The Application of Reasonableness Evaluation Model Based on Coefficient of Variation and Normal Distribution in Advanced Mathematics

Zhipeng Ma¹*, Shuwan Yao², and Yuming Fang²

¹College of Automotive Engineering, Jilin University, 130022 Changchun, China
²College of Mechanical Science and Engineering, Jilin University, 130022 Changchun, China

Abstract. In order to test the learning and teaching outcomes in a comprehensive and objective way, it is necessary to have a practical examination based on a reasonable proposal. On one hand, a comprehensive evaluation model is built in the paper. Firstly, we analyze the five evaluation indicators, including the relevance, importance, difficulty, extensibility and enlightenment of knowledge points in the database. Then, the weights of these indicators are calculated by Matlab. Lastly, with the method of coefficient of variation, the model is built to formulate a reasonable Question program for advanced mathematics BI. One the other hand, another evaluation model based on the Kuder-Richarson method is built to assess the reliability of the tests. According to the two quantitative models above, a reasonable solution to the test schemes can be obtained.

1 Introduction

Examination is the most common way of teaching evaluation. The quality of the papers will directly affect the results of the examination. The paper should be reasonable, which means it should take the difficulty, the quantity of questions and range of knowledge points into account. Furthermore, it should be instructive and innovative.

In the modern society, the status of the examination has been strengthened. The examination of the scientific and accuracy requirements are more stringent [1]. At present, experts and scholars both at home and abroad have carried out research on test schemes. Jiansheng Bao and Ningzhong Shi established the curriculum difficulty model [2]; Yaomei Zhu used the comprehensive difficulty mathematical model [3] to evaluate the tests, which has a good result; Suijun Jia with the revision of the above model, established the comprehensive difficulty coefficient model constituted by factors [4]. Their researches on the design of the test have played an instructive role. However, further study in this field is hardly needed.

2 Weights of Knowledge Points Evaluation Model

2.1 Determination of Evaluation Indicators

In the distribution of knowledge points given in the questions, each knowledge point is analyzed by the five indicators of relevance, importance, difficulty, extensibility and enlightenment. The 11-point scale of 0-10 is adopted, in which the extensibility and enlightenment are subjectively scored.

The "Association Standard" in ten aspects of the "Principles and Criteria of Mathematics Education in American Schools" promulgated by the American Mathematical Teachers Association puts forward the concept of associative mathematics [5], which means that students can "recognize Application of mathematical concepts of interrelationships." A related mathematical problem can lead students to tap the internal connection of knowledge through the synthesis of multiple knowledge points so as to enhance students' understanding.

In view of this, the relevance indicator is based on the knowledge points integrated in the test questions. After analysis, the scores of the relevance indicators are obtained, as shown in Table 1.

<table>
<thead>
<tr>
<th>Comprehensive knowledge points</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scores</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

When the sample is large enough, the importance of the knowledge point is approximately linear with the score. The indicator is based on the ratio of each knowledge point’s score.
The test scores should be normal distribution in a set of high quality papers [6]. According to the characteristics of the normal distribution curve, questions of medium difficulty should be the main body. According to the survey results, questions with 75%~80% accuracy are at a medium level of difficulty. (There is not 25%~45% correct rate of the questions in the database.).

The coefficient of variation is an important statistic that measures the degree of variation of each observed value in the data. The standard deviation and average value of the indicators are calculated and normalized to confirm the weight. Calculation method is shown as (1):

$$ V_i = \frac{\sigma_i}{\bar{x}_i} (i = 1, 2, 3, 4, 5, 6, 7) $$

$$ w_i = \frac{V_i}{\sum_{i=1}^{7} V_i} (i = 1, 2, 3, 4, 5, 6, 7) $$

Where $\sigma_i$ is the standard deviation of the $i$th indicator and $\bar{x}_i$ is the average of the $i$th indicator.

Using MATLAB to calculate, the weight matrix $W1$ is obtained:

$W1=[0.3179 \ 0.2419 \ 0.1782 \ 0.1086 \ 0.1534]$.

Knowledge points score vector $S$ is:

$$ S = W1X^T $$

Where $X$ is the scoring vector of the indicator.

### 3 The Application of Knowledge Points Evaluation Model

#### 3.1 Data Sources and Analysis

The data in this paper is from 20 final exam papers of advanced mathematics BI in Jilin University. The reference book is the University Mathematics - Calculus (the first volume) (second edition) [10]. Ai represents the $i$th chapter. Aij represents the $i$th chapter, the $j$th section. Aijk represents the $i$th chapter, the $j$th section, the $k$th knowledge point.

Scoring by the program above, the top 10 points of knowledge are shown as Table 4.

<table>
<thead>
<tr>
<th>Knowledge Point Ranking table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking</td>
</tr>
<tr>
<td>Knowledge point code</td>
</tr>
</tbody>
</table>

| Knowledge point code | A421 | A261 | A162 | A511 | A132 |

It can be seen from the above table that the most proportion of the scores of the results obtained by the coefficient of variation analysis are mainly concentrated on the applications on Chapter 2 Derivatives, Chapter 4 Indefinite Integrals and Chapter 3 Median Value Theorems and Application of Derivatives.
Based on the data analyzed, a set of test paper in advanced mathematics was provided and a sample of 500 people was selected to complete the proposed paper in 120 minutes.

### 3.2 Score Distribution Analysis and Evaluation

In order to reflect the degree of students' access to knowledge and teaching feedback objectively, a high-quality paper should be more difficult and appropriate to distinguish. Reflected in the distribution of scores, it should be the approximate normal distribution. The majority is intermediate grades and high or low scores account for the minority [9]. If it matches the normal curve better, the quality is better.

Fit through the Matlab function fitting toolbox with normal distribution formula (3):

$$f(x)=a_1 \exp\left(-\left(\frac{x-b_1}{c_1}\right)^2\right)$$

The fitting result is as follows:

![The Test Papers Normal Fitting Curve](image)

**Table 5.** Fitting results

<table>
<thead>
<tr>
<th>parameter</th>
<th>data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>73.5</td>
</tr>
<tr>
<td>$b_1$</td>
<td>9.412</td>
</tr>
<tr>
<td>$c_1$</td>
<td>3.937</td>
</tr>
<tr>
<td>Correlation coefficient of DOF</td>
<td>0.9768</td>
</tr>
<tr>
<td>Complex correlation coefficient</td>
<td>0.9807</td>
</tr>
</tbody>
</table>

As shown in Table 5, both of the complex correlation coefficient and the complex correlation coefficient of the adjustment degree of normal curve fitting are close to 1, that is, the fitting degree of the curve fitting is very good. Therefore, according to the evaluation model, the quality of papers is high.

### 4 The Establishment of Rationality Evaluation Model

#### 4.1 Establishment of Rationality Evaluation Model

Captions should be typed in 9-point Times. They should be centred above the tables and flush left beneath the figures.

The rationality of the test questions mainly depend on the reliability of test questions, which reflects its stability. The reliability of an examination refers to an equivalent test on multiple groups of samples, analyzing relevant statistics and comparing the degree of consistency obtained. The reliability of questions is a reflection of the objectivity of the test and the accuracy of the test results. It is also an important indicator to evaluate the quality of the test questions [8].

The Formula 20 of Kude-Richardson method is referred to when it comes to objectivity questions [6].

$$\gamma^2 = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^{\infty} p_i (1-p_i)}{\sigma^2}\right)$$

Where $n$ represents the number of objectivity questions; $p_i$ represents the correct rate of question I; $\sigma^2$ represents the variance of total score of the subjects questions obtained by the volunteers.

In the case of subjective questions, with reference to Ge Tao questions quality evaluation model [8], the reliability value $m$ is defined as

$$m = \sqrt{\frac{N}{N-1} \left(1 - \frac{\sum_{j=1}^{\infty} \sigma_j^2}{\sigma^2}\right)}$$

Where $N$ represents the number of subjective questions, $\sigma_j^2$ means the variance of the scoring rate of question j.

### 4.2 Results of Application and Discussion

500 samples were randomly divided into five groups of 100 sub-samples, which were respectively used to analyze the reliability of objective and subjective questions.

The reliability value is calculated by MATLAB programming as Figure 2, Figure 3.

![The Reliability of Objective Questions](image)
The standard deviations of objective and subjective questions of the test papers are 0.004 and 0.021. Therefore, the reliability and stability of test questions are high. The quality of objective questions is better than that of the subjective questions.

4 Conclusion

The knowledge point evaluation model in this paper utilized the variation coefficient method to determine the weight. It has the advantages of simple algorithm and practicality. Based on it, we analyzed the arrangement of the importance of knowledge points and designed a new test paper. Through the normal curve and reliability model to evaluate the rationality of the new papers, we found that new test papers are reasonable.

In summary, a set of excellent questions must take into account the distribution of knowledge points, difficulty, differentiation and some other factors, so as to effectively evaluate the quality of the learning and teaching quality. The rationale for analyzing a test needs to consider both of the reliability and stability of test questions.

References

5. Su Yaozhong, Huang Liaoqiu, Guo Xiaoyan, Embodies the foundation, for all students, focusing on inquiry, leading the teaching reform - 2012 Shanxi province senior high school entrance examination mathematics proposition ideas. Educational Theory and Practice, Vol 32, No. 29, pp. 3-6, 2012.