

# Research on the Construction of Test Questions Bank Based on Educational Measurement Theory

Fei Wang

Wenshan University, Wenshan, 663000, China

**Keywords:** Test questions bank, Test questions library, Educational measurement theory.

**Abstract.** The implementation of test question bank system has become an important part of college teaching system. The evaluation indexes of test questions quality based on educational measurement theory mainly include difficulty and distinguishability. The evaluation indexes of test quality based on educational measurement theory mainly include reliability, validity, test questions structure and the proportion of core test questions. Practice shows that the theory of educational measurement can effectively improve the quality of test questions bank construction.

## Introduction

The comprehensive evaluation of examination questions quality is an important content of educational surveying. It is an important tool to reform teaching and carry out scientific education management. It is impossible to talk about the education problems to leave the quality of the examination questions. Classical educational measurement theory has a set of methods to evaluate the quality of test questions. The commonly used quantitative indexes include difficulty, distinguishability, reliability, validity and so on. The effective measure of education is the purpose of our construction of examination questions database. Educational measurement should follow the law of education, and its theoretical foundation is educational statistics and educational measurement theory. The primary task of educational measurement is to accurately measure the level of testers and minimize measurement errors. To make the examination more scientific, it is necessary to make use of educational measurement, while the object of educational measurement is the whole process of examination, marking, analysis and evaluation. Examination is the most direct way to measure the level of education at present. It involves the concept and theory of educational statistics. The purpose of educational statistics is mainly to detect the scientific nature of education to explore the various rules and characteristics of teaching phenomena, and to study the phenomena and essence of education through statistical methods and theories. Examination paper is the main form of examination at present, and the result of measurement decides the quality of examination. Whether the design of paper is scientific or not will be an important step in teaching work. It can reflect the teaching level of teachers and the students' learning effect objectively, and it is also the most important thing to be studied in this paper. In the construction of modern examination question bank system, advanced educational measurement theory is introduced, advanced technology is guided by advanced theory, and the construction of examination question bank is ensured, which fully adapts to the development trend of modern teaching idea. This has become a problem that must be paid attention to in the construction and development of modern test questions database. It is necessary to recognize and evaluate the quality evaluation indicators of educational measurement theory for both education managers and teachers.

## Quality Evaluation of Test Questions Based on Educational Measurement Theory

**Difficulty.** The difficulty of test questions refers to the degree of difficulty. It is the index of students' knowledge and ability level. The difficulty index of the test questions is usually expressed by the passing rate of the examination answer, that is, the proportion of the person who answers or passes the question. The difficulty is mainly divided into relative value and absolute value. Relative difficulty, also called statistical difficulty, is the difficulty calculated by statistical test. Absolute

difficulty, also known as cognitive difficulty, has nothing to do with the actual level of the person being tested. There are three characteristics of difficulty, that is, relativity, measurement and interval. Relativity is based on different tests and different test subjects based on statistics. Specifically, if the examinee's level is different, even if the same test paper, test statistics after the difficulty value will not be the same. Measurement refers to the difficulty is a specific quantifiable indicator. In the calculation analysis, the difficulty of examination, often just perceptual and vague, there is no accurate theoretical basis based on the judgment, but in the classical theory of measurement, the difficulty is defined as the specific indicators, calculation is through mathematical statistics calculated from the. If the statistical diagram shows that the test results are in normal distribution, then the difficulty of the test paper tends to be in the middle. When the statistical results show negative skewed distribution, the test paper is difficult and the test scores are low. When the statistical results show positive skewed distribution, the test paper is simple and the test scores are higher. It is worth noting that the difficulty is not immutable and frozen, because each class level is not the same, can the right answer would be different; on the other hand, even if a problem is more difficult, but because of the number of students to do more, skilled, the problem is relatively turned easy questions. Therefore, in the construction of test questions library, according to the correct rate of each examination and the number of times to be extracted, dynamic correction is carried out according to the measurement algorithm. Only in this way can we scientifically and objectively reflect the difficulty of the examination paper.

**Distinguishability.** Distinguishability refers to the degree of identification or differentiation to test the actual level of a candidate. It is an important index and basis for evaluating the quality and selection of test questions. Tests with good distinguishability can make students who have high levels get high marks. The distinguishability of test questions in general should be more than 0.3. The grading of the test questions is mainly to identify the examinee's actual learning ability and to distinguish the subjects being tested. The higher the division is, the better. Measurement experts regard the index of division as an important measure of whether the test is effective, and as the main index and basis for evaluating the quality of the project and screening the project. The better the division of the questions, the better the ability to examine and test the corresponding test questions. The degree of distinguishability is a quantitative index for evaluating the quality of test questions. It refers to the ability to distinguish the mental characteristics of subjects at different levels. If the subjects who have high scores on a certain test, the actual ability level is also high, the low score of the actual ability level of the subjects is also low, then the problem will have higher division. The degree of distinguishability is closely related to the difficulty, reliability and validity. If the questions are too difficult or too small, the segmentation is very low. To improve the reliability and validity of the test and to ensure the quality of the test, it is of great significance to improve the division of questions. There are many methods to calculate the division degree, and the objective questions and the calculation methods of the questions are different. These methods are more rational. In practice, the improved algorithm is also more. In classical educational measurement theory, the algorithms about graduation also depend entirely on the score information of students. The students' scores are reliable, the scoring objective and the calculation results of distinguishability can be of reference value, otherwise they are not credible. Similarly, the calculation results of the distinguishability also exist random error and system error.

### **Test Quality Evaluation Based on Educational Measurement Theory**

**Reliability.** Reliability reflects whether the measurement results meet the stability requirements. As the measurement type varies, both physical and educational measurements, even in the face of the same measurement object, the measurement results may also vary. Each measurement results will always be affected by two major factors, that is, the actual level of the object being measured and the unavoidable error in the measurement process. Reliability includes the effect of random error on the actual test process, and the inverse relationship between them. The main factors affecting reliability include four aspects. The test takers are inevitably subject to varying degrees of difference, which is the main reason for the uneven scores. The number of test questions and the length of the test paper

will lead to the test scores, subject to random sampling, and, in contrast, appropriately increase the number of subjects in the test paper, which has a positive effect on reliability. Whether the content of the test paper is the same or not has a positive influence on the reliability, that is, the reliability of the test paper is low in general, while the reliability is higher on the contrary. The difficulty of the test paper is skewed distribution when the degree of test is too high or too low. Reliability refers to the degree of stabilization and reliability. The reliability of measurement results indicated by correlation coefficient. In reliability analysis, the most commonly used measure statistic is the Alpha Cronbach coefficient. It can be pointed out if the reliability coefficient is more than 0.8, the scale has high reliability. We generally take the correlation degree of the test for two times as the index of reliability test. It shows the reliability of a test when it reflects the actual level of the subjects.

**Validity.** Validity reflects whether the measurement is effective and how effective it is, which is the proximity of the test measurement and the characteristics to be measured. If the characteristics to be measured are accurately measured, the validity of the measurement is determined. Analysis of the following key factors has an important impact on validity. Like the reliability, the nature of the test content is common or not, and the validity of the inverse proportion of validity of the law, that is, the higher the content consistency between the questions, the lower the validity. The difference between candidates should lead to a difference in the true score difference. The length of the test paper is also like the effect on the reliability. The number of subjects in the test paper is positively proportional to the validity. The content validity was used to test the validity of the test. The content validity test is the test content, and materials are measuring scope and content of educational objectives are consistent, testing problems caused by the behavior whether can clearly reflect the properties to be measured, the test results can be used as a representative sample of behavior. From the content satisfying the content validity, these conditions are clear and indefinite, and they are not clear and clear. They have obvious grey characteristics, and they are difficult to measure. In practice, we use logic inference to estimate validity. The method is to determine the content validity of the expert based on the full conformity of the test subject and the scope of the content involved. Surface validity refers to whether the meaning of the test is really measured. Tolerance validity refers to whether the content of the test includes all aspects of the content that you want to measure. Constructing validity refers to the effective inference of the student's level according to the test result. Generally, if the validity of the test is greater than 0.4, we consider that the validity of the test is high.

**Test Questions Structure.** The question structure refers to the proportion of all kinds of questions and types contained in the test paper. In the field of cognitive testing, written tests can be broadly divided into two categories. One is standardized test, also called objective test, and the other is discussion test, also called subjective test. For the examination of objective questions and subjective questions, many scholars in China have made a comparative study of these two types of questions. They also have less each one has his good points. The advantages of objective questions are objectivity, which can assess the extent of students' mastery of knowledge and facilitate wide range of use. But cannot test the ability to solve problems. The subjective test content is broad, can students' ability of knowledge depth and comprehensive analysis, and problem solving, can give students a more space to play, but is affected by subjective factors is relatively large. Therefore, the two types of questions can be used in combination, complement each other, and strengthen their strengths and circumvent weaknesses. It is generally believed that the proper proportion of objective questions to subjective questions is 7:3. The proportion of objective questions is too large, while the proportion of subjective questions is too small. This situation makes the examination lack depth. Therefore, the weight of the questions should be assigned reasonably, and the proportion of subjective questions should be increased. The number of the test questions should be enough, and the distribution of the questions is balanced within the range of the index attributes. With the three attributes of knowledge, difficulty and the requirements of the syllabus as the core, a three-dimensional crossover network will be formed. Each cross node on the network has a reasonable amount of questions. Based on ensuring the core structure, we should also ensure that the distribution of the questions and the classification of cognition should be reasonable.

**Proportion of Core Test Questions.** The core test questions mean the questions can clearly reflect the students' knowledge of questions is a knowledge node. These questions have a fairly high level both in content and form, when the teacher has a clear knowledge of the students, which is selected in a core problem, first consider the core problem. Another common question, to test the number specified. The between subject knowledge logic is not strong, each teaching unit includes a lot of knowledge, with strict logic between knowledge, and a knowledge point often on behalf of a chapter or section, will not be included in other chapters. When we organize questions, especially in the design of question bank management system, we should fully consider and adapt to the differences of the knowledge structure of this subject. Therefore, they are important components of high quality test paper. If an examination lacks the necessary core questions, the quality of its propositions will inevitably be questioned. The appropriate proportion of core and non-core questions is 3:7. Sometimes, the proportion of core questions accounting for the total number of subjects is significantly higher than normal. If there are too many core questions, it will disturb the test paper of the test bank and reduce the quality of the examination paper. As the most independent unit of examination, the quality of examination questions directly affects the quality of examination. Therefore, better quality examination questions are the prerequisite to ensure the quality of the examination. The analysis of questions is the process of critically examining questions to select the best questions to produce high-quality papers. In the examination question bank, the subjective question quantity is insufficient, the most difficult and the difficult question examination are few, in the examination question explanation and the question solution question proportion are too few. Therefore, we should adjust the proportion of core questions and improve the quality of test questions according to the educational measurement theory.

## Conclusion

The construction of the test questions bank is a long-term and complex job, and it is also a meticulous work. Its establishment can automatically generate papers, score and help teachers to analyze the test the quality, which greatly reduces the burden on teachers. Practice has proved that the test papers generated by the test questions bank based on educational measurement theory is available, which can reflect the students' mastery level objectively.

## References

- [1] Wang Tianjian, Peng Zhongmei. The Application of Generalizability Theory and EduG to Measurement in Education: Illustrated with a Study of Test Item Difficulty[J]. Journal of Wenshan University, 2016, 29(6): 88-93.
- [2] Du Jiaxuan, Chen Ping, Xin Tao. Analyzing classification consistency indices in large-scale assessment based on item response theory [J]. Journal of Beijing Normal University (Natural Science), 2015, 51(6): 643-648.
- [3] Li Huailong. The Confirmation of Measurability: Epistemological Foundation of Psychological and Educational Measurement [J]. Psychological Exploration, 2015, 35(4): 303-306.
- [4] Su Aoxue. The Quality Assessment of Math Test Paper based on the Educational Measurement Theory [J]. Journal of Minzu University of China (Natural Sciences Edition), 2017, 26(1): 43-47.