

POWERFUL TESTS ,POWERLESS TEST DESIGNERS ? — CHALLENGES FACING THE COLLEGE ENGLISH TEST

Jin Yan

Shanghai Jiao Tong University

Abstract

The dilemma facing language testers ,especially designers of large scale high stakes tests ,is that the better the professional quality of a test ,the greater the possibility of its being misused . The higher the stakes of a test ,the more powerful it will become ,and the less power the testers will have over its uses . In other words ,test designers often feel increasingly powerless to influence the way a test is used or to prevent misuses of the test as the stakes of the test becomes higher . Our experience with the College English Test (CET) ,the largest EFL test in the world ,shows that designers of large scale tests are constrained by factors beyond their control despite their good intentions . The perplexing situation echoes the growing awareness in the field of language testing that washback does exist and that it is a hugely complex matter . This paper will address the new challenges facing the CET ,i. e. to revise the test design on the one hand ,so as to maintain the professional standards of the test ,and to improve the washback of the test on the other hand . The view to be argued for is that the power of a high stakes test must be harnessed in order for the test to positively impact teaching and learning . To achieve the goal ,every stakeholder has a role to play . It is important and urgent ,for example ,for the CET designers to make the intended uses and major characteristics of the test transparent to its users through detailed documentation on what is measured and explicit dissemination of the standards ,content outline ,theoretical framework ,and score interpretation of the test .

Key words

College English Test ;challenges ;test washback

1 . Introduction

Testing has a long history in China and has always played an immensely important role in education ,in politics ,and in society as a whole . In the preface to the book series *The History of Testing in China* (Yang 2003) ,China was acclaimed as “the home country of testing” ,and the contribution of China's imperial examinations system to human civilization was regarded to be as important as that of the four greatest inventions of China . Dr . Sun Yat sen (1985) discussed the imperial examinations system from a global perspective and considered it the oldest and best testing system in the world .

In the imperial examinations system ,there were so many levels of examinations that candidates had to spend years ,even their lifetime ,climbing up step by step the ladder leading to fame and wealth . As the test results could have a profound effect on the lives of these people and their families ,candidates sometimes attempted to cheat at the risk of a severe penalty or even capital punishment . Pictures taken in an Imperial Examinations Museum located in a Confucius Temple in the suburb of Shanghai show various types of evidence of cheating and penalties for cheating . Where there is a test ,there is cheating ,just like the flip side of the same coin . And this is even more true for such high stakes tests like the imperial examinations .

In present day China ,testing is still a very powerful tool used for making decisions of various kinds : key high school entrance ,college or university entrance , MA or Ph .D . program entrance ,language program exit ,and qualification requirement ,to name but a few . And China is also one of the major markets for a number of international language tests ,for example ,TOEFLiBT ,TOEIC ,IELTS ,BEC , BULLETS ,and so forth .

Wellintended as they are to serve as “ a catalyst for curriculum innovation ,a strategy to promote curricular innovation ,or a mechanism to drive instruction ” (Andrews 2004 :39 42) ,high stakes tests could have detrimental effects on individuals “ as they can create winners and losers ,successes and failures ,rejections and acceptances ” (Shohamy 2001 :113) . The most serious consequence of this unwelcoming aspect of testing is that teachers are likely to be disempowered for the very simple reason that they no longer have much control over what to teach and how to teach .

What concerns language testers in China more than those in other parts of the world is that most of our tests are large scale tests ,“large scale ”in its truest sense . For instance ,as many as 9 .5 million high school graduates took the national college and university entrance examination in 2006 . The College English Test (CET) was given to 12 million college and university students in 2006 . Designed as an optional test to measure the English language proficiency of students in tertiary institutions ,the CETis now taken by almost every college and university non -English -major student ,that is ,students who major in arts and humanities ,science and technology ,medical and life science ,and so forth . This is due in considerable part to the increasing recognition of the test by the institutions and by society .

In this paper ,the development of the CET in the past two decades will be briefly reviewed to contextualize the growing power of the CET . This will be followed by a discussion on the challenges and dilemmas facing the designers and administrators of this powerful test and the measures that have been taken by the test designers to tackle the power of the CET . Finally ,the paper will conclude with some thoughts on the future development of the test .

2 . An Overview of the CET :Contextualizing the Powerful Test

The CET was launched in 1987 ,immediately following the implementation of the first national College English Teaching Syllabus in the mid 1980s (State Education Commission 1985 ,1986) . The intended purposes of the test were to provide an objective evaluation of a student s overall English proficiency on the one hand and positively impact EFLteaching at the tertiary level in China on the other hand . Since its inception ,the CET has been revised several times to reflect and respond to the increasingly higher requirements of college and university students brought about by China s continuous opening up and reform policy (see Jin 2004 ,2005 ,2006 ;Jin & Yang 2006 ;Wu 2005 ;Yang & Weir 1998) .

A number of new test formats and tasks were introduced to the CETin the mid 1990s ,and the CET Spoken English Test (CET -SET) was started nationwide in 1999 when the first revised version of the national College English Teaching Syllabus (Ministry of Education 1999) was promulgated . The changes made to the CET test content and format and its score reporting system in recent years were meant to respond to the second major revision to the national teaching syllabus ,now the College English Curriculum Requirements (for trialimplementation) (Ministry of Education 2004) .

Through its 18 years of development ,the CET has had a positive impact on English language teaching at the tertiary level on the whole and has gained recognition among university teachers and administrative authorities . CET certificates have become a nationally recognized credential for employment of college and university graduates . Alongside the growth of the test ,the College English Testing Committee (hereafter referred to as the CET Committee or the Committee) ,which is the academic organization undertaking the mission of designing ,developing ,and revising the large scale test ,has become a professional team of English language testing . The CET Committee members are appointed by the Higher Education Department of the Ministry of Education and consist of college English teachers from a number of universities across the country who specialize in applied linguistics . The committee members work parttime for item writing and revision ,standard setting ,training of markers and examiners ,quality control of essay marking ,score equating and reporting ,test revision ,and so forth .

Over the years ,the Com mittee has been playing an increasingly important role in the field of language testing in China and in Asia . Friendly relationships have been established between the CET Com mittee and other language testing and assessment groups in China and in countries and areas in Asia and around the world . Regular academic seminars ,forums ,and conferences are organized or hosted by the CET Com mittee ,and its members frequently present at national and international conferences on language testing and assessment (see Jin ,in press) .

With the increasing recognition and use of the CET certificate ,for purposes intended or not intended ,the CET has become a test of extremely high stakes . The results of the test can determine ,for example ,whether college or university students can graduate or get a bachelor s degree ,whether they can find a good job ,or even whether they are eligible to be granted a residence permit in some major cities . The CET has been receiving more and more coverage ,positive and negative ,in the media . Test security is given the most serious attention by the groups of people involved in test paper printing ,delivery and storage . High tech cheating often becomes headline news after tests have been administered . University teachers were reported to have been charged for cases involving organized stealing of the test paper for making profits or organized cheating for their students better performances .

On top of the administrative concerns for test security and test fairness ,what has become a real challenge facing CET designers is teaching to the test (Li & Zeng 2002) . Coaching materials are sometimes used in classes by teachers to replace normal classroom teaching and by students for after class self learning ,leading to narrower teaching and learning content . College English teaching hours are not required in some universities after their students pass the test . Students are often satisfied with a narrow pass and stop learning English after they get a certificate . As the designers of the test ,we have to ask ourselves the question : What is the point of a powerful language test if the test cannot have a beneficial and happy effect on millions of EFL learners ?

3 . Major Revisions to the CET :Meeting the Challenges

All around the world today ,the number of English language learners keeps increasing . It is estimated that ,within two decades ,the number of people who speak English as a second or foreign language might well exceed the total number of native speakers of English in the world . The majority of these EFL learners are very likely to take some kind of English language tests in the process of learning . Therefore ,it is of utmost importance that the power of these language tests be controlled and the tests be used in an appropriate manner so that testing will serve to complement and promote teaching and learning . In the case of the CET ,continuous efforts have been made to address the issue of social consequences and impact on teaching and learning brought about by the power of the test . (See Figure 1 for a summary of the three stages of major developments of the CET in the past two decades .)

In the 1980s and early 1990s ,with a surging interest in educational and psychological measurement and testers over riding concern with marking reliability ,large scale tests in China and around the world employed mainly objective item formats . In the first phase of the development of the CET ,85 % of the items in the CET Band 4 (CET 4) were multiple choice questions ,the most typical type of selected - response items . Essay writing ,contributing to the remaining 15 % ,was the only component testing productive skills . In the CET Band 6 (CET 6) ,a higher level of the CET tests ,essay writing and editing (error identification and correction) constituted 25 % of the test .

As a type of objective test format with obvious advantages of marking efficiency and reliability ,multiple choice questions are the most popular format employed by almost all large scale standardized tests . The format has been proven especially useful and valid for testing reading comprehension ,listening comprehension ,knowledge of vocabulary and grammar ,and so forth . The attraction of the item type is outweighed ,however ,by its equally evident limitation ,i. e. ,incapability of testing productive skills .

From 1996 onwards ,a series of measures were taken by the CET Com mittee to remedy inadequate assessment of students productive skills as a result of over reliance on objective item formats . A variety of constructed response item types were introduced ,including ,for example ,short answer questions ,compound dictation tasks (dictation for words ,phrases ,clauses or sentences) ,and translation from English to Chinese tasks . To prevent the pursuit of a high pass rate by teachers and a narrow pass by

students ,the Average Graded Scoring system was developed to take into account the different levels of performances of those students who pass the test , which provides a more accurate evaluation of the overall performance of a group of students and proves to be a much more scientific indicator than the simple pass rate . To further encourage the improvement of students writing ability , a minimum score requirement was set for the writing component . Failing the writing part means a deduction of certain points from the total score ,and a zero in the writing part fails the test taker on the whole test no matter how well he or she performs in the other parts .

			STAGE 3
2007 onwards			Begin the project of the web based CET ; Improve the test washback
2006			Revise the test content and format
2005			Design and adopt a new score reporting system
		STAGE 2	
1999		Launch the CET Spoken English Test	
1998		Set a minimum requirement for WRITING	
1997		Introduce new item forms ; Report the average graded scores	
	STAGE 1		
1990	Use a separate paper for testing WRITING		
1985-1989	Design the test , Establish the norm		

Figure 1 . The development of the CET in the past two decades

The introduction of the CET Spoken English Test (CET SET) in 1999 was a milestone in the history of the CET development . The test is given in the authentic face to face interview format and employs interactive tasks such as answering the interlocutor's questions and group discussion among the three test takers . The training of the CET SET examiners and the process of establishing over 50 CET SET test centers in 35 major cities are in themselves effective means to impact the teaching and learning of EFL oral communication skills and strategies , which was much neglected in the 1980s and early 1990s (Jin 2000) .

Entering the 21st century , the Higher Education Department of the Ministry of Education saw the necessity of taking a step further to improve the quality of college and university teaching in order to better cater to the needs of the country's further reform and opening up policy . As part of the higher education reform package , the 1999 version of the College English Teaching Syllabus was revised and the National College English Curriculum Requirements were put into trial implementation in 2004 (Ministry of Education 2004) . To comply with the new requirements , the CET Committee revised the test content and introduced some new formats . The latest version of the CET (National College English Testing Committee 2006a , 2006b) includes new tasks such as comprehension of longer conversations , fast reading (i. e. skimming and scanning) , understanding the contextual meaning of vocabulary in a reading passage (banked cloze) , translation from Chinese to English , and so forth . Constructed response items contribute to 30-45 % of the total score . More emphasis has been given to the listening component , the percentage

rising from 20 % to 35 %. More attention is paid to contextualized language use instead of context free knowledge of language . The traditional component of vocabulary and structure , which used to be assessed in the format of multiple choice questions in the context of single sentences , was removed and replaced by the task of banked cloze , which requires not only knowledge of vocabulary and structure but also the skill of inferring contextual meanings of vocabulary at the discourse level . Table 1 provides a detailed summary of the revisions made in the past two decades to the CET test content and format . (For various versions of the CET and CET SET test syllabuses , please see College English Test Design Group 1987 , 1989 , 1994a , 1994b ; National College English Testing Committee 1999 , 2006a , 2006b)

Table 1 . Revisions to the CET Test Content and Format

	1987-1995	1996-2004	Since 2005
Test content	1 . Listening 20 % 10 short conversations 3 short passages	1 . Listening 20 % 10 short conversations 3 short passages OR 1 passage for dictation	1 . Listening 35 % 8 short conversations 2 long conversations 3 short passages 1 passage for dictation
	2 . Reading 40 % 4 short passages for in - depth comprehension	2 . Reading 40 % 3 short passages for in - depth comprehension	2 . Reading 35 % 2 short passages for in - depth comprehension
	3 . Vocabulary and structure 15 % 15 sentences for vocabulary recognition 15 sentences for structure recognition	1 passage for SAQs OR 5 sentences for translation from English to Chinese	1 passage for vocabulary comprehension 1 long passage for skimming and scanning
	4 . Cloze / Error correction 10 % CET 4 : 1 cloze passage ; CET 6 : 1 passage for error correction	3 . Vocabulary and structure 15 % 20-30 sentences for vocabulary recognition 0-10 sentences for structure recognition	3 . Cloze / Error correction 10 % 1 cloze passage OR 1 passage for error correction
	5 . Writing 15 % A short essay of 100 (CET 4) / 120 (CET 6) words in 30 minutes	4 . Cloze / Error correction 10 % CET 4 : 1 cloze passage ; CET 6 : 1 passage for error correction	4 . Writing and translation 20 % A short essay of 120 (CET 4) / 150 (CET 6) words in 30 minutes
Test format	75-85 % MCQs 15-25 % Writing (and constructed response items)	5 . Writing 15 % A short essay of 100 (CET 4) / 120 (CET 6) words in 30 minutes	5 sentences for translation from Chinese to English
		55-75 % MCQs 25-45 % Writing and constructed response items	55-70 % MCQs 30-45 % Writing and constructed response items

A most important decision made by the CET Committee in June 2005 was replacing the CET certificate with a score report form . Basically a norm referenced test , the CET established the norm for CET 4 in 1987 and for CET 6 in 1989 . After score equating and normalization , test takers used to be given a certificate indicating whether they had passed the test (60 points or above on the hundred point score scale with a mean of 72 and a standard deviation of 12) or had received a “ distinction ” (85 points or above) . Since June 2005 , the scores have been reported on a 710 score scale with a mean of 500 and a standard deviation of 70 . Instead of a CET certificate , test takers now receive a score report form indicating the overall and profile scores for each component .

The decision was not simply meant to prevent universities from using the CET certificate as an

indispensable requirement for graduation or to discourage employers from making the certificate a prerequisite for employment . Rather ,this was a purposeful approach to encourage the use of CET for EFLteaching and learning purposes and avoid putting undue pressure on the test and its designers . With changes in the score reporting system ,some university officials have loosened the ties between their students performances in the CET and the requirements of graduation or undergraduate degrees . Quite a number of universities have changed the policy completely .

Well meant as these changes may have been ,the sad truth and discouraging news for language testers ,especially designers of large scale tests ,is that “negative effects can result from a properly developed ,correctly targeted test that is being implemented in line with all available knowledge about best practice ” (Hamp -Lyons 1997 :297). Also ,“It should not be assumed that a ‘good’ test will automatically produce good effectsin the classroom ,or that a ‘bad’ test will necessarily produce negative ones ”(Wall 2000 :505 -506). The CET Com mittee s experience of designing and administering the CET in the past two decades has indicated that the better the quality of a test ,the more reputation it enjoys , and the higher the stakes of the test . What comes with the high stakes of a test is the power it possesses . A powerful test ,unfortunately ,is very likely to be over used or misused ,that is ,used for purposes not intended by its designers . The CET ,for example ,has had a powerful influence on what goes on in the classroom . In the meantime ,however ,teachers in some colleges and universities have been deprived to some extent of the power to decide on the teaching material ,teaching content ,or teaching methodology . Under the circumstances ,what should the test designers do ?Is it possible for test designers to take the responsibility for the consequences of the test without being given sufficient power over the use of the test ?

4 . The Way Forward :Stake holders Roles in the Power Battle

Shohamy raised several questions when talking about the dilemmas facing test designers :Is the test developer responsible for uses and misuses of tests ? What is the role of the tester once s/he notices misuses ?Or ,perhaps it is not the responsibility of the tester at all to worry about the test takers after a test had been handed to the users (Shohamy 2001 :115). Alderson explicitly ,though to the disappointment of test designers ,pointed out that “There are limits to what a test developer can achieve , and much more attention needs to be paid to the reasons why teachers teach the way they do ”(Alderson 2004 :xi).

Around the world ,in the field of language testing ,it has been increasingly well recognized that washback is a hugely complex issue . In the early 1990s ,Alderson and Wall (1993)asked the apparently simple yet most intriguing question in their seminal article on washback :Does washback exist ?Over the decade after the paper was published ,Alderson (2004 :ix)explicitly stated that “I believe there is no longer any doubt that washback does indeed exist The question today is not ‘does washback exist ?’ but much rather what does washback look like ?What brings washback about ?Why does washback exist ?’

Wall (2000 :499)pointed out that “There was a growing awareness of the importance of factors other than test design in determining whether tests would have the impact that was desired ”. Language testers seem to have come to the consensus that traditional understanding of test validity is inadequate for the dimension of test use . The growing concern over the social consequences of testing and the greater awareness of testers social responsibilities have given rise to the need for an extension of the concept of test validity to embrace the facet of the test s washback and impact . Messick s (1989 ,1996)unified and expanded theory of validity considers washback as the consequential aspect of test validity . Bachman and Palmer s (1996)test usefulness framework includes impact as one of the six interrelated qualities of language tests ,the other five being test reliability ,construct validity ,authenticity ,interactiveness ,and practicality . In a similar vein ,Kunnan s (2004)test fairness framework also emphasizes the aspect of social consequences ,in addition to the other essential facets like test validity ,absence of bias ,access ,and administration . McNamara (2000)makes a distinction between the macro -and micro level washback effects ,referring to “impact ” as the effects of tests at macro levels on education and society and “washback ”as the effects of tests at micro levels on teaching and learning .

It has been well accepted that ,as a complex yet important dimension of testing ,washback deserves

much more attention not only from testers but also from test users as well. Tests, as noted by Shohamy (2001: 113-114), are not neutral but rather embedded in political, social, educational, ideological, and economic contexts. Tests can be used in society as powerful tools for political, social, educational, and ideological agendas.

What is the way forward for the CET in the new century with respect to the improvement of its washback and impact? Messick (1996) considers "construct under representation" and "construct-irrelevant variance" as potential sources of negative washback. Therefore, the CET Committee has continuously put improving the quality of the test, so as to strengthen its intrinsic power, high on the priority list. One recent line of research which is likely to lend strong support to improving the CET is the work on the establishment of a common level framework for EFL learners in China. Encouraged and illuminated by the wide application and acceptance of the Common European Framework of Reference, the research into a framework with explicit descriptions of abilities to be achieved by Chinese EFL learners at various levels is of both theoretical and practical value. With an established level framework, English language teachers and testers in China will be able to build a closer link between the requirements of language teaching and the standards used in testing, which, in our view, may provide a viable and long-term solution to the problem of negative washback of large-scale high-stakes tests on teaching and learning.

Equally important is to directly address the issue of the test's social impact, that is, to deal with the extrinsic power of the large-scale test, which is a more serious challenge to the CET Committee. To achieve this goal, the Committee needs to further investigate in which direction, positive or negative, and in what way the implementation of the CET has impacted college English teaching and learning and China's tertiary-level educational policies. The investigation should start with the most direct stakeholders of the test, that is, college and university teachers and students. Aspects relevant to college English teaching programs, such as curriculum design, material development, implementation of teaching, test preparation, teachers and students' attitudes towards teaching and testing, are factors at the micro levels of the test washback. The studies of the macro-level educational and social impact may prove more difficult. The CET stakeholders, or rather, the users of the CET, such as parents, employers, school administrators, and educational policymakers, all have a vested interest in the use of the CET results because they are likely to be affected by students' performances in the test. But meanwhile, these groups of people may have the final say in the way the test is used.

Although it is difficult for test designers to directly influence how a test is to be used, conscientious efforts need to be made to introduce better testing and assessment practices to teachers and students. To better inform test stakeholders, both at the micro and macro levels, data need to be collected and made accessible to test users about the dimensions and aspects of how the test results can be used most appropriately. In other words, test designers should try to disseminate relevant information about the test and about the use of the test results.

What's more important, though, is to balance the power of each and every party of the stakeholders, that is, having testers and test takers, teachers, parents, educational administrators, and employers take and share the responsibility of making the test a better instrument. Experiences of the CET and the national college and university entrance examination in China indicate that the extrinsic power of a large-scale language test, instead of being built in by the test makers at the time when the test was designed, resides in considerable part in the high stakes brought about by the test users (see Gu 2004; Qi 2003). Users of a powerful test, willingly or unwillingly, derive authority from their control over the uses they make of the test. With authority comes responsibility. All stakeholders, therefore, should have a role to play in improving the educational and social consequences of a high-stakes test.

At the annual conference of International Language Testing Association held in Vancouver, Canada, in 2000, the Code of Ethics for the International Language Testing Association was drafted, representing an initiative to guard language testers against misuses of tests. The last of the code's nine principles directly addresses the issue of the use of tests: "Language testers shall regularly consider the potential effects, both short and long term, on all stakeholders of their projects, preserving the right to withhold their professional services on the grounds of conscience" (International Language Testing Association 2000).

The job of the CET Com mittee is ,therefore ,not simply of an academic nature . Instead ,as Prof Hamp -Lyons (2000 :587) said ,the widening of the lens through which we look at tests brings with it great ethical or moral problems about the responsibility of the language tester and of language testing as a profession . Therefore ,as the designers of the largest EFL test in the world ,the CET Com mittee must provide conscientious service to its test takers and test users . To bring the power of the test under control , however ,by necessity calls for concerted efforts by all the groups of people involved in test making ,test taking ,and test using .

References

- Alderson ,J . C . 2004 . Foreword . In L . Cheng ,Y . Watanabe &A . Curtis (eds .) . Washback in Language Testing :Research Contexts and Methods ,pp .ix xii . Mah wah ,NJ :Lawrence Erlbaum Associates ,Inc .
- Alderson ,J . C . &D . Wall .1993 . Does washback exist ?Applied Linguistics 14 :115 -129 .
- Andrews ,S .2004 . Washback and curriculum innovation .In L . Cheng ,Y . Watanabe ,&A . Curtis (eds .) . Washback in Language Testing :Research Contexts and Methods ,pp . 37 -50 . Mah wah , NJ :Lawrence Erlbaum Associates ,Inc .
- Bach man ,L . F . &A .S .Palmer .1996 . Language Testing in Practice . Oxford :Oxford University Press .
- College English Test Design Group .1987 . College English Test Band Four Syllabus and Sample Test . Shanghai : Shanghai Foreign Language Education Press .
- College English Test Design Group .1989 . College English Test Band Six Syllabus and Sample Test . Shanghai : Shanghai Foreign Language Education Press .
- College English Test Design Group .1994a . College English Test Band Four Syllabus and Sample Tests (Revised Version) . Shanghai :Shanghai Foreign Language Education Press .
- College English Test Design Group .1994b . College English Test Band Six Syllabus and Sample Tests (Revised Version) . Shanghai :Shanghai Foreign Language Education Press .
- Gu ,X . 2004 . Positive or negative ?An empirical study of CET washback on College English teaching and learning in China . Unpublished doctoral dissertation ,Shanghai Jiaotong University .
- International Language Testing Association . 2000 , March . Code of Ethics for ILTA . Adopted at the annual meeting of the International Language Testing Association ,Vancouver ,Canada .
- Jin ,Y . 2000 . Backwash effect of CET -SET on teaching EFLin China . Foreign Language World 4 :56 -61 .
- Jin ,Y . 2004 . The develop ment of the CET . Foreign Languages in China 1 :27 -29 .
- Jin ,Y . 2005 ,July . The National College English Test .In L . Hamp -Lyons (Chair) . Big Tests . Symposium conducted at the annual meeting of the International Association of Applied Linguistics , Madison , Wisconsin .
- Jin ,Y . 2006 . On the improvement of test validity and test washback :The CET washback study . Foreign Language World 6 :65 -73 .
- Jin ,Y . in press . The National College English Testing Com mittee .In L . Cheng &A . Curtis (eds .) . English Language Assessment and the Chinese Learner . Routledge ,Taylor &Francis Group .
- Jin ,Y . &H . Yang . 2006 . The English proficiency of college and university students in China :As reflected in the CET . Language , Culture and Curriculum 19 /1 :21 -36 .
- Hamp -Lyons ,L . 1997 . Washback ,impact ,and validity :ethical concerns . Language Testing 14 :295 -303 .
- Hamp -Lyons ,L . 2000 . Social ,professional and individual responsibility in language testing . System 28 :579 -591 .
- Kunnan ,A . 2004 . Test fairness .In M . Milanovic ,C . Weir &S . Bolton (eds .) . Europe Language Testing in a Global Context : Selected Papers from the ALTE Conference in Barcelona ,pp . 27 -48 . Cambridge : Cambridge University Press .
- Li ,X . &Y . Zeng . 2002 ,September . Language testing and teaching :Is the tail wagging the dog ?Paper presented at the First International Conference on English Language Testing in China ,Shanghai .
- McNamara ,T . 2000 . Language Testing . Oxford :Oxford University Press .
- Messick ,S . 1989 . Validity .In R . L . Linn (ed .) . Educational Measurement (3rd ed .) ,pp .13 -103 . New York :Macmillan .
- Messick ,S . 1996 . Validity and washback in language testing . Language Testing 13 :241 -256 .
- Ministry of Education . 1999 . National College English Teaching Syllabus (Revised Version) . Shanghai : Shanghai Foreign Language Education Press .
- Ministry of Education . 2004 . College English Curriculum Requirements (For Trial Implementation) . Shanghai :

- Shanghai Foreign Language Education Press .
- National College English Testing Committee .1999 . CET -SET Test Syllabus and Sample Test Paper . Shanghai : Shanghai Foreign Language Education Press .
- National College English Testing Committee .2006a . CET 4 Test Syllabus and Sample Test Paper (2006 Revised Version) . Shanghai :Shanghai Foreign Language Education Press .
- National College English Testing Committee . 2006b . CET 6 Test Syllabus and Sample Test Paper (2006 Revised Version) . Shanghai :Shanghai Foreign Language Education Press .
- Qi ,L .2003 . The intended washback of the National Matriculation English Test in China :Intentions and reality . Unpublished doctoral dissertation . Hong Kong :City University of Hong Kong .
- Shohamy ,E .2001 . The social responsibility of the language testers .In Cooper ,R . ,E .Shohamy &J . Walter (eds .) . Language Education . Volume in honor of Bernard Spolsky ,pp . 113 -130 . John Benjamins Press .
- State Education Commission . 1985 . College English Teaching Syllabus (For College and University Students of Science and Technology) . Shanghai :Shanghai Foreign Language Education Press .
- State Education Commission . 1986 . College English Teaching Syllabus (For College and University Students of Arts and Science) . Shanghai :Shanghai Foreign Language Education Press .
- Sun ,Y .1985 . Collected Works of Sun Yatsen . Volume 5 . Beijing :Zhong Hua Book Company .
- Wall ,D . 2000 . The impact of high stakes testing on teaching and learning :Can this be predicted or controlled ?System 28 :499 -509 .
- Wu ,Q .2005 ,February . On the Revision of the CET . Second news conference of the Ministry of Education . Retrieved February 25 ,2005 ,from <http://www.moe.edu.cn>
- Yang ,H . &C J . Weir .1998 . The CET Validation Study . Shanghai :Shanghai Foreign Language Education Press .
- Yang ,X .(ed .) .2003 . The History of Testing in China . Beijing :Higher Education Press .

(...continued from p .21)

- Stemler ,S .2001 . An overview of content analysis . Practical Assessment , Research & Evaluation ,7(17) . Retrieved September 15 ,2002 from <http://PAREonline.net/getvn.asp?v=7&n=17>
- Stern ,H .H .1983 . Fundamental Concepts of Language Teaching . Oxford :Oxford University Press .
- Vygotsky ,L .S .1962 . Thought and Language . Cambridge ,MA :MIT Press .
- Vygotsky ,L .S .1978 . Mind in Society . Cambridge ,MA :MIT Press .
- Wallace ,M .J .1991 . Training Foreign Language Teachers :A Reflective Approach . Cambridge :Cambridge University Press .
- Wenger E .1998 . Communities of Practice :Learning ,Meaning and Identity . Cambridge :Cambridge University Press .
- Williams ,M &R .L . Burden . 1997 . Psychology for Language Teachers :A social Constructivist Approach . Cambridge :CUP .
- 教育部 2002 《2001 年全国教育事业发展统计公报》 2002 年12 月下载网址http://www.edu.cn/jiao_yu_fa_zhan_498/20060323/20060323_54096.shtml
- 连榕 2004 新手—熟手—专家型教师心理特征的比较 《心理学报》第1 期。
- 刘润清、吴一安等 2000 《中国英语教育研究》。北京 :外语教学与研究出版社。
- 刘润清、戴曼纯 2003 《中国高校外语教学改革 现状与发展策略研究》。北京 :外语教学与研究出版社。
- 孙红 2004 高校教师培训发展探索 《中国成人教育》第11 期。
- 施黎明、周泽铭 2002 西南地区高校教师培训专题调研工作总结 《西南师资培训》第2 期。
- 徐金明等 2004 我国高校教师培训改革发展的前瞻与对策 《教师教育研究》第4 期。
- 叶澜 ,2001 ,《教师角色与教师发展新探》。北京 :教育科学出版社。
- 杨成伟、李广达 2003 ,辽宁省高等学校教师培训工作专题调研报告 《辽宁教育研究》第6 期。
- 杨秀玉 1999 教师发展阶段论综述 《外国教育研究》第6 期。
- 杨艳玲 2001 成长的烦恼 《聆听与倾诉—质的研究方法应用论文集》 杨璞 林小英编。北京 :教育科学出版社。
- 朱旭东 2000 国外教师教育改革动力分析 《高等师范教育研究》第12 期。