The Empirical Analysis and Evaluation of Examination Results of Security Investment Course

Yan Yang*

Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangdong, China

yangy3@sustech.edu.cn

*Corresponding author

Keywords: Test quality analysis, Reliability, Difficulty Index; Discrimination Index

Abstract: Examination and test are powerful tools to evaluate the teaching and learning effects. The quality of examination test will affect the assessment process and the analysis of high quality examination test will provide the theoretical bases for teaching reform. In this paper, we first introduced the quantitative indexes from the classical test theory, including reliability, difficulty index, discrimination index. Then based on the data of final examination of the course security investment from 2016 to 2021 in Sustech, we carried out the empirical analysis on these test results data and evaluated the quality of the examination tests. From the quantitative analysis results, we conclude that all those tests are reliable and consistent with medium difficulty and qualified discrimination.

1. Introduction

Examination and test are important assessment tools to measure the quality of teaching and learning [1]. The results of the test could give the instructors valuable information about the students' learning achievements and could also help the instructors to improve their teaching strategies and abilities in their future teaching practice. However, the quality analysis of the examination questions itself is also crucial for the teaching evaluation and teaching reform [2]. In addition, this analysis process could help us to identify the appropriate exam questions and build a reliable test bank in the future.

Security investment is a core course for undergraduate students majored in disciplines related to finance. After taking this course, the students should have a basic understanding of the financial theories including the modern portfolio theories, the theory of efficient market hypothesis, the common financial instruments such as stocks, bonds and derivatives [3].

In this paper, based on the data collected from the Southern University of Science and Technology, we will analyse the results of the final examination tests of the course security investment so that we could improve our teaching methods and skills when lecturing this course in the future.

2. Data and Methodologies

2.1. Data

The data covered the final examination paper of the course security investment in Sustech from 2016 to 2021. In each test, there are two type of exam questions, they are multiple choice and computational questions. Totally, 92 multiple choice questions and 41 computational questions are investigated, and 312 students test scores are used for our analysis. We also have detailed scores for each computational question for each student.

2.2. Methodologies

First, we will compute and compare the descriptive statistics for the overall test scores of each year. Then in order to give a quantitative analysis of the examination test, we would include the

  DOI: 10.25236/iche.2022.055

following three characteristic indexes from the classical test theory [2,4,5,6].

1) Reliability.

The reliability of a test is an index to evaluate whether the test is likely to produce consistent and stable scores, which means the students would have the same or nearly the same rank on the repeated test. The one of most used measure of reliability is called Cronbach's Alpha [5], which can be computed by the following formula [5]:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2}\right) \quad (1)$$

where k refers to the total number of items in the test, σ_i^2 refers to the variance associated with item i and σ_x^2 is the variance of the total score. In practice, we will use sample variance instead of population variance. In general, the range of α is between 0 and 1. Higher α indicates higher reliability.

2) Difficulty Index.

For each item in the test, difficulty index P is the ratio of the average score to the full score of that item. The higher is the difficulty index, the easier is that item. For the whole test, the difficulty index is just the weighted sum of the difficulty indexes of all the items. According to Allen & Yen [6], if the difficulty index is higher than 0.8, the item or the test is considered as easy. If the index is between 0.7 and 0.8, it's considered as moderate.

3) Discrimination Index.

Discrimination index measures the discrimination in the items between the students who perform good and not good on the overall test. Usually, the index is estimated by the difference between average scores of the students in the top and that at the bottom, adjusted by the full score of that item. Usually, the index is computed as

$$D = \frac{\bar{X}_H - \bar{X}_L}{m} \quad (2)$$

where D is the discrimination index, $\bar{X}_H$ is the average score for the students who ranked at top 27%. $\bar{X}_L$ is the average score for the students who ranked at the bottom 27%. m is full score of that item.

According to Ebel and Frisbie [7], if D>0.3, the examination question is considered a good question, while if 0.2<D<0.3, it would be better to modify it.

The discrimination index for the overall test is just the weighted sum of the indexes of all the items.

3. Results and Discussions

3.1. Basic Descriptive Statistics

Table 1 shows the basic statistics of the examination results for the six years. From the results, we can see that all the distributions are negative skewed.

Table 1 Descriptive statistics of examination results.

	2016	2017	2018	2019	2020	2021
0-59	24%	33%	27%	15%	26%	27%
60-69	10%	15%	24%	29%	10%	10%
70-79	22%	22%	16%	27%	15%	20%
80-89	33%	24%	32%	21%	28%	20%
90-100	11%	6%	2%	8%	21%	24%
Average	73	66	68	71	72	73
Median	78	71	70	72	77	77
Standard Deviation	19	19	18	15	21	18
Skewness	-0.96	-0.72	-1.70	-1.02	-0.90	-0.67
Kurtosis	0.40	-0.15	3.94	2.00	-0.17	-0.28

We also run the Jarque-Bera test for each year's data, the results indicate that the examination results of the year 2017 and 2021 are normally distributed while others are not.

3.2. Quality Analysis

We summarized the reliability, difficulty, and discrimination indexes for each year's data in Table 2. As we can see, the reliabilities are all above 0.8 except that in the year 2019, but even the reliability index for final examination test in 2019 is 0.73, which implies that all the tests are considerably consistent and reliable. The difficulty indexes are around 0.7. All the tests are relative appropriate from the aspect of difficulties. The tests in 2017 and 2018 are harder than the tests in other years. The discrimination indexes range from 0.34 to 0.49, which also means that all the test are fairly good in terms of discrimination index.

Table 2 Comparison of the Reliability, Difficulty Index and Discrimination Index.

	2016	2017	2018	2019	2020	2021
Reliability	0.82	0.82	0.88	0.73	0.87	0.81
Difficulty Index	0.73	0.66	0.68	0.74	0.72	0.73
Discrimination index	0.44	0.44	0.38	0.34	0.49	0.43

If we evaluate the difficulty index and discrimination index items by items, the comparison of those of multiple-choice questions are shown in Table 3.

Table 3 Comparison of the Difficulty Index and Discrimination Index for Multiple-choice questions.

	2016	2017	2018	2019	2020	2021
Difficulty Index	0.76	0.76	0.59	0.67	0.75	0.83
Discrimination index	0.31	0.35	0.28	0.27	0.31	0.27

The multiple-choice questions are hardest in the year 2018, which leads a relative higher overall difficulty index in 2018. However, in most years the difficult indexes are above 0.75, this is not surprising since most of the multiple-choice problems involves only remembering. The discrimination indexes for multiple-choice questions are between 0.27 to 0.35.

We plot the discrimination index against the difficulty index for computational questions. The result is displayed in Figure 1. About half of the computational questions have difficulty index in between 0.5-0.8 and have discrimination index higher than 0.4, which means they are good questions.

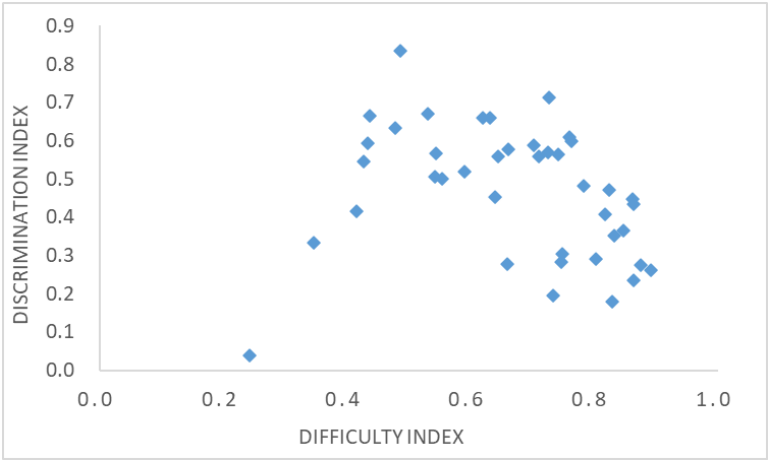

Figure 1 Difficulty Index and Discrimination Index for Computational Questions.

The hardest question has difficulty index 0.24 but with discrimination index 0.04. It's a problem to use the graph to prove the CPAM equation. It's too hard so that the discrimination index is very low, we should exclude this type of questions from our final exam paper in the future. Most of the hard questions require students to understand the mathematical derivation, which is in line with our expectations.

4. Conclusions

Only examination tests with high quality could reflect the true learning and teaching effects. From the quantitative analysis above, the quality of the final examination tests from 2016 to 2021 are quite good and consistent with moderate difficulty and excellent discrimination.

From item analysis we could identified the unqualified test questions. In the future, we should avoid putting such type of questions in the test.

More detailed analysis needs to be done in the future so that we could accumulate more excellent test questions to build our own test bank to make the future examinations more consistent and appropriate.

Acknowledgements

This research is funded by SUSTech Teaching Reform Project ("The practice of quantitative investment in teaching of security investment course for undergraduate", Grant No. Y01051946).

References

[1] Sutrisno, H. (2016) An Quality Analysis of the Mathematics School Examination Test, Jurnal Riset Pendidikan Matematika, Vol 3, pp. 162-177.

[2] Yuan, W.J., Deng, C.J., Zhu, H.X., Li, J. (2012) The Statistical Analysis and Evaluation of Examination Results of Materials Research Methods Course. Creative Education, Vol 3, Supplement, pp. 162-164.

[3] Bodie, Z., Kane, A. and Marcus, A. (2014) Investments. Global Edition, 10th Edition, McGraw Hill Higher Education.

[4] Crocker, L., Algina, J. (1986) Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

[5] Cronbach, L. (1951) Coefficient alpha and the internal structure of tests. Psychometrika, The Psychometric Society, vol. 16(3), pp. 297-334.

[6] Allen, M. J., Yen, W. M. (1979) Introduction to measurement theory. Monterey, California USA: Brooks/Cole Publishing Company.

[7] Ebel, R. L., Frisbie, D. A. (1991) Essentials of educational measurement. Englewood Cliffs, Prentice-Hall, Inc.