

# Analysis on the Main Factors Affecting the Reliability of Test Papers

Jiang Zhu

School of Foreign Languages, Changchun University of Science and Technology, Changchun, China  
Email: zhujiang98@163.com

Lemeng Han

School of Foreign Languages, Changchun University of Science and Technology, Changchun, China  
Department of Common Foreign Language Teaching, Guanghua College of Changchun University, Changchun, China  
Email: hanlemeng1982@sina.com

**Abstract**—Reliability is an important factor to evaluate test papers. This paper analyzes the factors which affect the reliability of test papers and discusses the methods to increase the reliability of test papers.

**Index Terms**—reliability, test paper, factor

## I. INTRODUCTION

Reliability refers to a measure which is reliable to the extent that independent but comparable measures of the same trait or construct of a given object agree. Reliability depends on how much variation in scores is attributable to random or chance errors.

It seems that it is difficult for us to trust any set of test scores completely because the scores will be different if the test has been administered on different situation. What can we do to reduce the declination of the tests? On one hand, we can construct and administer the score tests in such a way that scores actually obtained on a test on a particular occasion are similar to those which are obtained when the test is administered to the same students with the same ability, but at a different time. The more similar the scores are, the more reliable the test is. In this way the more reliability is obtained in the test.

On the other hand, we can give the candidates the same paper for many times. If the candidates get approximate scores on the same test paper regardless of the factors of time, place and so on, we can say that the paper has a high reliability. Reliability is a very important and necessary feature of test papers. Any paper without reliability can not reflect the true ability and level of the candidates.

## II. THE MAJOR FACTORS WHICH AFFECT THE SCORES OF TEST PAPERS

To analyze the factors which affect the reliability based on scores, let us see the factors which can affect the scores of test papers. Bachman (1997) considers that the scores of test papers are determined by the following four factors: the language ability of candidates, testing methods, personal features and random factors.

The scores will be decided by the language ability of the candidates, which is a basic factor. If candidates take the same language test over and over again and their performance is not affected, we may estimate their language ability through their scores. The scores are called true scores of a test paper. The more true scores we can get the more reliability the test paper will achieve. Testing methods mainly refer to testing environment, testing arrangement, or time division, which can also influence the testing scores. For example, if we use direct testing, we can get the marks on the spot, but the candidates may be affected psychologically by the circumstantial factors and can not show their real ability. If we use indirect testing, the candidates may get higher scores because of the relaxation. Personal features refer to some personal characteristics which are irrelevant to the language ability of candidates, like the teacher's point of view or the students' attitude. Even though the teacher is objective, the candidates themselves will be influenced by their own factors psychologically and physically. Random factors refer to the factors which can not be predicted beforehand. The last three factors are not as popular as the first one, for they disturb the reflection of how candidates achieve their scores. Therefore, we can reduce the effect of the last three factors in a language test.

## III. THE MAJOR FACTORS WHICH AFFECT THE RELIABILITY OF TEST PAPERS

In the previous part I analyze the factors which affect the scores of test papers. In this part I will discuss the factors which affect the reliability of test papers: the change of candidates, the process of testing, the features of tests and the methods of going over test papers.

#### A. *The Effects on Reliability through the Change of Candidates and Process of Testing*

The change of candidates themselves can affect testing scores by two methods. One is to change true score. The other is to mislead testing results. The former method can reflect the change of candidates' language ability. The latter one can affect candidates' real language level (Davies, 1990). For example, some factors such as sickness and emotions will interfere with the candidates' normal testing level so that the consistency of a test is reduced. Although these factors are unpredictable and there are also some accidents which are out of control, we still have to reduce the effects of these factors. We should inform candidates to rest instead of studying hard before they take tests. We should also improve the validity of the test paper (Liu, 2000). The content of the test paper should meet the needs of testing major students. No matter how perfect a test paper is it will still lose its consistency if there are some problems during the process of testing. This phenomenon often occurs and is easy to be neglected, such as testing environment and plagiarism.

Time distribution is also very important to reliability. In some tests which aim to test the speed of reading comprehension, scores not only reflect the candidates' comprehensive ability but also their reading speed. However, in some tests which aim at candidates' language ability, candidates have to prove their ability on a certain task. If time is not enough in such tests, the reliability of the tests will be influenced. Testing environment is also very important. Natural environment without noise and pollution will be the best choice. Candidates often work better on tests in the morning than in the afternoon.

#### B. *The Effects of Testing Features on Its Reliability*

Some features such as length and difficulty on the test paper itself will also affect its reliability. No matter whether the standard level of testing is norm-referenced or criterion-referenced, longer papers always show more reliability than short ones. This is because the more contents there are in the test paper, the bigger scale there is in it. In fact, if there are more representative contents in the test paper, the reliability of the paper will be more complete. Therefore, important tests should be longer papers instead of short ones. Of course the content of test papers can not be over-sized, or the reliability of the papers will also be affected.

Some research shows that when the quantity of testing content surpasses a certain degree of asymptote, the quantity of the paper can not make effects on its reliability anymore. The degree of testing difficulty and division will also affect testing reliability. To some candidates, some easy and simple questions in the test seem to be difficult. If the questions in the test are either very difficult or very easy, the reliability of test paper will both be influenced. Therefore, researchers suggest that the difficulty of tests must be during 0.33-0.67(Harrison, 1983). The difficulty of test papers is a relative notion. We can not make all tests by one standard. For example, a high qualified TOEFL test paper may not be very appropriate as a mid-term exam for middle school students.

#### C. *The Effects of Going over Test Papers on the Reliability of Tests*

No matter how well we have done in organizing test papers, the tests will also be unsuccessful if we make mistakes during the process of going over test papers. Objective questions don't require any subjective judgment so that it can achieve a high reliability and consistency. However, some subjective questions which need our subjective judgment will influence reliability of the tests for there are some differences between questioners and questionnaires.

In order to keep the reliability of the tests, we have to make a standard to go over the papers. If there is no such standard, a tester will make different comments on the same test paper in different time and place. For example, when we go over the writing part of a test paper, we may neglect some faults on grammar at the very beginning. But we will gradually notice the faults and focus more attention on its correctness. Finally we will change our attitude and be more critical with it (Heaton, 1988). Besides, personal attitude, ability and experience will also influence the consistency of going over a paper.

As far as we know, testing results are often influenced by the above four major factors. In order to keep the reliability of a test, we should be more careful of every point which is related to a test paper. Teachers should have some basic testing knowledge and trainings so that they can use test papers to test students' ability properly and effectively.

### IV. HOW TO INCREASE THE RELIABILITY OF TEST PAPERS

The reliability of a test paper is affected by some factors. In this part I will discuss the methods to increase the reliability of a test paper.

#### A. *Do Not Allow the Candidates Too Much Freedom*

The range of the possible answers should be restricted in the questions of a test paper. In some language tests there is a tendency to offer candidates a choice of questions and allow them to answer the questions they have chosen with a great deal of freedom. The disadvantages are that the more freedom the candidates can get, the more difficult it is to estimate the levels of the different candidates for their performance. A good test paper should control the freedom and require the candidates to do the similar or the same task so that we can make the results of the test more objective and easy to make comparison (Madsen, 1983).

#### B. *Do Not Write Ambiguous Items in Test Papers*

It is important for the candidates to get the items whose meanings are clear. The best way to get unambiguous items is to draft and check the items, and then subject them to the critical scrutiny of colleagues, who should try as hard as they can to find alternative interpretations to the ones intended. In this way most of the problems can be identified before the test is administered. The testers should also provide the explicit instructions to make sure that the candidates can understand the request and they should write the answers clearly. The test papers should be typed well so that the candidates can see the papers clearly.

#### *C. The Scorers Should Get Some Training*

This point is particularly important when the scoring is very subjective. The scoring of the composition, for example, should not be assigned to anyone who has not learned to score compositions accurately before. After going over test papers, patterns of scoring should be analyzed (Bachman, 1989). Individuals whose scoring deviates markedly and inconsistently from the norm should be told how to do it correctly. In addition, a detailed scoring key should be provided. It should be the outcome of efforts to anticipate all possible responses and should be subject to group criticism. Items can also be used to permit the scoring as objective as possible, such as those in multiple choices.

#### *D. Identify Candidates by Number Instead of Names*

Scorers inevitably have expectations of candidates whom they know. Except in purely objective testing, this will affect the way by which they score. Facts have shown that even when the candidates are unknown to the scorers, the name on a script will make a significant difference to the scores given (Henning, 1987). For example, a scorer may be influenced by the gender or nationality of a name and may make predictions which can affect the score given. The identification of the candidates by number will reduce such effects.

### V. CONCLUSION

In this paper I first mention the notion and importance of reliability. Then I discuss the main factors which affect the reliability of test papers. At last I analyze the way to increase the reliability of test papers. I hope that testers will make more and more reliable test papers in the future.

### REFERENCES

- [1] Bachman, L. F. (1997). *Fundamental Consideration in Language Test*. Oxford: Oxford University Press.
- [2] Bachman, L. F. & Palmer A. S. (1997). *Language Test in Practice*. Oxford: Oxford University Press.
- [3] Bachman, David. (1989). *Language Testing: A Critical Survey and Practical Guide*. London: Edward Arnold.
- [4] Davies, Alan. (1990). *Principles of Language Test*. Oxford: Oxford University Press.
- [5] Harris, David. (1969). *Testing English as a Second Language*. New York: McGraw.
- [6] Harrison, Andrew. (1983). *A Language Testing Handbook*. London: Macmillan.
- [7] Heaton, J. B. (1988). *Writing English Language Test*. New York: Longman.
- [8] Henning, Grant. (1987). *A Guide to Language Test*. Cambridge: Newbury House.
- [9] Madsen, Harold S. (1983). *Techniques in Test*. Oxford: Oxford University Press.
- [10] Liu Runqing. (2000). *Language Test and Its Methods*. Beijing: Foreign Language Teaching and Research Press.

**Jiang Zhu** was born in Changchun, China in 1971. He received his M.A. degree in linguistics from Changchun University of Science and Technology, China in 2003.

He is currently an associate professor in School of Foreign Languages, Changchun University of Science and Technology. His research field is applied linguistics.

**Lemeng Han** was born in Changchun, China in 1982. She works in the Department of Common Foreign Language Teaching, Guanghai College of Changchun University.

She is now studying to get her M.A. degree in linguistics in School of Foreign Languages, Changchun University of Science and Technology. Her research field is applied linguistics.