

# A Study on a Method of Estimating the Difficulty of Quizzes Focused on Quiz Types

Shinichi IKEDA<sup>a</sup>, Teruhiko TAKAGI<sup>b</sup>, Masanori TAKAGI<sup>c</sup>, Yoshimi  
TESHIGAWARA<sup>a</sup>

<sup>a</sup> Graduate School of Engineering, Soka University, Japan

<sup>b</sup> Graduate School of Information Systems, The University of Electro-Communications,  
Japan

<sup>c</sup> Faculty of Software and Information Science, Iwate Prefectural University, Japan  
{e11m5202, teshiga}@soka.ca.jp

**Abstract:** In general, Item Response Theory (IRT) is used to quantify the difficulty level of quizzes. In order to estimate the difficulty level of the quizzes, students need to answer the quizzes in advance. However, it is hard to make students answer all quizzes in the item bank. Therefore, a method of estimating the difficulty level of not answered quizzes is studied. This study estimates the difficulty level of not answered quizzes by focusing on quiz types. Firstly, the relationship between quiz types and the difficulty level are analyzed. Next, from the result of the analyses, a procedure for estimating the difficulty level of quizzes focused on quiz types is discussed. The experiment results of comparing the proposed method with an IRT method showed that the value of the difficulty level estimated using the proposed method is close to the value estimated using IRT.

**Keywords:** IRT, E-testing, Difficulty Level, Similar Quiz, Quiz Types

## Introduction

In recent years, the test which is carried out on the web - “e-testing” has been attracting much attention [1]. In e-testing, wider range of tests can be performed by preparing as an item bank managed statistical data [2], such as based on the difficulty level of quizzes. Therefore, in e-testing, it is important to fill the item bank with a large number of quizzes. In addition, a number of researches on the support of the composition of the test by using these statistical data are done [3][4]. In these researches, Item Response Theory (IRT) [5] is used to quantify the difficulty level of quizzes. In order to estimate the difficulty level of the quizzes, students need to answer the quizzes in advance; but, it is hard to make students answer all quizzes in the item bank. Therefore, a method of estimating the difficulty level of not answered quizzes is studied.

This study estimates the difficulty level of not answered quizzes by focusing on quiz types. The difficulty level of the quizzes could be changed depending on how knowledge is asked and the difficulty level between similar quizzes has a relation. Thus, between similar quizzes, the difficulty level of not answered quizzes might be able to be estimated by focusing on the difference according to the change how knowledge is asked. In particular, the method of estimating the difficulty level of the similar quizzes focusing on quiz types by estimating the difficulty level of some quizzes with IRT is discussed. In this paper, the multiple-choice quiz is targeted.

## 1. Item Response Theory

The method of estimating the difficulty level of the similar quizzes by using IRT is studied. In IRT, a statistical model called IRT model is used to reveal the statistical property of quizzes. The property of quizzes is shown by Item Characteristic Curve (ICC) [6]. In this study, *two-parameter logistic model* (2PLM) used most commonly is applied. The probability of answering quiz ( $j$ ) correctly by the student ( $i$ ) with learning ability ( $\theta$ ) is defined as  $P_j(\theta_i)$  where  $D$  is the constant 1.7.

$$P_j(\theta_i | a, b) = \frac{1}{1 + \exp\{-Da_j(\theta_i - b_j)\}} \quad (1)$$

By using such model, it is possible to estimate the learning ability level of learners ( $\theta$ ), discrimination level ( $a$ ) and difficulty level ( $b$ ) from the test answer of the students. On the other hand, it needs to make students answer the quizzes in advance for estimating these parameters. Therefore, in this paper, the unique formula to calculate the quiz difficulty level is studied from the analyzed results of the difficulty level of each quiz type.

## 2. Method of Estimating the Difficulty Focused on Quiz Types

### 2.1 Quiz Types

In the preceding study, the quizzes were classified into 11 quiz types according to how knowledge is asked [7]. We classified based on the contents of the quiz questions or the quiz answers. Besides, we classified based on whether a quiz request to select a correct or an incorrect answer. Table 1 shows the 11 quiz types and their examples. In table 1 “Others” include computational quizzes, fill-in-the-blank quizzes or flawed quizzes and etc.

Table 1 Quiz Types

Type ID	Quiz Types and Example
Pa+	Select a correct example or explanation about a technical term. (e.g. Select a correct explanation about Morse code.)
Pa-	Select a wrong example or explanation about a technical term. (e.g. Select a wrong example about the Real-time Distributed System.)
Pb+	Select a technical term of the same type or attribute as a technical term. (e.g. Select a choice of the same type as the visual communication.)
Pb-	Select a technical term of the different type or attribute as a technical term. (e.g. Select a choice of the different type as the visual communication.)
Pc+	Select a correct example or explanation about something to relate to a technical term. (e.g. Select a correct explanation about something relating to data management on distributed environment.)
Pc-	Select a wrong example or explanation about something to relate to a technical term. (e.g. Select a wrong choice about reasons of structuring computer network.)
Pd+	Select a correct technical word about something to relate to a technical term. (e.g. Select a choice suitable for the telephone communication.)
Pd-	Select a wrong technical word about something to relate to a technical term. (e.g. Select a choice not existing in the bank or the convenience store as practical applications of computer network.)
Pe+	Select a correct combination of a technical term and explanation of it. (e.g. Select a correct explanation about 4 layers of OSI reference model.)
Pe-	Select a wrong combination of a technical term and explanation of it. (e.g. Select a wrong combination of expression and explanation of it.)
Pf	Select a correct technical term from example to explanation of it. (e.g. What is the Hostcentralized System which uses a single host computer from multiple terminals.)
	Others.

## 2.2 Similar Quiz

We define the similar quiz as “a quiz which the knowledge questioned and the knowledge needed for solution in the quiz are similar” when compared with the other quiz. It is possible to analyze the relative changes of the difficulty level of same knowledge by comparing the difficulty level between the similar quizzes. This knowledge includes a sector-specific concept, law, figures, and history and so on. Most of this knowledge is expressed by technical words. A method of managing the similar quiz automatically by using computer is proposed [8]. In this study, the quiz is classified by this method. Examples of the similar quizzes defined by our study are as follow.

## 2.3 The Analysis of Quiz Types and the Difficulty

At first, the difference of difficulty level of each quiz type is analyzed to examine the method of estimating the difficulty level. However, the amount of the item are not many, the quiz types (+) and (-) is not classified. So we analyze the quiz types Pa to Pf. Firstly, the difficulty level of some quizzes is estimated by IRT. The quizzes made by “CollabTest” [9] between 2006 and 2008 in the class “Computer Networks” are targeted, and the difficulty level of these quizzes are estimated by IRT based on the answer history of about 80 students in each year. Secondly, the similar quizzes of the quiz that difficulty level is estimated are retrieved. Thirdly, pick out the quizzes which the difficulty level is already estimated from the similar quizzes. The difference between the difficulty level of the similar quiz and that of the quiz estimated by IRT is calculated. Finally, the average difference of difficulty level for each quiz type combination is calculated.

Table 2 shows the result of calculating average difference of difficulty level for each combination of quiz types. When paying attention to the difference of the difficulty level of each quiz types, the difficulty level of the quiz type Pd and the quiz type Pc are estimated higher than the difficulty level of other quiz types. It is assumed that the difficulty level rises because it is necessary to apply knowledge and use it. Then, we propose the unique method of estimating the difficulty level based on the result of the analysis.

Table 2 The average difference in the difficulty of each combination of quiz types

Quiz Types	Pa	Pb	Pc	Pd	Pe	Pf
Pa		0.024	-0.145	-0.471	0.248	-0.018
Pb	-0.024		0.022	-0.535		-0.333
Pc	0.145	-0.022		-0.384	0.261	0.311
Pd	0.471	0.535	0.384		0.721	0.291
Pe	-0.248		-0.261	-0.721		-0.539
Pf	0.018	0.333	-0.311	-0.291	0.539	

## 2.4 Difficulty Estimation Procedure Based on Quiz Types

In this study, the difficulty level estimation which based on the quiz type is calculated using the formula proposed as below

$$PD = \frac{1}{n} \sum_{j=1}^n (sw_{ij} + d_j) \quad (3)$$

where  $PD$  is the difficulty level of the reference quiz ( $i$ ) calculated from estimated similar quiz of ( $n$ ) quizzes.  $w_{ij}$  shows the weight of combination of quiz types, and ( $s$ ) is constant that shows the difference per weight. Concrete value is set from the result of the analysis in preceding section, and  $d_j$  shows the difficulty level of estimated similar quiz. The difficulty

level of the estimated quiz to compare quiz ( $j$ ) is calculated by adding the amount of the change of the difficulty level  $sw_{ij}$  to  $d_j$  based on the weight  $w_{ij}$  of the combination of the quiz types between the estimated quiz and the compared quiz.  $n-1$  remaining quizzes are executed with the same procedure. Finally, the average value is used as difficulty level of the reference quiz ( $i$ ).

Figure 1 shows the method of estimating the difficulty level proposed in this study. First, the difficulty level of quizzes used in the test are estimated using IRT (estimated quizzes -  $d_j$ ) which is shown in Figure 1-(1). Then, from the same item bank, the target quizzes which difficulty levels need to be checked are selected as reference quizzes ( $i$ ). Following by that, the reference quizzes is used to select the similar quizzes from the estimated quizzes (estimated similar quizzes) as shown in Figure 1-(2). Next, the amount of the change of the combination of two quiz types (combined quizzes) is calculated by comparing the quiz types between reference quizzes and estimated similar quizzes as shown in Figure 1-(3). After that, the proposed formula is used to calculate the difficult level of the reference quizzes ( $PD$ ) by using the result from step 3. Finally, the calculated result ( $PD$ ) is registered.

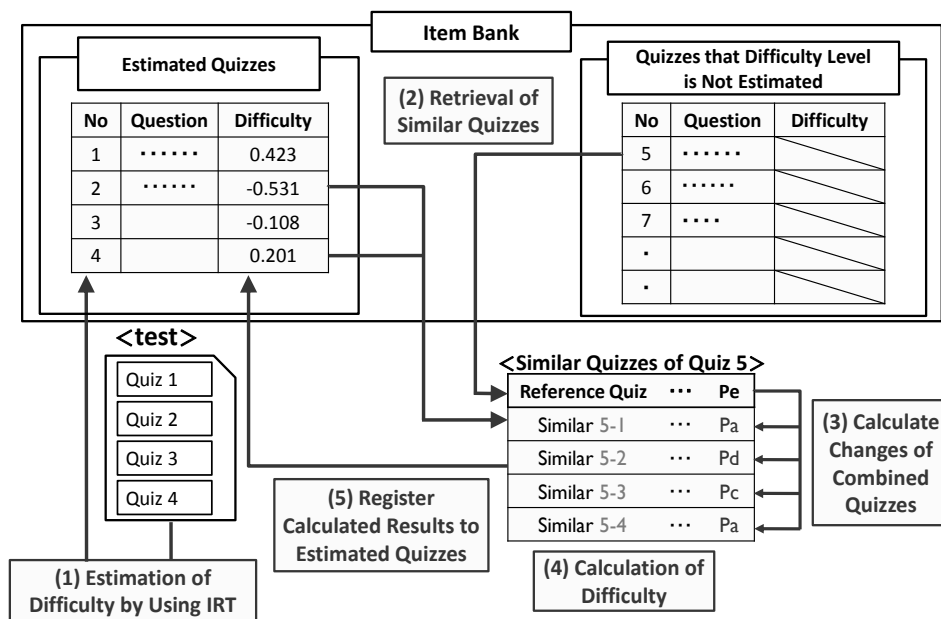


Figure 1 Calculation procedure of difficulty

### 3. Experiment

A comparative experiment is conducted to verify the relevancy of the proposed method. In this experiment, the quizzes that are analyzed in the section 2.3 are targeted, and the difficulty level of these quizzes is estimated by both IRT and proposed method. The absolute value of the difference of the difficulty level that is estimated by IRT and proposed method is calculated as a difference from the result of the analysis.

Table 3 shows the experimented result of each quiz type, and Table 4 shows the number of quizzes and the difference. From the result, the average of the difference is 0.359. From the result, there is little difference in the quiz types Pd and Pe. When the compared quizzes are three quizzes or more, the average of the difference is 0.231; the difference decreases as the number of compared quizzes increase. Therefore, in the case of some quiz types and a sufficient number of compared quizzes, the difficulty level which close to difficulty level estimated by IRT could be estimated by proposed method.

Table 3 Results for each quiz types

Quiz Types	Average of Comparison Number	Average of Difficulty Level		Average of the Difference	Number of Data
		Proposed Method	IRT		
Pa	2.6	-0.858	-0.834	0.417	20
Pb	2.6	-1.010	-0.823	0.525	6
Pc	3.8	-0.757	-0.757	0.236	5
Pd	5.5	-0.456	-0.552	0.096	4
Pe	6.7	-0.869	-0.809	0.091	3
Pf	3.5	-0.885	-0.967	0.401	11
All	3.4	-0.838	-0.833	0.359	49

Table 4 Difference for each quiz type

Comparison Number	1	2	3	4	5	6	7	8	9	10
Average of the Difference	0.58	0.38	0.25	0.26	0.00	0.26	0.29	0.33	0.04	0.01

#### 4. Conclusion

In order to estimate the difficulty level of not answered quizzes, we proposed a method of estimating the difficulty level of quizzes focused on quiz types. From the result of the compared experiment of the difficulty level, it shows that the proposed method could estimate the difficulty levels which are close to these estimated by IRT. However, the difference between IRT and the proposed method of some quizzes was large. One of the reasons of causing difference is because the proposed method does not consider the contents of the answer choices. In the multiple-choice quiz, the difficulty level could be changed depending on contents of the answer choices such as “obviously a correct choice” or “a mistakable choice”.

For the further works, we will discuss a method of estimating difficulty level of quizzes considering contents of the answer choices by focusing the selected probability and technical terms in the quiz. We pursue to propose an enhanced method to estimate the difficulty level more accurately.

#### References

- [1] Ueno, M., Nagaoka, K. (2009). e-testing: Baihukan. (in Japanese)
- [2] Japan Society for Educational Technology, (2000). Educational Technology Dictionary: Jikkyo Press. (in Japanese)
- [3] Songmuang, P., Ueno, M. (2008). Development of Prediction System of Score and Time in e-Testing: *Journal of IEICE* J91-D(9), 2225-2235. (in Japanese)
- [4] Otomo, K., (2009). Principles and Selected Applications of Item Response Theory: *Journal of IEICE* 92(12), 1008-1012. (in Japanese)
- [5] H. Toyoda (2002). Item Response Theory Introductory Version: Asakura Publisher. (in Japanese)
- [6] Baker, F. B. (2001). The Basics of Item Response Theory .2<sup>nd</sup> ed. USA, ERIC Clearinghouse on Assessment and Evaluation.
- [7] Takagi, T., Takagi, M. & Teshigawara, Y. (2008). A Proposal and Evaluation of a Similar Level Calculation Method by Type of Quizzes Created by Students. *Proc. IPSJ Summer Symposium in Setouchi*, 95-102.
- [8] Takagi, T., Takagi, M. & Teshigawara, Y. (2009). A Proposal and Evaluation of a Method of Calculating Similarity between Quizzes Created by Students. *Journal of IPSJ*, 50(10), 2426-2439. (in Japanese)
- [9] Takagi, M. Teshigawara, Y. (2006). A WBT System Enabling to Create New or Similar Quizzes Collaboratively by Students, Proceeding from ICET2006: *The Second IASTED International Conference on Education and Technology*, Calgary, Canada, 263-269.