Test Response Modeling Demonstrations for College Programming Language

Zhi Qiang Pu¹,a *, Wei Chen²,b

¹Network Information Center, Mianyang Normal University, Mianyang 621000, PR China
²School of Science, Southwest University of Science and Technology, Mianyang ,621010, PR China
³Email:pzq999@163.com,bEmail:chenweimy@yeah.net

Keywords: Test; Item Response Theory; Two-parameter Logistic model; Ability Evaluation.

Abstract. This article applies Item Response Models into College Computer Programming Language Test. According to the different characteristics of Programming Language Test, the author first get the Validity, the Discrimination Parameter and the Difficulty Factor of the test by using classical testing methods. Then, based on students accuracy of answering different examination questions, the author make use of Two-parameter Logistic Model to get students ability coefficient of answering different examination questions. The above mentioned model has been demonstrated and testified in a basic course test of the first-year students in a college. The results will be beneficial to improve both the teaching activities and raising teaching quality.

Introduction

Item Response Model is a significant model to test students' capabilities[1] in modern educational system. In teaching tests are an important factor[2] which can not only be used to facilitate feedbacks and evaluation of students'study, but also be used to evaluate students'information source entropy in study[3]. It can be applied both to the evaluation of students'learning and to the improvement of students'learning processes. Testing theories are scientific methods[4] used by researchers of learning in making effective and scientific judgments on the learning information collected and making decisions based on this information. Testing theories mainly consist of Classic Response Theory and Item Response Theory which offsets the deficiency of Classic Response Theory and makes a breakthrough in testing methods[5].

Item Response Theory and Item Response Model[6] are mainly applied in psychological tests, but now they have been widely used in different areas and have made great achievements. However, applying Item Response Model in college computer tests is a new topic which concerns mainly about students'ability of thinking in programming and theoretical operational capability. The evaluation of testing items includes the Discrimination Parameter and Difficulty Factor, of which the most important thing is to test students'thinking ability. In the following, the author will construct a College Programming Language Test Model and testify it.
Model Construction

Suppose the capability parameter of a student is $\theta$, according to Two-parameter Logistic Model\[7\], one can get the Probability of a student answering a question correctly as follows:

$$p(\theta) = \frac{1}{1 + \exp(-Da(\theta-b))} = \frac{1}{1 + e^{-\beta a(\theta-b)}}$$

(1)

where $a$ is Discrimination Parameter, $b$ is the Difficulty Factor of the question.

Suppose the testing items of a programming language course are Multiple Choice[8], Completion, Short Answer Questions and Programming and the accuracy rate of a student in answering Question $i$ is $x_i(i=1,2,3,4)$, according to the Logistic Model, one can predict that the probability of the student in answering four questions is:

$$P_j(\theta) = \frac{1}{1 + e^{-\beta a_j(\theta-b_j)}}$$

(2)

then, the Probability of a student answering a question wrongly is

$$Q_j(\theta) = 1 - P_j(\theta), i=1,2,3,4$$

(3)

Therefore, the Likelihood Function of the student is

$$L(\theta | x) = \prod_{j=1}^n P_j(\theta)^{x_j} Q_j(\theta)^{1-x_j}$$

(4)

According to the Maximum Likelihood Estimation Theory\[9\], the Maximum Likelihood Estimates of the Ability Parameter $\theta$ equals to the $L(\theta | x)$ when the $\theta$ is the highest.

Demonstrations

The testing subjects are two classes of first-year students from a four-year university and the total number of the students is 200. The testing paper of one of the four major courses in the final exam will be selected at random. The scores of the 200 students selected are all valid (See Table 1) and are normally distributed. (See Table 2). The examination proportion of the test paper is in accordance with the Teaching Curriculum of the course, including Multiple Choice, Completion, Short answer and Programming problem, covering all the required teaching content. The total score of the paper is 100 points, and the respective points of each part are 20 points, 30 points, 20 points and 30 points.

According the results of the testing paper (See Table 3), one can find out that the amount of calculation is a little bit large, and the Discrimination Parameter is good, which indicates that it is a good test paper and the final score of the students can basically reflect their true level in study. The students’ scores are normally distributed, which center around the average scores. The average scores of Part I and Part II are a bit high, those in Part III a bit low and the average scores of Part IV are the lowest. The scores in Part I, Part II and Part III are normal and the scores in Part IV are too low. The Standard Deviation and Degree of Dispersion in Part IV are a little big, that is, the scores of the students are uneven and the range of the distribution is wide, which indicates that the discrepancy of the students’ learning is great. The above results can be demonstrated by Item Response Theory effectively as follows:
If the accuracy rates of the student in answering the four Parts are \( x_1, x_2, x_3, x_4 \), the Likelihood Function will be[10]:

\[
L(\theta_j | x) = \prod_{i=1}^{100} P_i(\theta_j)^{n_i} Q_i(\theta_j)^{1-n_i} = \prod_{i=1}^{100} \left( \frac{1}{1 + e^{-\theta_j x_i}} \right)^{n_i} \left( 1 - \frac{1}{1 + e^{-\theta_j x_i}} \right)^{1-n_i} \quad \text{(5)}.
\]

Where \( j = 1, 2, 3, 4 \).

<table>
<thead>
<tr>
<th>Class 1</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
<th>( a )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
<th>( b_4 )</th>
<th>( b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7426</td>
<td>0.7592</td>
<td>0.7593</td>
<td>0.7123</td>
<td>0.3944</td>
<td>0.4675</td>
<td>0.4521</td>
<td>0.3375</td>
<td>0.4683</td>
<td>0.4271</td>
<td></td>
</tr>
<tr>
<td>Class 2</td>
<td>0.7407</td>
<td>0.8333</td>
<td>0.7407</td>
<td>0.7605</td>
<td>0.3792</td>
<td>0.4792</td>
<td>0.4181</td>
<td>0.3375</td>
<td>0.4203</td>
<td>0.4148</td>
</tr>
</tbody>
</table>

Where \( a_1, a_2, a_3, a_4 \) are discrimination Parameter of each class in each part are Ability parameters of Part I, Part II, Part III and Part IV, \( a \) is Comprehensive discrimination Parameter of each class. \( b_1, b_2, b_3, b_4 \) are Difficulty Factor of each class in each part are Ability parameters of Part I, Part II, Part III and Part IV, \( b \) is Comprehensive Difficulty Factor Parameter of each class.

Using computer programming and key in the data, one can get the maximum Likelihood Function of each student’s comprehensive ability. By using this formula, one can estimate and figure out the Ability Parameter of each student in each part and one can make a comparison of each student’s ability in answering different types of questions by Univariate Analysis of Dispersion. From the result of the comparison, one can see that the Significant Dispersion is less than 0.05, which illustrates that the ability of the same student in answering different types of questions is uneven.

Table 2  Average value and Standard deviation

<table>
<thead>
<tr>
<th></th>
<th>( A_1 )</th>
<th>( A_2 )</th>
<th>( A_3 )</th>
<th>( A_4 )</th>
<th>( A )</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( S_3 )</th>
<th>( S_4 )</th>
<th>( S )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>10.65</td>
<td>16.44</td>
<td>13.25</td>
<td>15.95</td>
<td>56.29</td>
<td>6.00</td>
<td>9.05</td>
<td>6.13</td>
<td>8.49</td>
<td>16.12</td>
</tr>
<tr>
<td>Class 2</td>
<td>10.42</td>
<td>17.46</td>
<td>13.25</td>
<td>17.39</td>
<td>58.52</td>
<td>6.03</td>
<td>9.96</td>
<td>5.96</td>
<td>9.21</td>
<td>15.43</td>
</tr>
</tbody>
</table>

Where \( A_1, A_2, A_3, A_4 \) are average values of each class in each part are ability parameters of Part I, Part II, Part III and Part IV, \( A \) is Comprehensive Average value of each class. \( S_1, S_2, S_3, S_4 \) are Standard deviation of each class in each part are Ability parameters of Part I, Part II, Part III and Part IV, \( S \) is Comprehensive Standard deviation Parameter of each class.

The comparison results of the students’ ability in answering the four different types of questions are as follows.

Table 3  Ability parameters

<table>
<thead>
<tr>
<th></th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \theta_3 )</th>
<th>( \theta_4 )</th>
<th>( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>0.5529</td>
<td>0.6012</td>
<td>0.8944</td>
<td>0.5731</td>
<td>0.6739</td>
</tr>
</tbody>
</table>
Where $\theta_1, \theta_2, \theta_3, \theta_4$ are Ability Parameter of each class in each part are Ability parameters of Part I, Part II, Part III and Part IV, $\theta$ is Comprehensive Ability Parameter of each class.

| Class 2 | 0.5457 | 0.6516 | 0.8731 | 0.6689 | 0.7004 |

The comparison results of the students’ ability in answering the four different types of questions are as follows: Class 1 is Stronger than Class 2 in Part I and Part III, Class 2 is Stronger than Class 1 in Part II and Part IV. The Comprehensive Ability of Class 2 is Stronger than Class 1 (See Fig. 1).

**Conclusion**

This paper makes a study on the testing theory and evaluation of College Computer Programming Language Test. It first obtains the Discrimination Parameter and the Difficulty Factor of the test by using classical methods. Then, it uses Item Response Model to get the students’ ability to Estimate between two different classes and analyzes the discrepancies in students’ ability. The results of the analysis are beneficial to the improvement of the research of Programming teaching and Programming teaching ability.

**Acknowledgements**

This work was supported by Sichuan educational Science Foundation under Grant No.13ZB0125

**References**


scale analysis for polytomous items. ProGARAMMA, Groningen, the Netherlands: iec
theory. London: International Educational and Professional Publisher.