	/	/	/
2007.6	/	/	1	/	/	/	/
2007.12	/	/	/	/	/	1	/
2008.6	/	/	/	/	/	1	/
2008.12	1	/	/	/	/	/	/
2009.6	/	/	1	/	/	/	/
2009.12	/	/	/	1	/	/	/
Total	1	0	2	2	0	3	0
Total			5			3	
Percentage			62.5%			37.5%	

Based on the test papers analysis, it is found that the topics of listening materials are extensive, authentic and familiar to students. Moreover, different part of the listening subtest focuses on different types of topics.

Genre refers to a type of stylistic structure into larger units. Testing syllabus (2006) stipulates that the passage genre should be varied, including narration, argumentation and exposition. So the writer of the paper tends to analyze the stylistic characteristics from narration, ar-

gumentation and exposition. A survey of the genres of the passages in Short Passages and

Compound Dictation is presented in Table 9 & 10.

Table 9. Genre of CET-4 Short passages

Test papers	Narration	Exposition	Argumentation
2006.6	/	2	1
2006.12	1	1	1
2007.6	2	1	/
2007.12	1	1	1
2008.6	2	/	1
2008.12	1	1	1
2009.6	1	/	2
2009.12	1	1	1
Total	9	7	8
Percentage	37.6%	29.1%	33.3%

Table 10. Genre of CET-4 Compound Dictation

Test papers	Narration	Exposition	Argumentation
2006.6	/	/	1
2006.12	/	1	/
2007.6	/	/	1
2007.12	/	/	1
2008.6	/	1	/
2008.12	/	/	1
2009.6	/	1	/
2009.12	/	1	/
Total	0	4	4
Percentage	0	50%	50%

Generally speaking, the genre of listening materials of Short Passages and Compound Dictation is varied, which meet well the requirements of genre in testing syllabus. Each genre can find its place.

Specifically speaking, the genre distribution as a whole is proper and reasonable. For Short Passages, narration, exposition and argumentation with the proportion of 37.6.3%, 29.1%, 33.3%, demonstrate the similar trends, that is, test-takers will encounter listening materials with different stylistic forms in real life, so it is necessary to have the ability or strategies to understand different stylistic listening materials. For Compound Dictation, exposition and argumentation occupy the same percentage (50%). Due to its higher requirements on lis-

teners to understand and memorize some specific information, expository and argumentative texts are advisable. Since they either objectively elaborate on a certain theory, method, and fact or try to be persuasive enough to bring the readers to their viewpoints, it may be a challenge to process.

Washback effects of the new CET-4 listening subtest

In 2006, the new CET-4 was carried out with the purpose to accurately measure college students' communicative language ability, especially listening and speaking ability so that they can communicate effectively in their work and life with both oral and written English in the future. The general intention of the new CET-4

constructors is to make college language-teaching concentrate more on communicative ability of non-English major undergraduates.

The results of both students' and teachers' questionnaires will be presented separately from four aspects: teachers and students' perception on the test, teaching and learning attitudes, teaching and learning content as well as teaching and learning methods. The purpose is to investigate whether the test encourages communicative approaches to teaching and learning.

To conclude, from the analysis on the questionnaires, we may find that both teachers and students are affected by the new CET-4 listening subtest, and many positive impacts are elicited. The trend is obvious: Influenced by the test, both CE teaching and learning are more or less changing from the exam-oriented mode into a real-life-oriented one. In other words, the new CET-4 listening test encourages the communicative teaching and learning.

CONCLUSION

Since 2006, great innovations have taken place in CET-4, especially in the listening comprehension subtest. Remarkable changes have been achieved in both input materials and test format. The present study tries to explore the validity of the new CET-4 listening subtest from the perspective of communicative language testing.

The major findings are briefly summarized to answer the three research questions:
1) The new CET-4 listening subtest mostly covers the major qualities of communicative language testing.

First, the new CET-4 listening subtest has the quality of reliability. The correlation between the scores of two objective tests in the present research is 0.867, within the reasonable range from 0.8 to 0.89. This result shows that, the new CET-4 listening subtest is comparatively reliable with regard to the testing environment.

Second, the new CET-4 listening subtest has satisfactory construct validity. On the one

hand, the logical analysis on test papers shows that the seven listening skills checked in this subtest cover the major elements of communicative language ability. The skills checked in each part have different focuses. On the other hand, the intra-correlation between different parts varies from 0.316 to 0.517 (see table 3.5), within the acceptable range from 0.3 to 0.5. It also proves that each part of the new CET-4 listening subtest focuses on different language abilities relating to listening.

Finally, the new CET-4 listening subtest has a reasonable and acceptable level of overall authenticity. The materials selected in the test embody extensiveness in terms of topics and genre. A great variety of topics close to students' daily life are selected in the test, which can not only contribute to reducing the bias of listening materials but also to broadening students' horizon and enriching their knowledge. The genre distribution is reasonable and meets both the requirements of different stylistic forms in real life and the teaching and testing syllabus. The actual input of materials, whose authenticity is measured in terms of accent, tone and speed, is close to real life and can be regarded as authentic. What's more, the students and teachers' questionnaires also demonstrate a general approval of the authenticity of test materials, actual input, and task type.
2)The new CET-4 listening comprehension test largely reveals the communicative listening ability of students.

The seven listening skills tested in this subtest cover the major elements of communicative language ability and the language competence tested in each part has different focuses. Through intra-correlation analysis, we find that correlations between each part of the listening test varies from 0.316 to 0.517, within the acceptable range from 0.3 to 0.5, which implies that different parts in the test check different language abilities relating to listening. The results of students and teachers' questionnaires also approve of the construct validity of the test, and regard it as a reliable measurement of their listening ability. Their responses show that the actual input of the new CET-4 listening subtest

119

is quite close to real life and can help to guide students to develop their communicative competence instead of doing examination-oriented learning for examination's sake.

Although the strategic competence and psycho-physiological mechanisms are also important components in Bachman's framework of communicative language ability, the analysis of the test paper puts more emphasis on analyzing the language competence it examines, so this study mainly focused on the language competence in communicative language ability. 3) The test does bring some positive washback effect to both students and teachers.

The results of questionnaires show that the washback effects of the new CET-4 listening subtest on both language teaching and learning are positive. Students can be encouraged to develop their communicative listening ability through the preparation of the test.

However, the study also discovers some problems affecting the validity of the test.

First, the item types in the test are limited, most of them are multiple-choice questions, and others are compound dictations. Despite the reliability, they are not adequate for the authentic tasks required in a communicative context.

Second, the aural input of the test is designed in an ideal situation, which does not thoroughly reflect the real nature of listening in terms of tone and pronunciation

Finally, despite its great innovation, the new CET-4 listening subtest does not avoid the chronic illness of any large-scale national examination — lack of interactiveness. Interactiveness is another characteristic of communicative testing. The internet-based CET-4 issued in 2008 (on trial) contributes greatly to the improvement in this aspect.

Since CET-4 is a large-scale national examination, which has great effect on both CE teaching and learning, more researches are needed to study its validity as a communicative test. Thus, CET-4 can serve well as a measurement of college students' English proficiency and a stimulus to encourage communicative teaching and learning. As a result, the efficiency of CE teaching and learning can be im-

proved. Therefore, more validity studies on other subtests in CET, such as reading and writing are needed to have a general picture of CET.

REFERENCES

Alderson, J. C. and Wall, D. (1993), Does washback exist? *Applied Linguistics, 14*(2).

Alderson, C., J., Clapham M. C. & Wall D (1995). *Language test construct and evaluation.* Cambridge: Cambridge University Press.

Bachman, F., L., & Palmer S. A. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Brumfit, J., C., & Johnson K. (2000). *The communicative approach to language teaching.* Oxford: Oxford University Press.

Gu, Xiangdong, Guan, Xiaoxian. (2003). A Sample Study on Readability of CET Reading Tests and College English Reading Textbooks. *Journal of Xi'an Foreign Language University 3,* pp. 39-42.

Heaton, J., B. (1988). *Writing English language tests (New Edition).* London: Longman Group UK Limited.

Henning, G. (1987). *A guide to language testing.* Cambridge: Newbury House.

Higher Education Department of the Ministry of Education. (1999). *The college English teaching syllabus.* Beijing: Higher Education Press.

Higher Education Department of the Ministry of Education. (2004). *College English curriculum requirement (On Trial) (CECR).* Beijing: Higher Education Press.

Hughes, A. (1989). Testing for language teachers. Cambridge: Cambridge University Press.

Littlewood, W. (2000). *Communicative language teaching.* Cambridge: Cambridge University Press.

Liu Runqing, Han Baocheng. (2000). *Foreign Language Testing and Approaches.* Beijing: Foreign Language Teaching and Research Press.

Prodromou, L. (1995). The backwash effect: from testing to teaching. *ELT Journal 49*, pp.13-25.

Testing Committee of CET-4&CET-6. (2006). *Testing Syllabus of CET-4 & CET-6 (revised edition).* Beijing: Tsing Hua University Press.

Xu Qiang. (2000). *Interactive English Teaching and Testing*. Shanghai: Shanghai Foreign Language Education Press.