

# Thoughts on College English Testing

Li Sun

School of Foreign Languages, Changchun University of Science and Technology, Changchun, China

**Abstract**—Testing and teaching are so closely related that it is virtually impossible to work in either area without being constantly concerned with the other. Testing must accompany English teaching and reflect how much English knowledge students have grasped in a certain phase of English study. Tests should be constructed primarily to reinforce learning, to motivate students, and to assess students' performance in language acquisition. Thus, it is necessary for instructors to design tests according to the features of the college English intensive reading course in order for validity and reliability to occur.

**Index Terms**—teaching and testing, testing procedures, compatibility and validity

## I. INTRODUCTION

Progress tests, also known as classroom tests, are tests that measure the extent to which students have mastered the material taught in the classroom. Each progress test is unique and can only be evaluated fully by the instructor in light of his or her knowledge of the students, the language program in which the students participate, and the instructor's own particular aims and goals. The instructor often measures students' ability by quizzes and small scale tests and the results obtained from them enable the instructor to become more familiar with the work of each student and with the progress of the class in general. A good progress test should encourage students to perform well in the target language and to gain additional confidence. In addition, the aim of classroom testing is to stimulate learning and to reinforce what has been taught.

Progress tests differ from achievement tests in that progress tests are less formal. Achievement tests measure achievement on a large scale and often contain more test items, test types, test time, and standard scoring. These tests are based on what the students are presumed to have learned, not necessarily what they actually learned or what they were actually taught.

## II. CLASSROOM TESTING

### A. *What Kinds of Classroom Practices and Exercises Should Be Considered for Testing Beginning Students?*

As we ponder the idea of classroom testing, the question arises: what kinds of classroom practices and exercises should be considered for testing beginning students? The college English intensive reading course is one highly practical course geared towards college students of various majors. Testing is not only reserved for written exams, but also for all sorts of classroom practices and exercised. In a broad sense, all exercises and practices, especially oral practices, can be regarded as a kind of testing, such as reading texts aloud or answering questions, which, to a certain degree, may reflect students' pronunciation, intonation, and the mastery of vocabulary and grammar. This type of testing, however, can be a disadvantage in that evaluating "finish products" may not reflect students' integrative English levels. For instance, a beginning student read a poem eloquently and passionately under the guidance of instruction and receives high praise in a speech contest. At first, it may appear that the students' performance is based on his or her proficiency in English when in actually, repetition and memorization brought about the appearance of proficiency instead of actually knowledge of the language.

Of course, it is neither reasonable nor practical for instructors to give assessment marks for all students' reading or conversations. After all, students should not be hindered from learning because of fear of being assessed. If students are assessed too much in the classroom, enthusiasm is greatly reduced, the relationship between students and instructor is decreased, the learning atmosphere in the classroom becomes tense, and students do not dare to think, speak and question freely.

It should be mentioned that multiple choice testing is often a component in many types of beginning tests; however, the marks obtained from multiple choice do not completely reflect students' true knowledge of English skills in that it fails to measure the degree that students answer question correctly or incorrectly. In addition, multiple choice testing fails to measure the ability to use language such as the ability of speaking, writing and translating. Thus, multiple choice should not be major component in classroom testing.

### B. *What Form of Classroom Tests would Be Most Suitable to Evaluate English Skills of Moderately Advanced Students?*

A second question arises: what form of classroom tests would be most suitable to evaluate English skills of moderately advanced students? For years teaching practice has used integrated testing forms involving listening

comprehension, dictation, and reading comprehension, which measure students' comprehensive knowledge of intonation, pronunciation, spelling, vocabulary and other areas. An experienced instructor can write reliable and valid test questions that accurately and consistently measure students' abilities in these areas. Take a Chinese student who graduated from the Chinese institute. The student always received good testing marks in listening comprehension, dictation and reading comprehension. When he studied at a New York university, instructors praised him for his excellent knowledge of English. This is one of many examples demonstrating the consistency of these testing areas. Hence, listening comprehension, dictation, and reading comprehension would be most suitable to evaluate moderately advanced students' work.

### C. *What about Reliable and Valid Classroom Tests for More Advanced Students?*

A third question arises: what about reliable and valid classroom tests for more advanced students? According to the requirements of Chinese students, outlined in the college English Teaching Outline, students learn English during four years of college study. Since during the third year university require students to learn large amounts of vocabulary, using words correctly in context, translation and interpretation skills are key points for English teaching. Thus, retelling stories is used as a valid classroom practice and test to measure these areas. After reading material and working out the exercises of the texts such as vocabulary, translation, and answering question, students should be able to retell information in a text. Retelling information can show how much a student has learned from a text and judging from the students' degree of achievement, the instructor can see if his or her teaching methods are effective. So, retelling stories using oral or written modes of expression is a way of testing for more advanced students.

## III. ANALYZING THE USED TEST PAPER

### A. *The Importance of Analyzing the Used Test Paper for Teachers*

From analyzing the used test papers, teachers could have a clear understanding whether the test papers have validity, reliability and discrimination.

Whether a test is good or not, first we must know whether it has validity or not. Validity means whether a test measures what it intends to measure, it includes face validity, content validity, concurrent validity, construct validity and predictive validity. No matter what a large-scale test it is, it does have these kinds of validity. As for the face validity, we can judge it soon after the test papers have been finished designing. But other kinds of validity could only be judged through analyzing the used test papers. Many teachers have such opinion that only the large-scale exam needs analysis, as for our final especially the mid-term exam there is no need to do that. But this is a kind of misleading opinion. Our purpose of carrying out the exam, no matter what a large-scale test it is, is to judge whether we can test what we want to test, if you did not make any analysis on the used test papers how could you know that. Maybe when you design the test papers, you have ignored something, and you need to make analysis. For example, in an exam, if you design a new type of items that you want to exam the students' comprehensive ability, but there are so many new words in it which have bad effect on the students' comprehension on it, so such kind of items lack of construct validity. If one of the exam questions intends to test the students' mastery of grammar, but it turns out to test the students' mastery of vocabulary, then it lacks of content validity. Also, if you put aside the used test papers and made no analysis on them, how can you know the concurrent validity of the test. So analyzing the used test papers is very important for the teachers to decide the various kinds of validity of a test.

Reliability is another important factor for a test. It refers to the extent to which test scores are consistent. So through analyzing the used test papers we can know whether there is any injustice scoring and whether all the teachers have the same scoring standard. Also we could know whether the test contains enough information and the items are representative. Because for the mid-term and final exam, the designers are always the markers, so they could know how to make the test most reliable next time through analyzing it. Besides that, if we want to make a comparison among the different years of graduates through their exam scores, first we must make sure that the tests have validity and reliability at the same level through analyzing the used test papers. Also because the mid-term and final exam belong to the achievement test, so we need a certain kinds of discrimination, whether a test has such kind of discrimination we need to compare the results of the students got in the exams. If it does not have, we need to analyze the test paper itself to find out the reason.

From the used test papers the teachers could know how to design a more successful test paper. As for this point, it is especially useful for the young teachers. Maybe at the very beginning of their work, they could get no chance to design the test papers. But eventually they will get the chance, so they need to accumulate experience. From analyzing the used test papers, they could know how to find out the most appropriate forms of the component parts of a test. For example, if they want to test the students' vocabulary, whether they should use the form of filling in blanks or choosing the synonyms or just writing down the words according to their Chinese meanings, as for the multiple choice they could know how to design the stems and the opinions, they could know how to weight each section, they could know how to make clear directions for each part, because for some new test designers, sometimes, they will create ambiguous directions. They could know which part is better and how to make similar items in their later designing.

The analyzing of the used test papers is not only useful for the young inexperienced teachers but also useful for the experience teachers. We call some teachers experience teachers but it does not indicate that they could always design

the perfect test papers. They could also design the invalid items which maybe they could not realize when they were designing. If they design such kind of items again, other people will probably doubt about their ability. So they have to analyze the used test papers to find out which parts are not very good and summarize the reasons, then they will be more careful next time. And also from analyzing the used test papers, the teachers could know which parts seems to be too difficult or too easy for the students, then they could design more practical test papers. And experienced teachers also know that they could find out the errors the students are easily to make from analyzing the used test papers. They could design such items as error corrections. In addition to what I talked above, through analyzing the same used test papers, different teachers could state their own opinions and exchange their ideas, they can point out which part needs to improve, which part should be deleted and which item is good enough to be used again.

The analyzing of the used test papers could have a positive effect on English language teaching. One of the effects of testing is to serve as an assistant of language teaching, it can diagnose teaching and guide teaching. From certain test the teachers can know many things. For example, if the test indeed tested what we want to test, we could know which part had the students mastered well and which part they had difficulties, so in our further teaching we could have a kind of guidance about we should delicate explanation and make more exercise on it.

We often say that a good test can have a positive feedback effect on teaching. Although it was a very good test, the teacher did not pay enough attention, just give the marks and made not analysis on it, how can it has such kind of effect. On the contrary, if the teachers made a careful analysis of the used test papers, they could have a clear consciousness on the focal points in their further teaching, they do not have to spend more time on explaining the knowledge the students had mastered well. And also their teaching could become happy and more successful. So, the teachers should keep their used test papers especially the young teachers, they should make analysis on them so that they could make a comparison among the students of different grade, work out different targets and give their own recommendations. Once there are more accumulation of such kinds of test papers, it will become the precious materials for studying testing. It will not only save time for the teachers in their further designing the exam question, but will also help them to make clear the focal points in teaching and to improve the teaching quality.

In a word, making analysis of the used test papers is not only helpful for the teachers to design more practical test paper but also has a beneficial feedback effect on teaching. Besides that, analyzing the used test papers is also important for the students.

#### *B. The Importance of Analyzing the Used Test Papers for the Students*

Nowadays, almost no university and college distribute their used test papers back to the students needless to say the teachers explain them to the students. One of the reasons is that the teachers want to store the used test papers for further research, but they could collect them back after analyzing them together with the students.

Most of the students want to know why they have got such marks, some students think they did the exam questions quit well but they have got low marks, some students do not think they did well but they have got high marks so they want to find out the reasons.

There are also some careful and diligent students who want to have their used test papers to see which part they lost marks and why. Is it because the items are too difficult or because they did careless working. And if they lost marks really because they could not work out the items, they will find similar items and so more such kind of questions so as to improve their own ability, or they can as the teachers in time so h twill not make the same mistakes again. And sometimes the students could not work out some items in an exam, they are eager to know the answers but they could not get the original materials so maybe they will not know the answers for a long time, maybe it will affect the students' accumulation of other knowledge just because language is a continuous system. Some of my students had suck kind of experience, after they had taken the cet-4 exam, they said "we had seen the similar exam question in our former final exams but we did not work out it at out that time and this time we have not work either. If our teacher had explained it to us we would have done it well". So at that time I realized the importance of analyzing the used test papers. Teachers should not only distribute the used test papers back, but also should help the students to analyze them. They do not have to analyze each question in detail but should help the students make it clear why they make suck kind of errors and explain the part that most of the students made mistakes. As long as the students analyze the students analyze the used test papers with the help of their teachers, they would know how to work out the similar questions next time and how to use the special method to work out a question quickly and correctly. All in all, it is necessary for the students to make analysis on the used test papers.

Also, besides what I have talked above, because the test is designed for the students, they have used the test papers so they have their own opinion about the test papers. The teachers could ask the students to give their opinions on the test papers and listen to their suggestions. It is very useful for both the students and the teachers to do so.

### IV. RELIABILITY OF CLASSROOM TESTS

#### *A. Factors Affecting the Reliability of a Test*

There are two important factors affecting the reliability of a test. They are the extent of the sample of material selected for testing and the administration of the tests. As validity is concerned chiefly with the content of the sample, so reliability is concerned with the size. The larger the sample, the greater the probability that the test as a whole is

reliable. And the administration of the test is the same test administered to different groups under different conditions or at different times? Obviously, this is an important factor in deciding reliability, especially in tests of oral production and listening comprehension. In order to differ from validity, we can give you the following example: if a recording for a listening comprehension test is initially poor in quality, then it is poor in quality for all candidates. This will result in invalidity. But if the quality of the recording is good and if certain groups hear it played under good acoustic conditions while the other groups hear it under bad acoustic conditions, this will result in unreliability and invalidity. There are also some other factors affecting the reliability of a test, such as test instructions, personal factors and scoring the test, but the important factors are that we just talked about.

#### *B. Methods of Measuring the Reliability of a Test*

One method of measuring the reliability of a test is to re-administer the same test after a lapse of time. It is assumed that all have been treated in the same way in the interval---that they have either all been taught or that none of them have. If such assumptions can be made, comparison of the two results would then show how reliable the test has proved. Clearly, this method is often impracticable and, in any case, a frequent use of it is not to be recommended, because certain students will benefit more than others by a familiarity with the type and format of the test. Moreover, in addition to changes in performance resulting from the memory factors, personal factors such as motivation and differential maturation will also account for differences in the performances of certain students. Another means of estimating the reliability of a test is by administering parallel forms of the test to the same group. This assumes that two similar versions of a particular test can be constructed, such tests must be identical in items contained in the nature of their sampling, difficulty, length, rubrics. Only after a full statistical analysis of the test and all the items contained in them can the test safely be regarded as parallel. Of the correlation between the two tests is high, then the tests can be termed reliable. The split-half method is yet another means of measuring test reliability. This method estimates a different kind of reliability from that estimated by test producers. The split-half method is based on the principle that if an accurate measuring instrument were broken into two equal parts, the measurements obtained with one part would correspond exactly to those obtained with the other.

#### *C. The Importance of Reliability*

Imagine that forty students take a 100-item test at one o'clock one Friday afternoon. The test is not impossibly difficult or ridiculously easy for these students, so they do not all get zero or a perfect score of 100. Now what if in fact they had not taken the test on the Friday but had taken in at one o'clock the previous afternoon or the following afternoon? Each student would not have got the same score on the Friday as they actually did on the Tuesday or on the Saturday. Because human beings simply do not behave in exactly the same way on every occasion, even when the circumstances seem identical. We talked about some factors affecting the reliability and some methods of measuring the reliability. Here we will talk about the importance of reliability. First, we will introduce the reliability coefficient. Reliability coefficients are like validity coefficients. They allow us to compare the reliability of different tests. There is a test with a reliability coefficient of one is one which would give precisely the same results for a particular set of candidates regardless of when it happened to be administered. A test which had a reliability coefficient of zero would give sets of results quite unconnected with each other, in the sense that the score that someone actually got on a Friday would be no help at all in attempting to predict the score he or she would get if they took the test the day after. It is between the two extremes of one and zero that genuine test reliability coefficients are to be found. While the reliability coefficient allows us to compare the reliability of tests, it does not tell us directly how close an individual's actual score is to what he or she might have scored on another occasion. With a little further calculation, however, it is possible to estimate how close a person's actual score is to what is called their true score. If a test is not reliable then we knew that the actual scores of many individual are likely to be quite different from their true score. This means that we can place little reliance on those scores. Even where reliability is quite high, the standard error of measurement serves to remind us that in the case of some individual there is quite possibly a large discrepancy between actual score and true score. This should make us very cautious about making important decisions on the basis of the test scores of candidates whose actual score place them close to the cut-off point. We should at least consider the possibility of gathering further relevant information on the language ability of such candidates.

### V. CONCLUSION

Testing plays a key role in promoting effective college English teaching. Its purpose lies in examining students' understands and mastery of word, expression, sentence structures and contents of certain texts during a designed period of time; determining if teaching methods are effective, and finding out if students are receiving instruction effectively. Certainly in order to make progress tests serve teaching better, instructors must reinforce the validity and reliability of testing in college English intensive reading course and make testing an effective means of measuring teaching students learning and promoting the teaching and learning progress. Testing is a very useful assistant of teaching and also an indispensable method of inspecting the teaching result so we must give it enough attention. And if we want to get useful information from testing, we must make analysis on the used test papers. On the one hand it can help the teachers to design more practical and successful test papers in their future work and help them to absorb each other's good points

and on the other hand it can help the students to have a clear understanding of their own weak points and learn a better way in language study and help them to be more successful in their study.

#### REFERENCES

- [1] Aikhenvald Alexandra. (2001). *Problems in Comparative Linguistics*. Oxford: Oxford University Press.
- [2] Aitchison Jean. (2001). *Language Change: Progress or Decay*. Cambridge: Cambridge University Press.
- [3] Campbell Lyle. (2004). *Historical Linguistics*. Edinburgh: Edinburgh University Press.
- [4] Clarke Smith. (1990). *Sources of semiotic*. Carbondale: Southern Illinois University Press.
- [5] Croft William. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- [6] Tim McNamara. (2003). *Language Testing*. Shanghai: Foreign Language Teaching and Research Press.

**Li Sun** was born in Jilin, China in 1978. She received her Master degree in foreign linguistics and applied linguistics from Changchun University of Science and Technology, China in 2005.

She is currently a lecturer in School of Foreign Languages, Changchun University of Science and Technology, Changchun, China. Her research interests include linguistics and teaching methodology.