

# 1 | College English Education in China: From Testing to Assessment

Yan Jin  
(Shanghai Jiao Tong University, China)

## INTRODUCTION

With a tradition of using tests as a fair and objective means for selective purposes, education in China has long been influenced by its testing-oriented culture. Selecting talents by means of examination has been an important element of Chinese culture since the Sui Dynasty (581-618 AD). Sun (1985) viewed the examinations system in ancient China from a global perspective and hailed the 1300-year Imperial Examinations as “the oldest and best examination system in the world” (p. 511).

In pursuit of fairness and objectivity in a competitive world, testing remains a powerful tool for decision-making in present-day China. What makes the situation in modern China even more significant is that tests are often of an alarmingly large scale. Tests involving over one million testees a year include the National College Entrance Examination, the Graduate School Entrance Examination, the Public English Testing System, and the College English Test (See Cheng & Curtis, 2009; Yang, 1999 for major testing programs in China today). Admittedly a necessary evil, testing has had strong impact, both positive and negative, on education in China today (Yang & Gui, 2007).

The College English Test (CET), the focus of discussion in this paper, is a typical case in point. In the past two decades, College English<sup>1</sup> has witnessed its fastest development, and the CET, the program’s exit test, has gained a high academic reputation as well as general social recognition for its validity and reliability. However, the development of College English education has been, to some extent, overshadowed by its testing-oriented teaching and learning practices. This paper takes a close look at the influence of the CET, which currently has an annual

test population of over 16 million, on English language teaching and learning at the tertiary level and advocates a shift of paradigm from an overreliance on testing to systematic assessment for student improvement and curriculum development.

## **BACKGROUND AND KEY CONCEPTS**

### **College English**

College English is a compulsory English language education program for tertiary level students in China<sup>2</sup>. The program was started in the late 1970s, when the National College Entrance Examination was resumed after a hiatus of over a decade during the devastating Cultural Revolution. In the 1980s, the program was, in fact, only a foundational language course in the college curriculum, aiming at improving students' English language skills with an explicit emphasis on the ability to read in English for Academic Purposes (EAP) (Working Group on College English Teaching Syllabus, 1985, 1986). The other language skills, listening, speaking and writing, were considered of secondary importance. The course was made compulsory in the Syllabus for all non-English college majors. Since the 1980s, China's socio-economic reform has provided a strong impetus for English language teaching and learning. To meet the changing needs of society, major revisions were made in the Syllabus in the late 1990s. While still giving top priority to the EAP reading ability, the revised Syllabus stressed the importance of EAP listening and writing and suggested that the two skills be given more attention (Working Group on College English Teaching Syllabus, 1999). The revised Syllabus also suggested a variety of optional courses to enhance students' language learning skills and strategies, their cultural awareness and cultural knowledge.

Upon entering the 21st century, the Department of Higher Education of the Ministry of Education saw the necessity of setting higher requirements on college students' ability to use English as a tool of communication. The newly promulgated College English Curriculum Requirements (Department of Higher Education, 2007) stipulated that "the objective of College English is to develop students' ability to use English in a well-rounded way, especially in listening and speaking, so

that in their future studies and careers as well as social interactions they will be able to communicate effectively, and at the same time enhance their ability to study independently and improve their general cultural awareness so as to meet the needs of China's social development and international exchanges" (p. 25). College English, therefore, has now evolved from a skill-based language course to a typical language program, which, as defined by Lynch (1996), "consists of a slate of courses designed to prepare students for some language-related endeavor" (p. 2).

Among all the foundational courses for tertiary level education, College English enjoys the most credit hours, and receives the greatest attention from administrators, teachers, students and society in general. College English teachers are provided with more chances to receive professional training. Teaching facilities such as audio- or video-tapes, multimedia classrooms, or satellite radio receivers are purchased with special budgets from the university. The rising status of College English in tertiary level education has led to significant improvement of the overall English language proficiency of non-English college majors (Jin & Yang, 2006).

### **College English Test**

To complement College English teaching and learning, the CET, a standardized English proficiency test, was designed for college students and officially launched in the late 1980s by the CET Design Group (now the National College English Testing Committee, hereafter NCETC). As a national test of an optional nature, the CET was intended to motivate teachers and students at a time when English teaching and learning was much neglected and promote the implementation of the national College English Teaching Syllabus. Since its inception, the CET, which had only one band in 1987 when it was administered for the first time, has grown into a test series with CET Band 4 (CET-4), CET Band 6 (CET-6), and the CET Spoken English Test (CET-SET), and the scale of the test has increased from 100,000 in 1987 to 16 million in 2008 (See Table 1).

**TABLE 1****The Number of College Students Registered for the CET in Recent Years**

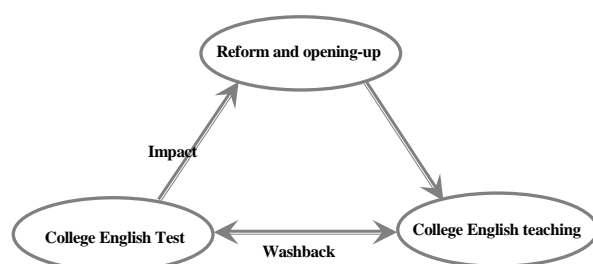
	2000	2001	2002	2003	2004	2005	2006	2007	2008
CET-4	2864738	4042631	5449869	6756402	8205600	7717742	8652566	9753786	10480209
CET-6	1161124	1589509	2092746	2742302	3498069	3242501	3505811	4443129	5618722
Total	4025862	5632140	7542615	9498704	11703669	10960243	12158377	14196915	16098931

Notes.

1. The CET is administered twice a year. The table shows the number of students registered for the CET-4 and CET-6 in recent years.
2. The CET-SET is a separate test in the form of face-to-face oral interview. The test is also administered twice a year and has an annual test population of less than 100,000.

The development of the CET, in fact, exemplified the relationship between social needs, language teaching and language testing (See Figure 1). The compulsory instruction of College English was necessitated by the growing demand of a more open society for university graduates with communicative competence in English. The teaching syllabuses defined and refined the teaching objectives and served as guidelines for establishing teaching models and developing teaching materials. With College English becoming a compulsory language course, a reliable and valid test was considered necessary as a tool for promoting teaching and learning, and exerting positive impact on society. Take the development of the CET-SET as an example. At the beginning stage of China's opening up and reform, English speaking skill was considered somewhat less important than the other major English language skills and was therefore seldom taught and never assessed. In the mid-1990s, the increasing demand for college students' ability to communicate orally in English led to the inception of the CET-SET in 1999, which, as expected by its designers, has had huge impact on the instruction of the speaking component in the College English curriculum (Jin, 2000). Now there are 58 CET-SET test centers in 35 major cities in China and the test is administered twice a year, in May and November respectively.

**FIGURE 1**  
**Social Needs, College English Teaching and College English Test**



The implementation of the CET in the past two decades has had profound influence on the teaching and learning of English in tertiary institutions, and impacted English language teaching and learning in a wider social context (Jin, 2009; Zheng & Cheng, 2008). The test, for example, has proved to be an effective means to ascertain and recognize the proficiency levels of students on a standardized basis. The test results describe in objective terms the overall English proficiency level of students in an institution, a city, a province or a greater part of the country, which is useful for educational policy-making at different administrative levels. This explains why the implementation of the large-scale test has gained strong support from provincial and municipal educational authorities and institutional administrators. The test has also proved useful to developing a shared understanding among College English practitioners, including curriculum designers, teachers, and material developers, about the requirements of College English teaching and learning. This shared understanding is essential to maintaining the standards of College English teaching and ensuring the achievement of the teaching objective. At a news conference held by the Ministry of Education on the reform of the CET, the vice-minister favorably commented on the success of the test as follows:

The fact that such a large-scale test has been developing steadily in the past 17 years is in itself solid evidence to show that the CET has met social needs, won social recognition, produced

beneficial effects on society, and contributed significantly to the continual improvement of the quality of College English teaching in China. (Wu, 2005)

### Testing and Assessment

In the field of educational measurement, *testing* and *assessment* share overlapping shades of meaning when used to refer to measurement in a general sense. In fact, the two terms have been used interchangeably until very recently. For the discussion in this paper, however, distinctions between the two types of measurement activities are emphasized to call attention to their different functions and purposes.

*Testing*, as used in the paper, refers to measurement of students' language proficiency for summative purposes. Test results are reported mainly in scores and students can be rank-ordered according to their performances in the test. *Assessment*, however, is designed for understanding and improving learning, involving such tasks as "making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing and interpreting evidence to determine how well performance matches those expectations and standards; and then using the resulting information to document, explain, and improve performance" (Angelo, 1997, p. 73). Assessment can support learning as well as measure it (Askham, 1997; Black & Wiliam, 1998; Stiggins, 2002). To put it in simpler terms, *assessment* refers to systematic collection of both qualitative and quantitative data for the purpose of monitoring the process and controlling the quality of teaching and learning. When assessment activities are properly blended with classroom instruction, formative assessment will become part of the pleasant learning process rather than frustrating testing experiences and the line between teaching and assessment blurs.

## CONSEQUENCES OF HIGH-STAKES USES OF THE CET

### Uses of the CET Not Intended by Its Designers

The consensus in the field of language testing is that large-scale high-stakes language tests could have huge impact, both positive and negative, on individuals as well as society (e.g., Alderson & Wall, 1993; Cheng, Watanabe & Curtis, 2004; Shohamy, 1997, 2001a, 2001b; Wall, 2000, 2005). The case of the CET illustrates the complexities involved in ethical test use, a topic of growing interest in language testing since the 1990s (Bachman, 2000, p. 1). Amid the rising reputation of the CET as a domestically produced English proficiency test and the first large-scale test that was empirically validated (Jin, 2000; Jin & Wu, 1998; Yang & Weir, 1998), there has been serious concern over the high-stakes uses of the CET that were not intended by its designers, placing the test in a complex context with wide-ranging consequences.

The CET is used by society for a variety of purposes across different levels – national, provincial, and institutional. At the national level, the test results are used by educational authorities as one of the key factors for decisions on the Project 211, a project initiated in the early 1990s by the Ministry of Education to cultivate high-level elite in 100 key universities to prepare for China's socio-economic development in the 21st century. Inclusion in the list of the Project 211 would greatly increase the university's educational resources invested by the government. At the provincial level, CET results are used by the local governments in major cities like Shanghai and Beijing to determine whether a college graduate can become a permanent resident of the city. The overall pass rate of the students in a college is also an important determinant for the college league table. Colleges themselves often use the test results for institutional accountability purposes, quoting the pass rate to stakeholders or external inspectors as the indicator of their high quality of teaching. Overseas, the CET-6 is used by some institutions as an alternative to the TOEFL or IELTS for applying to their graduate programs.

## Consequences of the Over-Uses of the CET

### *College English Teaching and Learning*

Since the mid-1990s, the currency carried by the CET certificate has turned the optional test into a ‘must’ and the test has been attached undue importance by educational policy-makers, teachers, learners and other stake-holders like parents and employers. High-ranking officials of the Ministry of Education and members of the NCETC reiterated, on many occasions, the optional and summative nature of the test. However, for college students, a good CET score nowadays could determine their opportunities for a satisfactory job and even legality to be granted a residential certification. The competitive environment created by the imperative for colleges to progress on the league table has also made the CET a prerequisite for graduation or awarding of a bachelor’s degree in quite a number of universities (Wang, 2008).

College English teachers, as a result, are overridingly concerned with their students’ performance in the test and often modify teaching to suit the test. Some universities emphasize the maximization of students’ performance in the test at the expense of the actual improvement of their communicative language ability. For example, in a few universities, normal teaching activities are replaced by practices of mock tests of inferior quality to train students in test-taking strategies. A more worrying phenomenon is that with a disproportionate focus on what teachers believe is tested in the CET, the curriculum of College English is narrowed (Gu, 2004). Some critics even pin the blame for some university graduates’ unsatisfactory performance in workplaces on the negative impact of the CET and question the necessity of the test (e.g., Liu, 2007). Like it or not, the “tail-wagging-the-dog” phenomenon (Li & Zeng, 2002) may, in the long term, seriously hamper the sustainable development of College English education and the CET as well.

### *College English Curriculum Development*

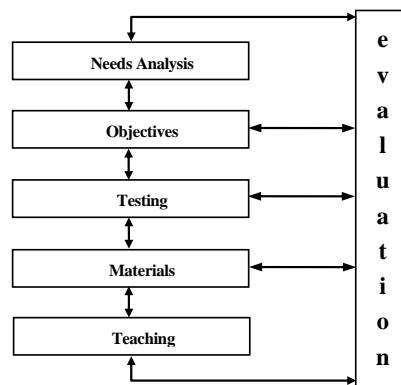
Added to the undesirable phenomenon of teaching-to-the-test is the over-reliance of College English curriculum design and implementation on the CET, that is, the CET is often used as the predominant means of assessment and evaluation in College English education, making the process of College English curriculum design and implementation a



largely one-way system instead of a cyclical one as described in Brown (2001, p. 20).

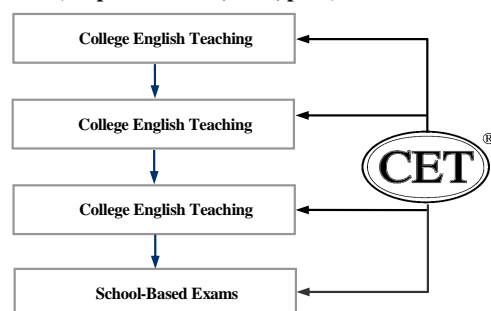
According to Brown's systematic curriculum development framework, designing and maintaining a language curriculum should be an on-going process, which involves a very dynamic interaction between the key components (See Figure 2). *Testing* and *evaluation* play a key role in such a systematic process. The information provided by *testing* informs *needs analysis* and serves to operationalize *objectives*. *Testing* also provides data to show whether *materials* fit teaching objectives as well as students' levels and needs. *Testing* is used by teachers to monitor the teaching process, diagnosing deficiencies and demonstrating achievements. All these components are interrelated and linked through constant and regular *evaluation*, which demonstrates how effectively the system is working. Poorly-achieved objectives should lead program designers to examine the entire system in order to identify places where improvements might be made, which could involve a change in the objectives, a revised assessment of students' entry level and needs, a critical review of the instructional methods, a review of testing methods, and so on.

**FIGURE 2**  
Systematic Approach to Designing and Maintaining Language Curriculum (Brown, 2001, p. 20)



In the system of College English curriculum development, as shown in Figure 3, *needs analysis* is often a missing component. The assumption that College English learners' levels and needs are the same or similar, however, may seriously affect decisions on teaching objectives, materials, and methodology. The objectives set for the College English program of a specific institution are often a direct copy of the national teaching objectives or curriculum requirements, and remain unchanged for years. Textbooks or courseware developed by established publishers are usually adopted, occasionally adapted, but rarely created by colleges. In reality, students studying in different universities are at various levels and may have different learning needs. Even those from the same university are likely to have diversified educational backgrounds and will be working in different fields. With respect to *testing*, placement tests are used in some universities to put students into classes of different levels upon entering the College English program. School-based exams are administered for achievement purposes. But the results of these institutional tests mainly serve the purpose of order-ranking the students and providing them with external motivation.

**FIGURE 3**  
College English Curriculum Development and Implementation  
(adapted from Jin, 2008, p. 61)



The point to be made here is that College English curriculum development lacks the essential components of *testing* and *evaluation* to

link all the components of the system into a meaningful whole. The absence of these two components has been filled in by school-based exams and the national test. However, as mentioned above, school-based exams are not used as a type of formative assessment for developmental purposes; and the CET can only inform College English teaching objectives, material development and classroom teaching and testing at a very general level. As a basically norm-referenced proficiency test designed for summative purposes, the CET is inadequate when employed by individual institutions for assessment and evaluation of the program. On the one hand, the test could not possibly take into consideration the contextual background of a specific institution, and therefore may not be particularly tailored to the teaching and learning needs of the institution and may not adequately inform its teaching objectives. On the other hand, the test lacks sensitivity to the learning gains of individual learners and is insufficient for diagnosing individual learners' difficulties and deficiencies. As a result, the CET may not lend itself to decisions on such essential elements in the curriculum of College English as the analysis of students' learning needs, the method of classroom instruction, the adaptation or creation of teaching materials, and may not support students' individualized and autonomous learning, which is stressed in the recently implemented national curriculum requirements (Department of Higher Education, 2007). What is urgently needed for a healthy development of College English education, therefore, is the assessment of a formative nature designed for improving teaching and learning and for informing decisions on curriculum development.

### **College English Assessment: The Status Quo**

To have a clear idea of the status quo of formative assessment in College English education, eleven journal articles reporting the empirical studies on formative assessment conducted as part of the College English Reform Project<sup>3</sup> were reviewed and summarized with respect to the assessment instruments employed, participants involved and effects on teaching and learning (See Table 2).

**TABLE 2****A Summary of Eleven Studies on College English Assessment**

<b>Study</b>	<b>Instrument</b>	<b>Participant</b>	<b>Effect</b>
Cai & Li 2007	CO, SJ/R, QZ, QS, LC, SP, ST	NG	Facilitative
Li 2005	QS, INT, SP; Final score=60%FA+40%ST	175 students	Facilitative
Liang & Gao 2007	CO, INT, S/PA, SJ/R, SP	NG	Facilitative
Qu & Wang 2007	SP, CO, S/PA	74 students	Welcomed by students
Tang 2005	NG	NG	Facilitative
Tang & Zhang 2007	S/PA, SP, SJ/R; Final score=70%ST+20%Oral+10%FA	599 students; 8 teachers	Not significant
Wang 2006	CO, QS, INT, S/PA, SP	NG	Facilitative
Yan & Zhang 2005	CO, QS, ST, INT	Inspectors, program designer, teachers, students	NG
Yang 2006	CO, ASSIG, QZ, SJ/R; Final score= 50%ST+50%FA	243 students	Low dependability
Zhou & Qin 2005	SP, CO, S/PA, ST	78 students	Facilitative
Zou & Cai 2006	SJ/R, SP, S/PA, CO; Final score=70%ST+30%FA	NG	Low dependability

Notes: ASSIG: assignment; CO: classroom observation; FA: formative assessment; INT: interview; LC: learning contract; NG: information not given; QS: questionnaire survey; QZ: quiz; SJ/R: student journal/report; SP: student portfolio; S/PA: student self-/peer-assessment; ST: summative test.

The brief review of these journal articles shows a general consensus among College English practitioners and learners about the necessity of incorporating formative assessment in College English education and the facilitative effects of formative assessment on teaching and learning. Some of the favorable comments are presented below as supporting evidence.

- Consultation-based assessment is more beneficial to students than the CET (Cai & Li, 2007).
- Formative assessment facilitates the implementation of the

- computer-and classroom-based teaching model (Li, 2005).
- Multi-method formative assessment is facilitative to improving students' multiliteracies (Liang & Gao, 2007).
- A systematic and comprehensive evaluation model facilitates College English curriculum development and implementation (Tang, 2005).
- The multi-dimensional online formative assessment system facilitates teaching and learning, especially students' autonomous learning (Wang, 2006).
- A dynamic process-oriented systematic evaluation system ensures high teaching quality (Yan & Zhang, 2005).
- Formative assessment stimulates learners' motivation and interest, makes it possible for teachers to monitor students' learning and facilitates students' autonomous learning (Zhou & Qin, 2005).
- Formative assessment cultivates learners' autonomous and cooperative learning and improves their communicative skills (Zhou & Qin, 2005).

The instruments frequently adopted in formative assessment include classroom observation (CO), student portfolio (SP), student journal or report (SJ/R), student self- and peer-assessment (S/PA), classroom quizzes (QZ), questionnaire survey (QS), interview (INT), and so on. Multiple methods were reported to have been employed for data collection and analysis and individualized feedback was reported to have been provided in most of the studies. Suggestions on how to improve the effectiveness of formative assessment included:

- Consulting and involving students are essential in formative assessment (Cai & Li, 2007).
- Technical support is essential to the success of the evaluation model for computer-assisted instruction (Qu & Wang, 2007).
- It is essential to involve all participants and employ multi-methods in formative assessment (Tang, 2005).
- More attention is needed for qualitative evaluation for developmental purposes (Tang, 2005).

- It is necessary to first establish learning objectives, and to systematically collect and analyze quantitative and qualitative data (Tang & Zhang, 2007).
- It is necessary to design a systematic formative assessment system and have logistic support for implementation (Tang & Zhang, 2007).
- Measures such as inspection by university administrators and setting up an evaluation centre help standard setting and implementation (Yan & Zhang, 2005).

Formative assessment, as reported in the articles, however, was rarely designed as part of the systematic evaluation of College English curriculum development. Except for Tang (2005) and Yan and Zhang (2005), the remaining studies investigated classroom assessment of students, leaving behind the other equally important curriculum components like the analysis of learners' needs, setting the objectives of teaching, and developing teaching materials. With respect to the participants involved, with the only exception of Yan and Zhang (2005), no studies involved education administrators, program designers, and material developers.

It is also disappointing to note that the type of data collected to prove the effectiveness of formative assessment in these studies were mostly students' and teachers' responses to survey questions and their comments given at interviews; that is, teachers' and students' perceptions, instead of the actual effects of formative assessment that could be proved, were collected and reported. Most of the studies did provide 'hard' empirical evidence to demonstrate the effects of formative assessment. Tang and Zhang (2007) was the only one among the eleven studies that had a proper research design and employed a principled method to investigate the effects of formative assessment on students' performances. Their findings, however, were somewhat discouraging because no significant effects were statistically confirmed of formative assessment on students' performances in proficiency tests. The ineffectiveness of formative assessment was mainly attributed to the lack of experience with assessment techniques on the part of College English teachers. Young teachers were found to be more used to formative assessment, but College English teachers were, in general,

not well trained to conduct assessment of a formative nature and, as a result, not confident about its effects on teaching and learning (Tang & Zhang, 2007). It is also worth noting that a large discrepancy between students' performances in formative assessment and summative tests was found in Yang (2006), which led to the conclusion that formative assessment, as performed by the College English teachers in the study, had low dependability. This was supported by Zou and Cai (2006), which also mentioned that formative assessment was extremely time-consuming to implement.

## **COLLEGE ENGLISH EDUCATION: FROM TESTING TO ASSESSMENT**

### **Efforts to Revise the CET**

Like any other high-stakes language test, the CET is a double-edged sword. We gain the benefits of providing an accurate description of students' proficiency levels and an extrinsic motivation to improve student performances in the test, but we have to pay the price for its negative impact on teaching and learning as a result of overuses of the test. Facing the challenges, the government has been pressed to promote stakeholders' awareness of sharing and taking the responsibility of preventing overuses of the test, and the NCETC, the organization appointed by the Department of Higher Education, the Ministry of Education, to be in charge of the design and development of the CET, has also been pressed to justify the test design.

In the past twenty-odd years, the NCETC has been making continuous efforts to revise the content and format of the CET in the hope of avoiding the so-called "construct under-representation" and "construct-irrelevant variance" (Messick, 1996) and improving the test's washback and impact (Jin, 2006, 2009). The CET scoring scale and score reporting method have also been adjusted to encourage ethical uses of the large-scale test. What follows is an overview of the content, the format and the scoring system of the CET at the key stages of its development (See Tables 3 and 4).

**TABLE 3**  
**Test Content and Format at the Key Stages of the CET Development**

Stage	Content	Format
PB CET: 1987-1996	Listening comprehension	20% MCQ
	Reading comprehension (careful reading)	40% MCQ
	Vocabulary and structure	15% MCQ
	Cloze or Error correction	10% MCQ, Error identification & correction
	Writing	15% Guided essay writing
PB CET: 1997-2005	Listening comprehension	20% MCQ, Dictation
	Reading comprehension (careful reading)	40% MCQ, SAQ, Translation
	Vocabulary and structure	15% MCQ
	Cloze or Error correction	10% MCQ, Error identification & correction
	Writing	15% Guided essay writing
CET-SET: 1999-now	Warm-up activities	- Question and answer
	Individual presentation and group discussion	- Presentation and discussion
	Further-check questions	- Question and answer
PB CET: 2006-now	Listening comprehension	35% MCQ, Dictation
	Reading comprehension (reading in depth)	25% MCQ, Banked cloze
	Reading comprehension (fast reading)	10% Sentence completion, True/false, MCQ
	Cloze or Error correction	10% MCQ, Error identification & correction
	Translation	5% Translation (Chinese to English)
IB CET: 2008-now	Writing	15% Guided essay writing
	Listening comprehension	25% MCQ
	Listening-based integrated task: Dictation	15% Dictation



Listening-based integrated task: Listening & repeating	10%	Listening to and repeating sentences
Listening-based integrated task: Writing	15%	Summarizing and making comments
Reading comprehension (reading in depth)	20%	MCQ
Reading comprehension (fast reading)	10%	MCQ, Sentence completion
Grammar & structure	5%	Blank filling

Notes.

1. MCQ: multiple choice question; PB CET: paper-based CET; IB CET: internet-based CET.
2. The CET-SET adopts a graded scoring system and issues a certificate indicating the test-taker's final grade (A+, A, B+, B, C+, or C).
3. The IB CET-4 trial test was first implemented in June 2008, so the test content and format are subject to change based on the results of further trial implementations.

**TABLE 4**  
Scoring System and Score Report at the Key Stages  
of the CET Development

	<b>PB CET (1987-2005)</b>	<b>CET-SET (1999-now)</b>	<b>PB CET (2006-now)</b>	<b>IB CET (2008-now)</b>
Scoring system	Norm-referenced; Norm: 6 top universities Mean=72, Sd.=12	Graded: A+, A, B+, B, C+, C, D	Norm-referenced; Norm: 16 upper-middle universities Mean=500, Sd.=70	Norm-referenced; Norm: to be established Mean=500, Sd.=70
Score report	Certificate indicating pass or distinction	Certificate indicating grade and grade descriptions	Score report: total and profile scores(listening, reading, cloze, writing and translation)	Score report: total and profile scores(listening, integrated tasks, reading)

Notes. IB CET: Internet-Based CET ; PB CET: Paper-Based CET.

The guiding principles of the test revision at each stage were to respond to the changing requirements of College English teaching on the one hand, and to reflect more closely the achievements made in the

field of language testing on the other. It can be seen that with a better understanding of the construct of communicative competence and a better knowledge of the theory and practice in language testing, the CET has been moving from a test of isolated language skills using discrete-point objective items towards a performance test of integrated skills with an increasingly higher percentage of constructed-response items.

Progress has also been made in recent years in the application of modern information and communication technology to improve the test's validity and fairness, which are, in Bachman's words, "issues at the heart of how we define ourselves as professionals, not only as language testers, but also as applied linguists" (Bachman, 2000, p. 25). With a view to further improving the authenticity of CET test tasks, the trial test of the IB CET-4, for example, uses audio and video clips of radio and TV programs as input materials for testing listening comprehension. Integrated tasks (dictation, listening and repeating, listening-based integrated writing) constitute 40% of the total score in the IB CET-4.

With respect to the CET scoring system, certificates were issued by the Department of Higher Education, the Ministry of Education, to those who passed the test from 1987 to 2005. As a measure to discourage overuses of the CET, the 18-year practice was abandoned in June 2005, and certified score reports replaced the pass-or-fail certificates. Both the total and profile scores are provided to test-takers to help them identify their strengths and weaknesses. In addition, a new score scale with a mean of 500 and a standard deviation of 70 was adopted to prevent the misinterpretation of the 60-point passing score in the traditional one-hundred-point score scale.

## **Suggestions to Improve College English Assessment**

### *A Framework of Systematic Assessment*

Merely revising the content and format of the CET and advocating shared responsibility for the consequences of the uses of the test, however, may not necessarily affect what is going on in College English classrooms. A constructive way out of the dilemma facing CET designers and College English teachers and learners, as argued in this

part, is to design and incorporate systematic assessment into the curriculum system. As part of regular curriculum activities, systematic assessment as proposed in the framework in this part should play a central role in linking the key components of the College English curriculum. The framework shown in the form of a table details the participants, purposes, contents, and instruments of a systematic and comprehensive assessment of College English teaching and learning (see Table 5).

**TABLE 5**  
**A Framework of Systematic Assessment**  
**for College English Curriculum Development**

Participant	Purpose	Content	Instrument
National or provincial level educational authorities	Provide guidance to policy-making Supervise policy implementation	Social needs for English language competence The status quo of teaching and learning Students' general English proficiency Suitability of curriculum requirements	To be employed by different groups of participants for different purposes: questionnaire survey group or individual interview
Institutional level administrators	Supervise curriculum design Support curriculum implementation	Students' entry level and learning needs Level and experience of teachers Adequacy of teaching facilities Suitability of teaching objectives Achievement of teaching objectives	document analysis classroom observation classroom quiz student journal or report teacher journal or report student portfolio
College English program designers	Design and improve the curriculum Improve teaching and learning Evaluate teaching materials Improve assessment and evaluation	Suitability of curriculum design Suitability of teaching models/methods Organization of teaching activities Suitability of teaching materials Effectiveness of assessment and evaluation	peer-assessment summative test

College English teachers	Tap into the teaching process Control the teaching quality Reflect on and improve teaching Observe and supervise learning Encourage and motivate learners	Understanding of teaching objectives Planning of teaching activities Teaching attitude Teaching methodology Classroom assessment Teaching outcomes
College English learners	Tap into the learning process Demonstrate learning progress Check learning outcome Reflect and improve learning	Understanding of learning objectives Planning of learning activities Learning attitude and motivation Learning skills and strategies Learning outcomes

Note: For a summary of the methods that can be used for a systematic evaluation of language programs, see Weir and Roberts (1994, p. 26-31).

To effectively inform and promote the design and implementation of College English curriculum, College English assessment, as proposed in the framework, should follow a set of guidelines which are formulated and presented below.

- 1. Contextualize assessment activities with the specific situation of a particular institution, including such aspects as support from administrators, human and financial resources available, teachers' knowledge and skills of assessment.
- 2. Address each key component of the curriculum, including the analysis of students' learning needs, defining the teaching objective, developing teaching materials, and the process of teaching and assessment per se.
- 3. Adopt a systematic procedure that standardizes the process of assessment into discrete phases of data collection, analysis, communication of feedback information, and decision-making on revisions to teaching and curriculum design and implementation.
- Involve all parties of stakeholders of College English

저석 있음: 글머리 기호 및 번호 매기기

education, from national-level educational authorities to teachers and learners, each playing a different role in assessment.

• Employ a variety of instruments such as survey, interview, group discussion, document analysis, classroom observation, student report, self- or peer-assessment, and so on.

• Collect information of different types, both quantitative and qualitative.

• Provide immediate and informative feedback to stakeholders concerned.

• Last but not least, the information gathered by each group at every stage should feed back into the curriculum system to enable constant revisions and adjustments of each component of the system.

서식 있음: 글머리 기호 및 번호 매기기

#### *Assessment for Learning and Criterion-Referenced Testing*

When *testing* and *assessment* are differentiated, the ideology of ‘assessment for learning’ lies at the heart of the differentiation. Assessment for learning is defined as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” (Assessment Reform Group, 2002). Assessment for learning takes place throughout the teaching and learning process and the information gained from assessment activities should be used to shape the process. For assessment to be a regular part of teaching and learning, assessment for learning should be regarded as a key professional skill for teachers, and teachers should be supported in developing these skills through initial and continuing professional development (Qualifications and Curriculum Development Agency, 2007).

An encouraging finding from the review of the eleven journal articles in the paper is that these teachers recognize the value of assessment for learning and have made conscious efforts to put into practice what they understand as formative assessment. In reality, however, the majority of College English teachers are deficient in knowledge and skills, and necessary power and authority, to design and implement systematic

assessment for improving College English education. The proposed framework and the set of guidelines suggested, therefore, serve as a principled procedure to put the ideology of assessment for learning into the practice of College English education. However, for College English assessment to achieve its purposes, educational authorities need to be fully aware of the central importance of assessment for learning, and College English teachers need to be trained for professional knowledge and skills in performing effective assessment.

It is also worth noting that *testing* in Brown's (2001) model for designing and maintaining a language curriculum refers mainly to criterion-referenced testing (CRT), which is most useful to classroom teachers and curriculum developers. Brown and Hudson (2002) argued that "CRTs are specifically designed to assess how much of the content in a course or program is being learned by the students" (p.xiv). To incorporate formative assessment into College English teaching and learning, by necessity, involves setting up detailed criterion-referenced standards. These standards would form the basis of defining and refining teaching objectives and provide guidance to the development of teaching materials. None of the eleven studies reviewed in this paper, however, attempted to set up such kind of standards and implement them in the design of assessment tasks. To some College English teachers, formative assessment may simply be equated to allocating a certain percentage of weighting to students' class attendance and their performance in quizzes or assignments in the decision of their final grades.

CRT is also useful for the development of large-scale testing. With clear-cut standards, test specifications could describe the knowledge, skills and abilities being tested in explicit terms. Score reporting forms could provide information on what learners at a certain level can do. With well-defined standards, educators could design assessment instruments and conduct assessment on a sample of learners instead of relying on full-cohort testing for decision making. Large-scale testing is a very challenging task, for test designers as well as educational authorities. In this respect, testing organizations could draw on the experience of projects like Programme for International Student Assessment (PISA) and the National Assessment of Educational Progress (NAEP), both representing efforts of testing organizations to

make use of assessment for the purpose of improving teaching and learning (Wang, 2007).

## CONCLUSION

The College English Test, currently the largest English as a Foreign Language (EFL) test in the world, enjoys a good reputation in the language testing circle both at home and abroad. However, the test has been overburdened in recent years with uses for unjustified purposes. The issue of its negative washback on teaching and learning as a result of the high-stakes uses of the test and the over-reliance of the curriculum development on the test discussed in this paper highlights the central importance of incorporating process-oriented formative assessment in the design and implementation of College English curriculum for improving teaching and learning. To this end, a framework of systematic assessment was proposed in the paper, which details the participants, purposes, contents and instruments of College English assessment activities. The implication of the argument is that if systematic formative assessment could become a regular part of teaching and learning, College English education would be less reliant on large-scale testing and the CET would be decoupled from the multiple purposes not intended in its original design. More importantly, College English teachers would adopt a more rational attitude towards the CET, and eventually be relieved of the imperative to pursue good CET scores at all costs, which, in return, would ensure a sustainable development of both the College English program and the CET. As advocated by the International Language Testing Association (2000) in its Code of Ethics, “(L)anguage testers in their societal roles shall strive to improve the quality of language testing, assessment and teaching services, promote the just allocation of those services and contribute to the education of society regarding language learning and language proficiency.” This paper represents a major step taken by the designers of the CET in this direction.

## NOTES

1. While *college education* often refers to higher education of 3-year vocational-technical training, no distinction is made between *college* and *university* in this paper. In most places, *college* is used to refer to all types of tertiary level education in China, including 4-year university education, in order to conform to the use of the word in the title of the program *College English*.
2. For tertiary level education in China, there are two types of English language programs: English major programs (for students majoring in English language and literature) and non-English major programs (for students majoring in other disciplines). Non-English college majors are required to take a total of about 240-280 hours of College English classroom instruction during the first two years and earn an average of 12 to 16 credits.
3. The project was launched in 2002 by the Department of Higher Education, the Ministry of Education. The three major tasks of the project were to 1) define the objective of College English education in accordance with the social needs for college and university students in the new century, 2) establish a new model of College English education that makes full use of modern information technology, and 3) revise the College English Test for better washback and impact.

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Yang Zhizhong from Nanjing University for his careful reading and editing of the manuscript.

## REFERENCES

### Publications in English

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Angelo, T. A. (1997). Reassessing and defining assessment: a second try. In Bonwell, C. C. (Ed.), *Using active learning as assessment in the*



- postsecondary classroom* (pp.73-74). Washington, DC: Clearing House.
- Askham, P. (1997). An instrumental response to the instrumental student: assessment for learning. *Studies in Educational Evaluation*, 23(4), 299-317.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Retrieved August 15, 2008, from the World Wide Web: <http://www.assessment-reform-group.org.uk>.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- Black, P. J., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Brown, J. D. (2001). *The elements of language curriculum: A systematic approach to program development*. Beijing: Foreign Language Teaching and Research Press.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.
- Cheng, L., & Curtis, A. (Eds.). (2009). *English language assessment and the Chinese learner*. London: Routledge.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gu, X. (2004). *Positive or negative? An empirical study of CET washback on College English teaching and learning in China*. Unpublished doctoral dissertation. Shanghai Jiao Tong University, Shanghai.
- International Language Testing Association. (2000). *Code of ethics for ILTA*. Adopted at the annual meeting of the International Language Testing Association, Vancouver, British Columbia, Canada.
- Jin, Y. (2009). The national college English testing committee. In Cheng, L. & Curtis, A. (Eds.), *English language assessment and the Chinese learner* (pp.44-59). London: Routledge.
- Jin, Y., & Yang, H. (2006). The English proficiency of college and university students in China: As reflected in the CET. *Language, Culture and Curriculum*, 19(1), 21-36.
- Li, X., & Zeng, Y. (2002). *Language testing and teaching: Is the tail wagging the dog?* Paper presented at the First International Conference on English Language Testing in China, Shanghai.
- Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge: Cambridge University Press.
- Messick, S. (1996). Validity and washback in language testing. *Language*

- Testing*, 13, 241-256.
- Qualifications and Curriculum Development Agency. (2007). *The 10 principles: Assessment for learning*. Retrieved July 20, 2009, from the World Wide Web: <http://www.qcda.gov.uk/4336.aspx>.
- Shohamy, E. (1997). Testing method, testing consequences: Are they ethical? Are they fair? *Language Testing*, 14(3), 340-349.
- Shohamy, E. (2001a). The social responsibility of the language testers. In R. Cooper, E. Shohamy & J. Walter (Eds.), *Language education, volume in honor of Bernard Spolsky* (pp.113-130). Amsterdam: John Benjamins Press.
- Shohamy, E. (2001b). *The power of tests: A critical perspective on the use of language tests*. New York: Longman.
- Stiggins, R. (2002). Assessment crisis: the absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28, 499-509.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge: Cambridge University Press.
- Weir, C. J., & Roberts, J. (1994). *Evaluation in ELT*. Oxford: Blackwell.
- Zheng, Y., & Cheng, L. (2008). College English Test (CET) in China. *Language Testing*, 25(3), 408-417.

#### Publications in Chinese

- Cai, K., & Li, X. (2007). Consultation-based assessment for College English teaching. *Higher Education Research*, 23(1), 59-60.
- Department of Higher Education, Ministry of Education. (2007). *College English curriculum requirements*. Shanghai: Shanghai Foreign Language Education Press.
- International Language Testing Association. (2000). *Code of ethics for ILTA*. Retrieved July 20, 2009, from the World Wide Web: <http://www.iltaonline.com/code.pdf>.
- Jin, Y. (2000). Backwash effect of CET-SET on teaching EFL in China. *Foreign Language World*, 4, 56-61.
- Jin, Y. (2006). On the improvement of test validity and test washback: The CET washback study. *Foreign Language World*, 6, 65-73.
- Jin, Y. (2008). On the reform of College English: The role of assessment and evaluation. *Foreign Language Education in China*, 1(3), 57-66.
- Jin, Y., & Wu, J. (1998). Examining the validity of CET reading comprehension by introspection. *Foreign Language World*, 2, 47-

52.

- Li, C. (2005). An empirical study of the evaluation model for College English internet-based instruction. *Foreign Languages and Their Teaching*, 7, 33-36.
- Liang, A., & Gao, H. (2007). On multi-method assessment for College English teaching. *Shandong Foreign Language Teaching Journal*, 4, 68-71.
- Liu, R. (2007). *Liu Runqing's anthology on English language education*. Beijing: Foreign Language Teaching and Research Press.
- Qu, X., & Wang, J. (2007). Research and practice of a new evaluation model for College English CAI. *China Educational Technology*, 249, 83-86.
- Sun, Y. (1985). *Collected works of Sun Yat-sen, Vol. 5*. Beijing: Zhong Hua Book Company.
- Tang, X. (2005). College English Test and the evaluation of College English teaching. *Foreign Language Education*, 26(1), 56-59.
- Tang, X., & Zhang, S. (2007). New forms of assessment in College English teaching: Implementation and problems. *Foreign Languages and Their Teaching*, 1, 14-19.
- Wang, H. (2008). A systems approach to the reform of College English Testing: report on the 'Survey of College English Testing Reform'. *Foreign Language in China*, 5(4), 4-12.
- Wang, L. (2007). *Exploration in educational assessment and evaluation*. Xi'an: Xi'an Jiao Tong University Press.
- Wang, Y. (2006). Research of the evaluation system for the internet-based College English teaching. *Foreign Language World (Supplement)*, 96-107.
- Working Group on College English Teaching Syllabus. (1985). *College English teaching syllabus (For college and university students of science and technology)*. Shanghai: Shanghai Foreign Language Education Press.
- Working Group on College English Teaching Syllabus. (1986). *College English teaching syllabus (For college and university students of arts and science)*. Shanghai: Shanghai Foreign Language Education Press.
- Working Group on College English Teaching Syllabus. (1999). *National College English teaching syllabus (revised version)*. Shanghai: Shanghai Foreign Language Education Press.
- Wu, Q. (2005). *Reform of the CET*, Second news conference of the Ministry of Education. Retrieved February 25, 2005, from the

World Wide Web: <http://www.moe.edu.cn>.

- Yan, M., & Zhang, X. (2005). Teaching process research on College English quality appraisal. *Heilongjiang Researches on Higher Education*, 137(9), 131-133.
- Yang, H., & Gui, S. (2007). Language testing from a sociological perspective. *Modern Foreign Languages*, 30(4), 368-374.
- Yang, H., & Weir, C. J. (1998). *The CET validation study*. Shanghai: Shanghai Foreign Language Education Press.
- Yang, X. (Ed.). (1999). *History of examination in China*. Beijing: Higher Education Press.
- Yang, Z. (2006). Research on the practice of formative assessment in College English. *China Science and Technology Information*, 23, 260-261.
- Zhou, P., & Qin, X. (2005). The application of formative assessment in multimedia-assisted language learning. *Computer-Assisted Foreign Language Education*, 105, 9-13.
- Zou, H., & Cai, Z. (2006). An experiment on implementation of formative assessment and the reform of the current assessing system of college English teaching. *Journal of Jiangxi Normal University (Social Science)*, 39(1), 114-118.