International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 7, July 2014 **ISSN 2319 - 4847**

A NEW PROPOSAL FOR APPLYING RASCH GSP METHOD IN ASSESSMENT OF EDUCATIONAL TESTING

Tian-Wei Sheu [1], Phung-Tuyen Nguyen [2], Duc-Hieu Pham [3], Ching-Pin Tsai [4], Phuoc-Hai Nguyen [5], and Masatake Nagai [6]

[1,2,3,4,5,6] Graduate Institute of Educational Measurement and Statistics,
National Taichung University of Education, Taichung, Taiwan

ABSTRACT

In the current trend of dwindling number of students in a class, educational assessment method is required to be flexible in evaluating such as possible to handle the small samples as well as the large samples to create the unity in all cases. The purpose of this paper is to propose a new assessment method which is based on combining Grey Student Problem chart and Rasch GSP method to evaluate the question difficulty and student ability in the test. This method utilized the advantages of two mentioned model to create the creative evaluation based on the arrangement in order of statistical data according to priority criteria. The results processed by this method have been compared with results processed by previous model, the Item Response Theory one-parameter logistic, to determine the suitability of method. This comparison result showed that the fit between two methods was high, moreover, the advantage of proposed method was simple and easy to apply. Especially, it was not only best for handling small samples but also applicable for large samples. This proposed method is a more convenient way for educators to perform the evaluation process after a teaching stage.

Keywords: Grey Student Problem; question difficulty; student ability; small sample; large sample

1. INTRODUCTION

For teaching in any form, educational assessment is always essential, not only to guide the development of individual students but also to monitor and continuously improve the quality of programs and to inform and provide evidence of accountability to prospective students and their parents [1]. Whether within the scope a school or within wider scope, evaluating academic achievement of students occurs frequently to ensure the provision of timely feedback results in service of teaching, it requires evaluation results to be obtained quickly but accurately and objectively. Therefore, the design of assessment method that has high effectiveness is very interesting. An important requirement for assessment methods is that they must be valid, reliable, and practicable to meet the needs of teaching [2]. In the trend of dwindling number of students in a class in some countries such as Japan and Taiwan, a good educational assessment method not only satisfies the important criteria such as valid, reliable, and feasible but also has to be flexible in evaluating, that is it can well handle the small samples as well as the large samples to create the unity in all cases. In the classroom space, the Student-Problem chart (S-P chart) is an analytical and diagnostic method, which effectively diagnoses the results of students' learning, has been in use for many years. The main purpose of the S-P chart is to get the diagnostic data of each student, and teachers can provide better advise for each student academically depending on the analyzed data [3, 4]. In 1982, Deng proposed grey system theory wherein grey relational analysis is an effective mathematical tool, this analysis method can measure the degree of similarity or difference between two sequences based on the grade of relationship between them [5, 6]. In order to overcome the weaknesses of S-P chart which only processed dichotomous data, Nagai proposed Grey Student-Problem (GSP) chart in 2010. GSP chart is a combination of S-P chart and grey system theory to analyze S-P chart data more specifically. With GSP chart analysis, the uncertainty factors in the study are analyzed clearly [7]. In 1960, Rasch model was proposed for analyzing the test data to assess an examinee's level of ability in a particular domain such as math or reading. The aim of this model is to measure each examinee's level of a latent trait that underlies his or her scores on items of a test [8, 9]. Going abreast with Rasch model, the concept of Item Response Theory (IRT) was known during the 1960s and 1970s. The purpose of IRT is to provide a framework for evaluating how well assessments work, and how well individual items on assessments work. The most common application of IRT is in education, it was used for developing and designing exams, building item banks for exams, and equating the difficulties of items for successive versions of exams [10]. It is clear that both Rasch model and IRT are suitable for the analysis of large data, they need to have the support from computer for processing data. In practical application, the parameters of IRT are estimated by computer programs because of the vast number of parameters that must be estimated, and BILOG-MG is one of the most popular software used [11]. The view of the Rasch model was applied in GSP chart to form Rasch GSP method that was a creativity suggestion by Nagai in 2010. This method has been used to judge uncertain factors as a nonparametric statistics method [12]. It can make problem analysis more specific and clear, so it can help the students and questions be classified through the test [13]. S-P chart analysis method has been used popularly for last many years, however it is often applied for analyzing small samples. Meanwhile, BILOG-MG software processed data based on IRT to put out the results of evaluating parameters of the questions and the ability of students very clearly and specifically, but the samples which satisfied its assumptions are always large statistical data sets. It is hypothesized that it was possible to design an assessment method which not only applied for small samples but could also handle large samples. The present study proposes the combination of GSP chart with Rasch GSP method to construct a new assessment method in order to evaluate the difficulty of questions and ability of students in the test, it could be applied for both small and large samples in order to solve the faced problem as mentioned above. In that way, the proposed method would in turn be used to handle small and large samples, the obtained results are then compared to the results processed by the previous methods to test the validity and reliability of the proposed method.

2. BASIC THEORY

This study proposes a new assessment method which is based on the combination of GSP chart and Rasch GSP to evaluate difficulty of questions and ability of students, its experimental results are compared with results processed by Classical Test Theory (CTT) and IRT (using BILOG-MG 3). Thus, the basic theories needed to introduce as follows:

2.1 Rasch model

The Rasch model was named after the Danish mathematician Georg Rasch. The model shows what should be expected in responses to items (also called questions) if measurement is to be achieved [9]. The model assumes that the probability of a given respondent affirming an item is a logistic function of the relative distance between the item location and the examinee location on a linear scale. Georg Rasch first announced this model for analyzing the response of the answerers to obtain an objective interval scale that can measure the latent trait of an answerer [14]. In the Rasch model, the correct response probability of a student is a logistic function of the difference between that student's ability and the item difficulty [15]. The relation between latent trait (theta) and correct response probability is described by an item characteristic curve (ICC) (in Fig. 1).

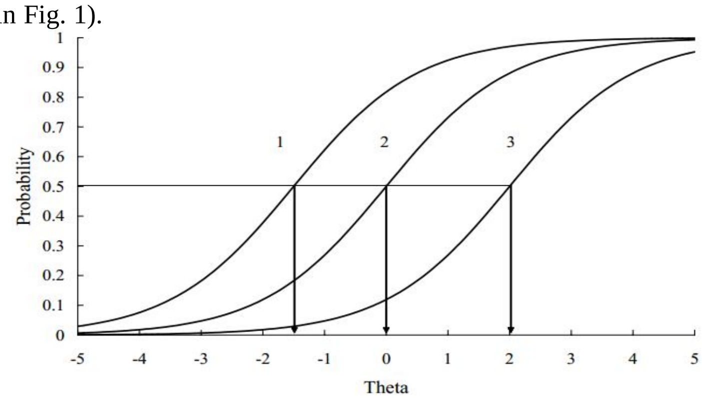


Figure 1: ICC for three different items in Rasch model

Note: Adapted from "Rasch measurement theory and application in education and psychology," by W. C. Wang, 2004, Journal of Education & Psychology 27(4), p. 644.

The Rasch model has descriptive function and predictive function. In descriptive function, this model can clearly explain the relationship between student's ability and item difficulty, the difference between students and the difference between items. In predictive function, this model can predict the probability of a student who has a specified ability to answer a specified item correctly [14].

2.2 Student - Problem chart

S-P chart is known as a method that can analyze, process, and arrange data in a defined order, it is very useful for diagnosing the learning state of student and quality of problem (also called question) [16, 17].

Definition 1: (The S-P chart matrix) Let $X = [x_{ij}]_{m \times n}$ be the S-P chart matrix, where $i = 1,2,\cdots,m$ is the order of student,

$j = 1,2,\cdots,n$ is the order of question, $m,n \in N$, and $x_{ij} = \begin{cases} 0, \text{ if answer is wrong} \\ 1, \text{ if answer is right} \end{cases}$ (1)

Caution Index for Student (CS)

$$CS_i = 1 - \frac{\sum_{j=1}^{n}(x_{ij})(x_{\bullet j}) - (x_{i\bullet})(\bar{x})}{\sum_{j=1}^{l}(x_{\bullet j}) - (x_{i\bullet})(\bar{x})} \quad \text{where } \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_{\bullet j} \text{ and } l = x_{i\bullet} = \sum_{j=1}^{n} x_{ij} \qquad (2)$$

Caution Index for Problem (CP)

$$CP_j = 1 - \frac{\sum_{i=1}^{m} (x_{ij})(x_{i\bullet}) - (x_{\bullet j})(\overline{x'})}{\sum_{i=1}^{l'} (x_{i\bullet}) - (x_{\bullet j})(\overline{x'})} \quad \text{where } \overline{x'} = \frac{1}{m}\sum_{i=1}^{m} x_{i\bullet} \text{ and } l' = x_{\bullet j} = \sum_{i=1}^{m} x_{ij} \tag{3}$$

The students would be diagnosed and classified according to the criteria of *CS* value and the rate of problems answered correctly by students, similarly, the problems would be diagnosed and classified according to the criteria of *CP* value and the rate of students answering problem correctly [18].

2.3 Grey relational analysis (GRA)

In grey system theory, grey mathematic is an effective tool to treat the uncertain, multiple, discrete and incomplete information [19]. This study refers the localized grey relational grade (LGRG) which is proposed by Nagai [20]. Its procedure as follows:

Establishment of original vectors

Definition 2: (Original vectors) The reference vector x_0 and inspected vectors of original data x_i are established:

$$\begin{aligned} x_0 &= (x_0(1), x_0(2), \cdots, x_0(j), \cdots, x_0(n)) \\ x_1 &= (x_1(1), x_1(2), \cdots x_1(j) \cdots, x_1(n)) \\ x_2 &= (x_2(1), x_2(2), \cdots x_2(j) \cdots, x_2(n)) \\ &\vdots \\ x_i &= (x_i(1), x_i(2), \cdots x_i(j) \cdots, x_i(n)) \\ &\vdots \\ x_m &= (x_m(1), x_m(2), \cdots x_m(j) \cdots, x_m(n)) \\ i &= 1,2,\cdots,m;\ j = 1,2,\cdots,n;\ m,n \in N \end{aligned} \tag{4}$$

Let $X = [x_i(j)] = [x_{ij}]_{m \times n}$ representing original data in S-P chart, if S_i is called a vector of answered states by the *i*-th student for all questions and P_j is a vector of answered states by all students for the *j*-th question, then:

$$\begin{aligned} S_i &= (x_i(1), x_i(2), \cdots, x_i(j), \cdots, x_i(n)) \\ P_j &= (x_1(j), x_2(j), \cdots, x_i(j), \cdots, x_m(j)) \end{aligned} \tag{5}$$

The data of S-P chart which have been shown in (5) will be applied for GRA [21].

Calculation of GRA: In original formula, x_0 is a reference vector of localized grey relational grade, and x_i are the inspected vectors. The established vectors have to satisfy three conditions: non-dimension, scaling, polarization. In this paper, grey relational generation uses the way: larger-the-better (the expected goal is the bigger the better).

Definition 3

(LGRG) The localized grey relational grade is defined:

$$\Gamma_{0i} = \Gamma(x_0, x_i) = \frac{\overline{\Delta}_{\max} - \overline{\Delta}_{0i}}{\overline{\Delta}_{\max} - \overline{\Delta}_{\min}} \tag{6}$$

where $\overline{\Delta}_{\max}$ and $\overline{\Delta}_{\min}$ are maximum and minimum value of $\overline{\Delta}_{0i}$ respectively, $\overline{\Delta}_{0i}$ is the absolute distance between x_0 and x_i [18].

The formula for $\overline{\Delta}_{0i}$ as follows: $$\overline{\Delta}_{0i} = \left\| x_0 - x_i \right\|_\rho = \left(\sum_{j=1}^{n} (x_0(j) - x_i(j))^\rho\right)^{1/\rho} \tag{7}$$

$\overline{\Delta}_{0i}$ is called Minkowski distance. This study applies $\rho = 2$, so $\overline{\Delta}_{0i}$ is also known as Euclidean distance. When Γ_{0i} is close to 1, it means that x_0 and x_i are highly correlated, in contract, Γ_{0i} is close to 0, the relationship between x_0 and x_i is lower.

Grey relational ordinal

The whole decision-making is made by the comparison of the grey relation Γ_{0i}. Through the ordinal, different causes can be identified, and the most important influence can be found, becoming the relational standard in the system.

2.4 GSP chart and Rasch GSP

GSP chart is the combination of GRA and S-P chart, it was developed in order to overcome the weaknesses of the S-P chart. GSP chart can make the analysis more concrete and accurate, and the uncertain factors in the studies can also be analyzed [22]. Its description is shown in Table 1.

Definition 4: (Gamma value)

In GSP chart, GS_i is the localized grey relational grade of the i-th student, and GP_j is the localized grey relational grade of the j-th problem. They are general called *Gamma* value, and in specific, LGRG-*S* is called *Gamma* value for student and LGRG-*P* is called *Gamma* value for problem [18].

$$GS_i = \Gamma_{0i} = \frac{\overline{\Delta}_{\max} - \overline{\Delta}_{0i}}{\overline{\Delta}_{\max} - \overline{\Delta}_{\min}}, i = 1,2,\cdots,m \quad (8)$$

$$GP_j = \Gamma_{j0} = \frac{\overline{\Delta}_{\max} - \overline{\Delta}_{j0}}{\overline{\Delta}_{\max} - \overline{\Delta}_{\min}}, j = 1,2,\cdots,n \quad (9)$$

Table 1: GSP chart

S \ P	**Problem-number** P_j $j = 1,2,\cdots,n$	**Total score**	**LGRG-*S***	***CS***
Student-number S_i, $i = 1,2,\cdots,m$	$X = [x_{ij}]_{m \times n}$	$SS_i = \sum_{j=1}^{n} x_{ij}$	High ↕ GS_i Low	CS
Number of correct answer	$PP_j = \sum_{i=1}^{m} x_{ij}$	$\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}$		
LGRG-*P*	More ⟷ Less GP_j			
CP	CP_j			

Note. Adapted from “An Applied Research of Rasch GSP for Evaluating Difficulty of Test Questions,” by T. W. Sheu, P. T. Nguyen D. H. Pham, P. H. Nguyen and M. Nagai, 2014, **International Journal of Application or Innovation in Engineering and Manegement** 3(3), p. 217. Copyright 2014 by the IJAIEM. Nagai applied the view of Rasch model in GSP chart to propose the Rasch GSP method that can analyze the relationship between two sets of data which were sets of the order value of students (or problems) and the localized grey relational grades. The purpose was to find a function that represented the characteristics of the entire data. This function is called Rasch GSP function, and its graph is called the Rasch GSP graph [23, 24].

Definition 5

(Rasch GSP function) Let $y = f(x) = \gamma + \frac{1-\gamma}{1+e^{-\alpha(x-\beta)}}$ be the three-parameter logistic regression function, where $\alpha, \beta, \gamma \in R$ are regression coefficients. When x is the order of student ability or the order of item difficulty and y is the localized grey relational grade, the above function $y = f(x)$ is called Rasch GSP function. If x is the order of student ability and y is LGRG-*S* then the above function $y = f(x)$ is called Rasch GSP function for students, similarly, if x is the order of item difficulty and y is the LGRG-*P* then $y = f(x)$ is called Rasch GSP function for problems [18].

3. METHODOLOGY

3.1 Item response theory with BILOG-MG software

IRT has become one of the most popular scoring frameworks for measurement data. IRT models are used frequently in computerized adaptive testing, cognitive diagnostic assessment, and test equating. One of the programs that perform for this purpose is BILOG-MG, this has proven particularly useful and reliable over recent decades for many applications. In order to fit into IRT models estimated with BILOG-MG, experimental data have to satisfy three assumptions, these are local independence, mono-tonicity, and uni-dimensionality [25]. For estimation of the IRT model parameters in BILOG-MG, the degree of bias and estimation error for parameter estimates depends on factors such as the number of parameters, number of examinees and test length. If any general guidelines can be given, it appears that for tests with between 15 and

50 items, approximately at least 250 examinees are required for the one-parameter logistic (1PL) model, and two-parameter logistic (2PL) model approximately at least 500, maybe even 1,000 examinees are required for the three-parameter logistic (3PL) model [26].

3.2 Research design with proposed method

This study creatively applies Rasch GSP method to propose new assessment method in evaluating the academic achievement of students, the procedure of this research is presented in the following flowchart (Fig. 2).

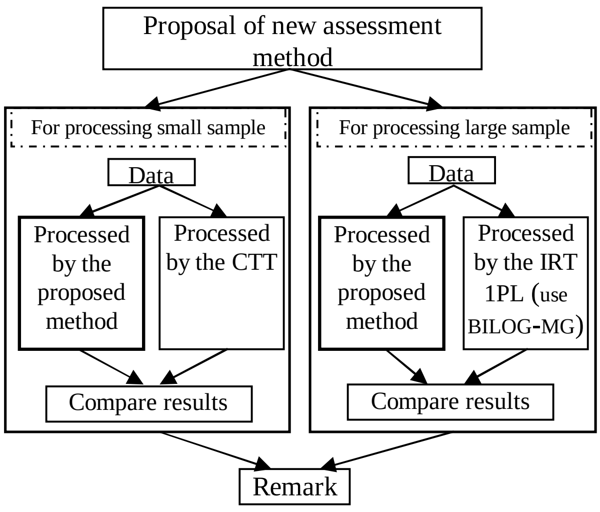


Figure 2: Flowchart for the procedure of this research

And the flowchart for applying the proposed method as follows

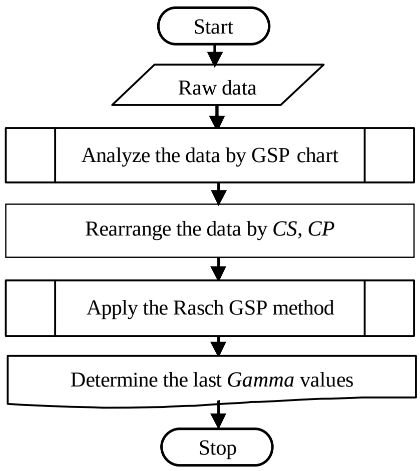


Figure 3 Flowchart for the process of proposed method

The process of processing data to evaluate the difficulty of questions in the test:

Step 1: Calculate and arrange the data with GSP chart. The original data are processed and arranged according to the rule of GSP chart, the results are presented according to specifications as in GSP chart (in the basic theory section).

Step 2: Arrange the data according to **Gamma** value. The questions are arranged in the order of their **Gamma** values (*GP*) from small to large.

Step 3: Rearrange the order according to the caution index for problem (*CP*). The questions, which have the same **Gamma** value (*GP*), are rearranged in order of *CP* value from large to small.

Step 4: Apply the Rasch GSP method for **Gamma** values arranged according to the above steps. The Rasch GSP method is applied, Rasch GSP curve is plotted for the entire data set.

Step 5: Match the new **Gamma** values for all the questions. Each question corresponds to a point on the Rasch GSP curve, the value corresponding to each question on the *y* axis is its new **Gamma** value that needs to be determined. The difficulties of questions are calculated according to their **Gamma** values, the larger the **Gamma** value a question obtains the higher its difficulty presents.

The process of processing data to evaluate the ability of examinees

Step 1: Calculate and arrange the data with GSP chart. The original data are processed and arranged according to the rule of GSP chart, the results are presented according to specification as in GSP chart.

Step 2: Arrange the order according to **Gamma** value. The examinees are arranged in the order of their **Gamma** value (*GS*) from large to small.
Step 3: Rearrange the order according to caution index for student (*CS*). The examinees who have the same *Gamma* value (*GS*), are rearranged in order of *CS* value from small to large.
Step 4: Apply the Rasch GSP method for **Gamma** values arranged according to the above steps. The Rasch GSP method is applied, Rasch GSP curve is plotted for all data.
Step 5: Match the new **Gamma** values for all the examinees. Each examinee corresponds to a point on the Rasch GSP curve, the value corresponding to each examinee on the *y* axis is its new **Gamma** value that needs to be determined. The abilities of examinees are calculated according to their **Gamma** values, the larger the **Gamma** value an examinee obtains the higher his (her) ability presents [27].

3.3 Design of Experiment

The study has conducted two different examinations to collect the data for verifying the proposed method about the ability to assess. The first examination was a 25-question Math test for junior high school student, and the collected data was a small sample with 33 answer sheets of student in Taichung, Taiwan, the obtained result has Cronbach's Alpha reaching 0.864. Using the program written in MATLAB (**Program for Rasch GSP**) to analyze the data and applied the proposed method, the obtained result would be compared with the result processed based on the Classical Test Theory (CTT). The second examination was a 48-question English test for 320 high school students also in Taichung, the test result also had high Cronbach's Alpha reaching 0.839. This satisfied the assumptions of IRT and estimation of IRT model parameters in BILOG-MG. Data were analyzed using the **Program for Rasch** *GSP* and the proposed method to give assessment results, then these data were also analyzed with the BILOG-MG 3 software, the two analytical results were compared with each other.

4. RESULTS AND DISCUSSIONS

4.1 Results

For evaluating on small sample, the results of Rasch GSP analysis were Rasch GSP curve for problems (Fig. 4) and Rasch GSP curve for students (Fig. 5) mentioned in step 4. To show the result of evaluating difficulty of each problem, a coordinate method was applied. The difficulty of questions were calculated and exported from *Program for Rasch GSP* and presented in Table 2 (in appendix), they were arranged in order from small to large value. This result was compared with the result calculated by CTT (in Table 3). The questions were highlighted in two tables to compare together to show that the order of their difficulties were arranged in the same position for the two methods. In Table 3 (seen in appendix), the difficulty p_j of the *j*-th question was calculated by the following formula:

$$p_j = \frac{n_j}{N}$$

where n_j is the number of students who answered the *j*-th question correctly, *N* is the number of students taking this test. The larger the p_j value is the smaller the question difficulty gets, so the questions are arranged by value of p_j from large to small. Similarly, the student abilities estimated by two methods were compared with each other (shown in Tables 4 and 5), students are sorted according to their ability value in order from small to large.

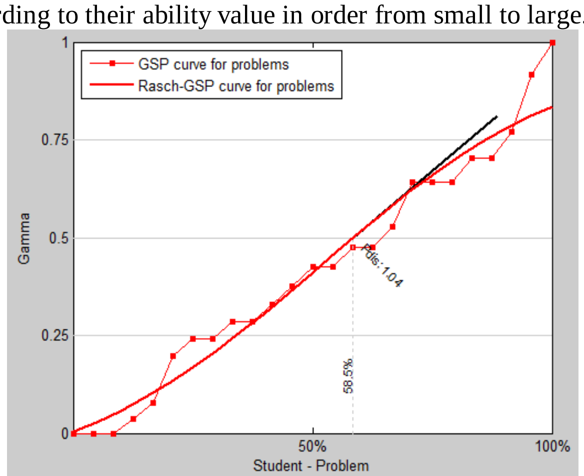


Figure 4: Rasch GSP curve for problems in the case of small sample

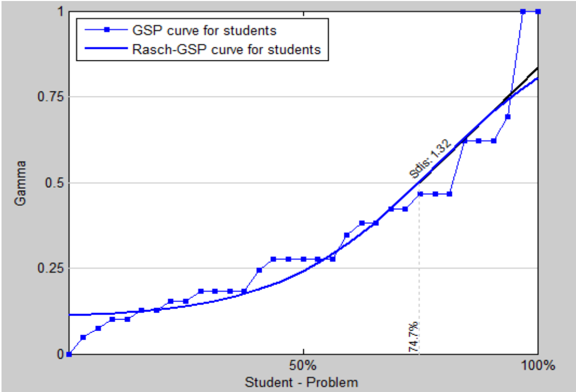


Figure 5: Rasch GSP curve for students in the case of small sample

For evaluating on large sample, the results of Rasch GSP analysis were Rasch GSP curve for problems (Fig. 6) and Rasch GSP curve for students (Fig. 7) presented. The difficulty of questions were exported from Program for Rasch GSP and shown in Table 6, they were arranged in order from small value to large value. This result was compared with the result processed by BILOG-MG 3 which presented in Fig. 10. The difficulty of questions represented by the parameter b (according to IRT) were also arranged from small value to large value (in Table 7) to compare with the results of the proposed method.

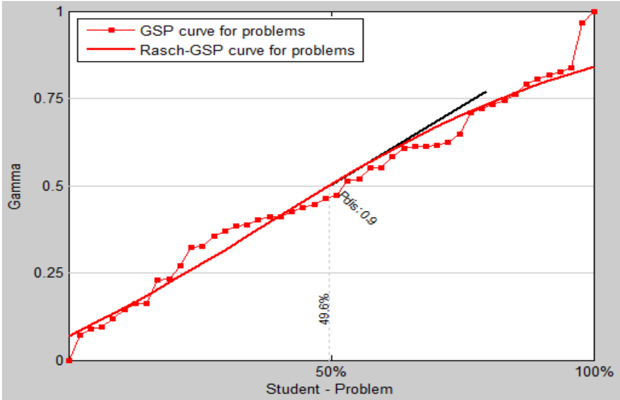


Figure 6: Rasch GSP curve for problem in the case of large sample

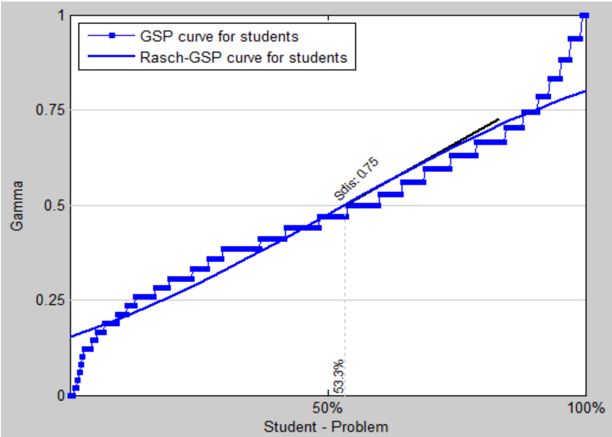


Figure 7: Rasch GSP curve for students in the case of large sample

In order to compare the result evaluated by proposed method with result evaluated by IRT 1PL model, Person-Maps were presented for both methods. Person Map (PM) gave a picture of the frequency distribution of the examinees on their ability measurement scale. In the PM, the frequency distribution of the measured ability of the students was from most able at the top to least able at the bottom. The evaluation results of two methods were presented on PM (as described in Figures 8 and 9) satisfying for the convenience of comparison. In Fig. 8, the map was plotted by the data exported from BILOG-MG 3 and modeled according to WINSTEPS software. Similarly with Fig. 9, the map was plotted by using the data exported from Program for Rasch GSP and modeled by WINSTEPS.

4.2 Discussion

There was a problem posed whether an evaluation method can be applied to both small and large sample to create unity in all cases. The present study has proposed a method to perform this duty. The experimental result showed that the proposed method has evaluated and given the difficulty of each question in the test, the ability of each student has been also estimated with determined value at the same time. In the case of small sample, the sorting the order of question difficulty by the proposed method and the CTT were similar (Tables 2 and 3), namely at highlighted position of 09 questions. At these positions, the difficulties of these questions are completely different, so they are arranged in logical order. At the positions which are not highlighted, the proposed method distinguished the difficulty of questions obviously, so they are also arranged in logical order. Meanwhile, according to the CTT, these questions are not arranged in order clearly, namely at two questions number 1 and 7, three questions number 22, 2 and 6, two questions number 9 and 19, etc., because they have the same difficulty values. This weakness is due to the fact that the CTT did not consider the priority criteria, the questions which have the same correctly answered rate, always have the same difficulty while the ability of students are different from each other. This shows the advantages of the proposed method which has already considered the quality and reliability of questions. In the same way, the students' ability are evaluated and differentiated by the proposed method clearly, so the students are arranged in order according to their ability reasonably even their correct answer rates are similar (seen in Tables 4 and 5), the students numbered 10, 14, and 21 or 17 and 22, for example. This work has not been done by the CTT, because they have been assessed having the same ability (the same correct answer rate), but their correct answer status are different from each other. In the case of large sample, the results of arrangement of questions according to their difficulties were completely similar for the results processed by two methods. This confirms the rationality and reliability of the proposed method which can be used to handle large sample. Comparing the PMs in the Figures 8 and 9, two frequency distributions of the measured ability of the students are very similar, the students who are placed in the same row are also identical in two maps, confirming again the ability to handle large samples of the proposed method. Especially, the proposed method have shown the feature to differentiate the ability of students who are on the same row in PM, they are arranged in the more logical order (Fig. 9) based on the new *Gamma* value, because these students had the correct answer rate similarly but their correct answer status were different while the questions which were answered correctly might have the difficulty differently. For example, the third line contains the students coded D08, D39, E03, A37, F01, and D16 similarly for two PMs, they were not arranged in order by processed result of BILOG-MG 3 because they were estimated having the ability relatively similar. However, they were arranged in order from the right to the left by proposed method, because they were estimated beyond with different stability of ability according to new *Gamma* value: 0.784, 0.783, 0.781, 0.780, 0.778, and 0.776 respectively. *T*his is striking advantage of the proposed method.

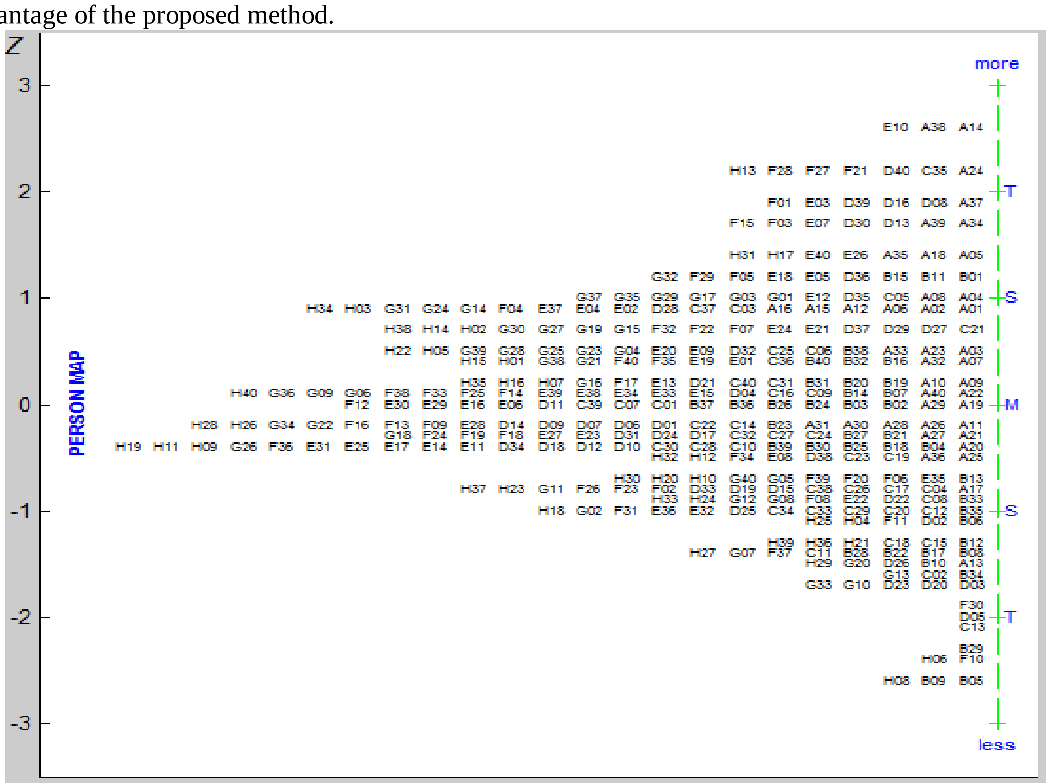


Figure 8: The Person Map for evaluation result processed by IRT 1PL model, it is modeled according to WINSTEPS (M: mean, S: one standard deviation (SD), T: two SDs).

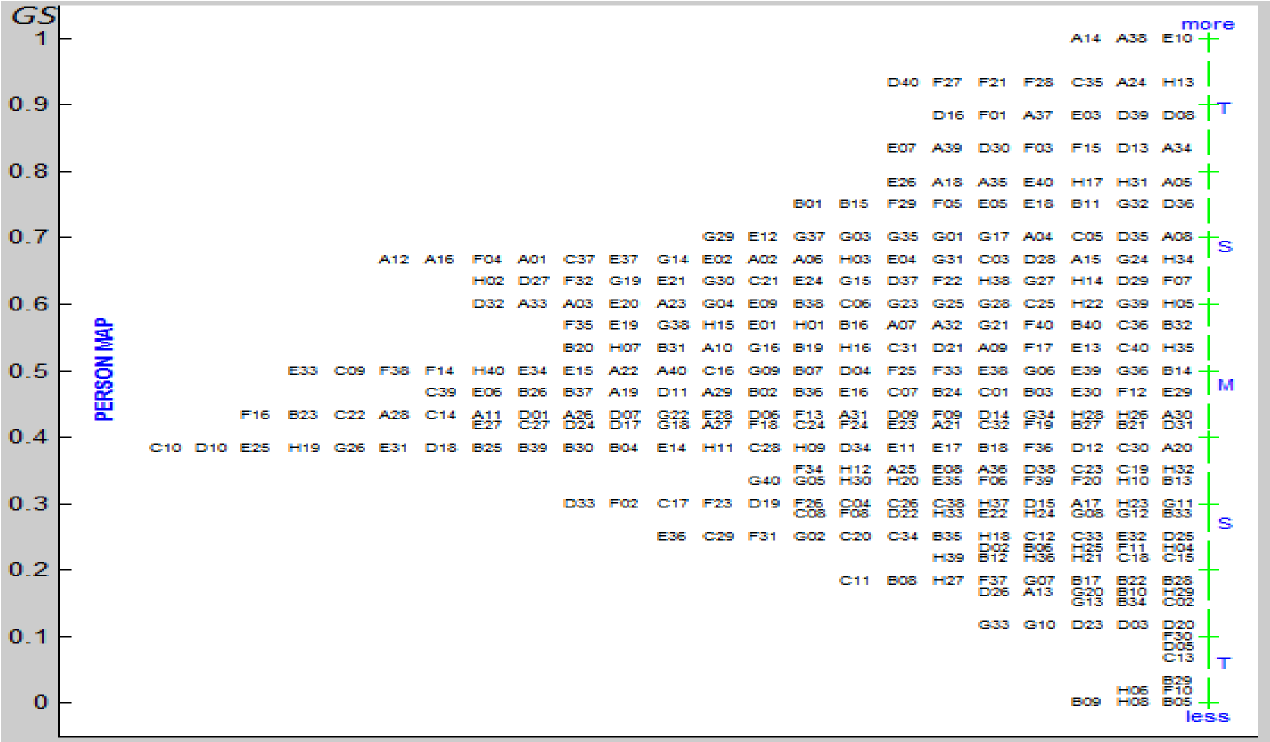


Figure 9: The Person Map for evaluation result processed by proposed method, it is modeled according to WINSTEPS (M: mean, S: one standard deviation (SD), T: two SDs)

This study has randomly conducted the tests that the participants were students of the same class, the same school or the wider range of students who learned at the same grade in a region. In these cases, the proposed method can handle data to give output results accurately and objectively. Of course, in any case, the questions of test must be designed suitably for the obtained results to have high reliability. Because the proposed method can be applied to both small and large sample form, so wherever and whenever the teachers and education managers can use it to assess the student ability and question difficulty through the tests after a period of teaching. Especially with the advantage in the evaluation of small samples, the proposed method will be useful and meets the needs of assessment and measurement in the situation of the number of students increasingly declined in a class. This creates favorable conditions for the educators to perform their tasks to achieve teaching objectives.

Advantages and limitation of the study: The proposed method overcame the disadvantages in the calculation of question difficulty and student ability by the CTT, especially the ability that not only handled small samples well but also can apply for large samples. The limitation of this study was that it has not built the scale of question difficulty and student ability scale that can compare with the scale in IRT. However, this task is enormous beyond the scope of the research.

5. CONCLUSION

There was a hypothesis that it was possible to build an educational assessment method which can be applied to both small and large statistical samples to create the unity in all cases. The present study proposed a new assessment method based on the combination of GSP and Rasch GSP to evaluate the difficulty of questions and the ability of students through the test. The findings of this study are as follows:

(1) The study has developed an assessment method that does not only well apply for small statistical samples, but also can handle large statistical samples.

(2) Evaluation results of the proposed method are compared with the evaluation results of the previous methods. The fit between the results is relatively high and shows the advantages of the new method.

(3) The advantage of the proposed method is that it is simple and easy to apply; the processed results are accurate and objective.

The striking advantage is the ability to logically arrange the objects (persons or items) that have the same correct answer rate. The feature of processing small sample helps teachers and education managers to have more convenience to perform their evaluation in the new context.

REFERENCES

[1] L. F. Gardiner, Redesigning Higher Education: Producing Dramatic Gains in Student Learning. ASHE-ERIC Higher Education Report No. 7: ERIC, 1994.
[2] E. D. Pulakos, "Selection assessment methods," SHRM Foundation's Effective Practice Guidelines, Society for Human Resource Management, Alexandria, VA, p. 55, 2005.
[3] M. N. Yu, Educational Testing and Assessment, Third ed. Taiwan: Psychology Publisher, 2011.
[4] D. McArthur, Analysis of Test Score Patterns: The Student-problem (SP) Technique: University of California, Center for the Study of Evaluation, 1983.
[5] J. L. Deng, "Introduction to Grey System Theory," Journal of Grey System, vol. 1, pp. 1-24, 1989.
[6] K. L. Wen, C. Chao, H. Chang, S. Chen, and H. Wen, Grey system theory and applications. Taipei: Wu-Nan Book Inc, 2009.
[7] T. W. Sheu, C. P. Tsai, J. W. Tzeng, D. H. Pham, H. J. Chiang, C. L. Chang, et al., "An Improved Teaching Strategies Proposal Based on Students' Learning Misconceptions," International Journal of Kansei Information, vol. 4, pp. 1-12, 2013.
[8] G. Karabatsos, "The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory," Journal of Applied Measurement, vol. 2, pp. 389-423, 2001.
[9] A. Tennant and P. G. Conaghan, "The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?," Arthritis Care & Research, vol. 57, pp. 1358-1362, 2007.
[10] S. E. Embretson and S. P. Reise, Item response theory: Psychology Press, 2013.
[11] M. Du Toit, IRT from SSI: Bilog-MG, multilog, parscale, testfact: Scientific Software International, 2003.
[12] T. W. Sheu, J. W. Tzeng, J. C. Liang, B. T. Wang, and M. Nagai, "The Use of Rasch Model GSP Chart and Grey Structural Model Analysis for Vocational Education and Training Courses: Taking Enterprise Ethics and Professional Ethics Courses as an Example," Journal of Educational Research and Development, vol. 8(4), pp. 53-80, 2012.
[13] J. W. Tzeng, T. W. Sheu, J. C. Liang, B. T. Wang, and M. Nagai, "A new proposal based on Rasch model GSP chart and grey structural model with analysis and discussion," International Journal of Advancements in Computing Technology, pp. 111-121, 2012.
[14] W. C. Wang, "Rasch Measurement Theory and Application in Education and Psychology," Journal of Education & Psychology, vol. 27, pp. 637-694, 2004.
[15] F. B. Baker, The basics of item response theory, Second ed. United States of America: ERIC Clearinghouse on Assessment and Evaluation., 2001.
[16] Y. C. Ho, "An Experimental Study of the Effects of Mastery Learning Combined with the Diagnosis of Microcomputerized S-P Chart Analysis on Students' Learning," Educational Psychology, vol. 22, pp. 191-214, 1989.
[17] S. C. You and M. N. Yu, "The Relationships among Indices of Diagnostic Assessments Knowledge Structures and S-P Chart Analysis," Education and Psychology, vol. 29, pp. 183-208, 2006.
[18] T. W. Sheu, P. T. Nguyen, D. H. Pham, P. H. Nguyen, and M. Nagai, "An Applied Research of Rasch GSP for Evaluating Difficulty of Test Questions," International Journal of Application or Innovation in Engineering & Management, vol. 3, pp. 213-223, 2014.
[19] Y. Li, L. Zhao, and R. Li, "Novel model on assessment of construction project investment environment," Pak. J. Statist, vol. 29, pp. 989-998, 2013.
[20] D. Yamaguchi, G.-D. Li, and M. Nagai, "Verification of effectiveness for grey relational analysis models," Journal of Grey System, vol. 10, pp. 169-181, 2007.
[21] P. T. Nguyen, P. H. Nguyen, D. H. Pham, C. P. Tsai, and M. Nagai, "The Proposal for Application of Several Grey Methods in Evaluating and Improving the Academic Achievement of Students," Journal of Taiwan Kansei Information, vol. 4, pp. 179-190, 2013.
[22] T. W. Sheu, T. L. Chen, J. W. Tzeng, C. P. Tsai, H. J. Chiang, C. L. Chang, et al., "Applying Misconception Domain and Structural Analysis to Explore the Effects of the Remedial Teaching," Journal of Grey System, vol. 16, pp. 17-34, 2013.
[23] J. W. Tzeng, T. W. Sheu, J. C. Liang, B. T. Wang, and M. Nagai, "A New Proposal Based on Rasch Model GSP Chart and Grey Structural Model with Analysis and Discussion.," International Journal of Advancements in Computing Technology, vol. 4, pp. 111-121, 2012.
[24] T. W. Sheu, D. H. Pham, P. T. Nguyen, C. P. Tsai, P. H. Nguyen, and M. Nagai, "RGSP Toolbox 1.0 for Educational Achievement," Open Journal of Communications and Software, vol. 1, pp. 1-10, January 2014 2014.
[25] A. A. Rupp, "Item response modeling with BILOG-MG and MULTILOG for Windows," International Journal of Testing, vol. 3, pp. 365-384, 2003.
[26] C. L. Hulin, R. I. Lissak, and F. Drasgow, "Recovery of two-and three-parameter logistic item characteristic curves: A Monte Carlo study," Applied psychological measurement, vol. 6, pp. 249-260, 1982.

[27] Y. C. Hsieh, C. I. Lee, and K. K. Chu, "Effect of an ontology-based reasoning learning approach on cognitive load and learning achievement of secondary school students," Pak. J. Statist, vol. 29, pp. 561-572, 2013.

APPENDICES

Table 2: Evaluation result for question difficulty by the proposed method in the case of small sample

Question-number	**11**	**13**	**10**	**1**	**7**	**22**	**2**	**6**	**24**	**9**	**19**	**12**	**18**
Difficulty	0.164	0.187	0.213	0.241	0.271	0.304	0.340	0.378	0.418	0.459	0.502	0.546	0.590
Question-number	**17**	**14**	**20**	**4**	**16**	**23**	**21**	**25**	**3**	**8**	**15**	**5**	
Difficulty	0.633	0.676	0.718	0.759	0.797	0.833	0.866	0.898	0.926	0.952	0.975	0.996	

Table 3: Evaluation result for question difficulty by the classical theory test in the case of small sample

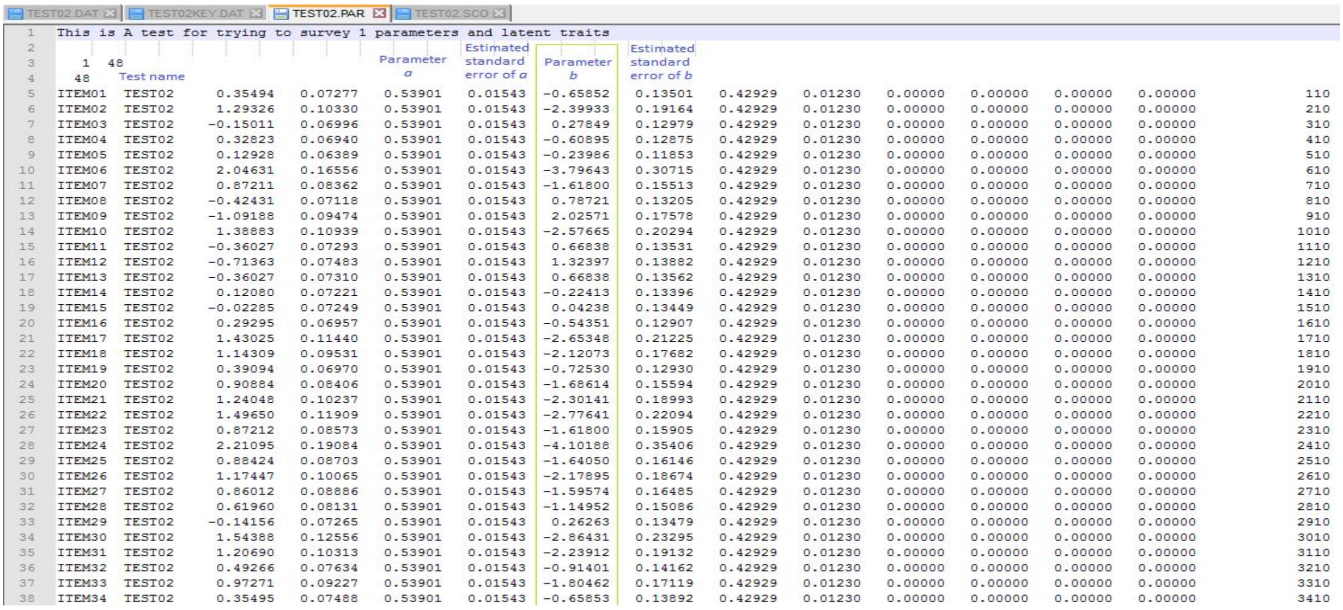
Question-number	**11**	**13**	**10**	**1**	**7**	**22**	**2**	**6**	**24**	**9**	**19**	**12**	**18**
p_j	0.878	0.848	0.788	0.758	0.758	0.727	0.727	0.727	0.667	0.636	0.636	0.606	0.606
Question-number	**17**	**14**	**20**	**4**	**16**	**23**	**21**	**25**	**3**	**5**	**8**	**15**	
p_j	0.576	0.545	0.515	0.515	0.485	0.485	0.455	0.364	0.333	0.303	0.303	0.303	

Table 4: Evaluation result for student's score by the proposed method in the case of small sample

Student-number	**2**	**7**	**6**	**18**	**24**	**30**	**10**	**14**	**21**	**17**	**22**	**16**	**25**
Score	0.790	0.790	0.742	0.706	0.649	0.649	0.587	0.546	0.504	0.463	0.424	0.387	0.352
Student-number	**33**	**1**	**9**	**20**	**27**	**29**	**31**	**13**	**15**	**28**	**32**	**3**	**8**
Score	0.320	0.291	0.265	0.241	0.221	0.203	0.188	0.175	0.163	0.154	0.146	0.139	0.133
Student-number	**5**	**11**	**12**	**26**	**23**	**19**	**4**						
Score	0.128	0.124	0.121	0.118	0.116	0.114	0.112						

Table 5: Evaluation result for student's score by the classical theory test in the case of small sample

Student-number	**2**	**7**	**6**	**18**	**24**	**30**	**10**	**14**	**21**	**17**	**22**	**16**	**25**
Score	1.000	1.000	0.920	0.880	0.880	0.880	0.760	0.760	0.760	0.720	0.720	0.680	0.680
Student-number	**33**	**1**	**9**	**20**	**27**	**29**	**31**	**13**	**15**	**28**	**32**	**3**	**8**
Score	0.640	0.560	0.560	0.560	0.560	0.560	0.520	0.440	0.440	0.440	0.440	0.400	0.400
Student-number	**5**	**11**	**12**	**26**	**23**	**19**	**4**						
Score	0.360	0.360	0.320	0.320	0.280	0.240	0.160						


```
TEST02.DAT   TEST02KEY.DAT   TEST02.PAR   TEST02.SCO
This is A test for trying to survey 1 parameters and latent traits

   1   48                                   Parameter  Estimated   Parameter  Estimated
  48   Test name                            a          standard    b          standard
                                                       error of a             error of b
ITEM01 TEST02  0.35494 0.07277 0.53901 0.01543 -0.65852 0.13501 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  110
ITEM02 TEST02  1.29326 0.10330 0.53901 0.01543 -2.39933 0.19164 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  210
ITEM03 TEST02 -0.15011 0.06996 0.53901 0.01543  0.27849 0.12979 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  310
ITEM04 TEST02  0.32823 0.06940 0.53901 0.01543 -0.60895 0.12875 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  410
ITEM05 TEST02  0.12928 0.06389 0.53901 0.01543 -0.23986 0.11853 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  510
ITEM06 TEST02  2.04631 0.16556 0.53901 0.01543 -3.79643 0.30715 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  610
ITEM07 TEST02  0.87211 0.08362 0.53901 0.01543 -1.61800 0.15513 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  710
ITEM08 TEST02 -0.42431 0.07118 0.53901 0.01543  0.78721 0.13205 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  810
ITEM09 TEST02 -1.09188 0.09474 0.53901 0.01543  2.02571 0.17578 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000  910
ITEM10 TEST02  1.38883 0.10939 0.53901 0.01543 -2.57665 0.20294 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1010
ITEM11 TEST02 -0.36027 0.07293 0.53901 0.01543  0.66838 0.13531 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1110
ITEM12 TEST02 -0.71363 0.07483 0.53901 0.01543  1.32397 0.13882 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1210
ITEM13 TEST02 -0.36027 0.07310 0.53901 0.01543  0.66838 0.13562 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1310
ITEM14 TEST02  0.12080 0.07221 0.53901 0.01543 -0.22413 0.13396 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1410
ITEM15 TEST02 -0.02285 0.07249 0.53901 0.01543  0.04238 0.13449 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1510
ITEM16 TEST02  0.29295 0.06957 0.53901 0.01543 -0.54351 0.12907 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1610
ITEM17 TEST02  1.43025 0.11440 0.53901 0.01543 -2.65348 0.21225 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1710
ITEM18 TEST02  1.14309 0.09531 0.53901 0.01543 -2.12073 0.17682 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1810
ITEM19 TEST02  0.39094 0.06970 0.53901 0.01543 -0.72530 0.12930 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 1910
ITEM20 TEST02  0.90884 0.08406 0.53901 0.01543 -1.68614 0.15594 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2010
ITEM21 TEST02  1.24048 0.10237 0.53901 0.01543 -2.30141 0.18993 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2110
ITEM22 TEST02  1.49650 0.11909 0.53901 0.01543 -2.77641 0.22094 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2210
ITEM23 TEST02  0.87212 0.08573 0.53901 0.01543 -1.61800 0.15905 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2310
ITEM24 TEST02  2.21095 0.19084 0.53901 0.01543 -4.10188 0.35406 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2410
ITEM25 TEST02  0.88424 0.08703 0.53901 0.01543 -1.64050 0.16146 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2510
ITEM26 TEST02  1.17447 0.10065 0.53901 0.01543 -2.17895 0.18674 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2610
ITEM27 TEST02  0.86012 0.08886 0.53901 0.01543 -1.59574 0.16485 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2710
ITEM28 TEST02  0.61960 0.08131 0.53901 0.01543 -1.14952 0.15086 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2810
ITEM29 TEST02 -0.14156 0.07265 0.53901 0.01543  0.26263 0.13479 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 2910
ITEM30 TEST02  1.54388 0.12556 0.53901 0.01543 -2.86431 0.23295 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 3010
ITEM31 TEST02  1.20690 0.10313 0.53901 0.01543 -2.23912 0.19132 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 3110
ITEM32 TEST02  0.49266 0.07634 0.53901 0.01543 -0.91401 0.14162 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 3210
ITEM33 TEST02  0.97271 0.09227 0.53901 0.01543 -1.80462 0.17119 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 3310
ITEM34 TEST02  0.35495 0.07488 0.53901 0.01543 -0.65853 0.13892 0.42929 0.01230 0.00000 0.00000 0.00000 0.00000 3410
```

Figure 10: Estimated result for item parameter from BILOG-MG 3 in the case of large sample

Table 6: Evaluation result for question difficulty by proposed method in the case of large sample

Question-number	**24**	**6**	**30**	**22**	**38**	**17**	**10**	**2**	**21**	**31**	**26**	**18**
Difficulty	0.160	0.169	0.178	0.189	0.199	0.210	0.221	0.233	0.246	0.259	0.272	0.286
Question-number	**33**	**20**	**25**	**23**	**7**	**27**	**35**	**42**	**41**	**46**	**28**	**37**
Difficulty	0.300	0.315	0.330	0.346	0.363	0.379	0.396	0.414	0.432	0.450	0.468	0.487
Question-number	**32**	**43**	**36**	**19**	**34**	**1**	**4**	**16**	**45**	**40**	**44**	**5**
Difficulty	0.506	0.526	0.545	0.565	0.585	0.604	0.624	0.644	0.664	0.684	0.703	0.722
Question-number	**14**	**15**	**29**	**3**	**13**	**11**	**8**	**48**	**47**	**39**	**12**	**9**
Difficulty	0.742	0.761	0.779	0.797	0.815	0.833	0.850	0.867	0.884	0.899	0.915	0.930

Table 7: Evaluation result for question difficulty from BILOG-MG 3 in the case of large sample

Question-number	**24**	**6**	**30**	**22**	**38**	**17**	**10**	**2**	**21**	**31**	**26**	**18**
Difficulty	-4.102	-3.796	-2.864	-2.776	-2.734	-2.653	-2.577	-2.399	-2.301	-2.239	-2.179	-2.121
Question-number	**33**	**20**	**25**	**23**	**7**	**27**	**35**	**42**	**41**	**46**	**28**	**37**
Difficulty	-1.805	-1.686	-1.641	-1.618	-1.618	-1.596	-1.487	-1.323	-1.323	-1.168	-1.150	-0.949
Question-number	**32**	**43**	**36**	**19**	**34**	**1**	**4**	**16**	**45**	**40**	**44**	**5**
Difficulty	-0.914	-0.827	-0.776	-0.725	-0.659	-0.659	-0.609	-0.544	-0.527	-0.463	-0.383	-0.240
Question-number	**14**	**15**	**29**	**3**	**13**	**11**	**8**	**48**	**47**	**39**	**12**	**9**
Difficulty	-0.224	0.042	0.263	0.278	0.668	0.668	0.787	0.963	1.129	1.186	1.324	2.026

AUTHORS

Tian-Wei Sheu received the Ph.D. degree in Mathematics from National Osaka University, Japan in 1990. He is the Dean of College of Education and a professor of Graduate Institute of Educational Measurement, National Taichung University, Taichung, Taiwan. His studies focus in IRT, Educational Measurement, and e-Learning, etc. He is the director of TKIA (Taiwan *Kansei* Information Association).

Phung-Tuyen Nguyen is currently a Ph.D. candidate in Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taiwan. He received Master's degree in Physics, in 2003 from Hanoi University of education, Vietnam. His research interests focus on item response theory, grey system theory, and educational measurement.

Duc-Hieu Pham received Master's degree of education at Hanoi Pedagogical University N^{o}2 of Vietnam in 2009. He works as a lecturer the Primary Education Faculty of Hanoi Pedagogical University N^{o}2, Vietnam. He is currently a Ph.D. candidate in Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taiwan. His research interests include grey system theory, educational measurement and primary education.

Ching-Pin Tsai received his Ph.D degree in Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education, Taiwan in 2014. He is a Mathematics Teacher in Changhua County Hsiu Shui Junior High School, Taiwan now. His research interests include numerical analysis, grey system theory, MSM theory, RGSM theory, kansei engineering, actions for testing of teaching, mathematics education and chaotic behavior of electron tunneling.

Phuoc-Hai Nguyen received Master's degree in Biology from Hanoi University of education of Vietnam in 2006. He is currently a Ph.D. candidate in Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taiwan. His research interests include biology, item response theory, grey system theory, ordering theory and educational measurement.

Masatake Nagai received his Master's degree in Engineering from Toukai University of Japan in 1969. He worked in Oki Electric Industry Co., Ltd. for 18 years and was mainly engaged in the design development of ME systems, communication network systems, OA systems, etc. He was also a researcher (Dr. Matsuo research) at the TohokuUniversity while working toward his Ph.D in Engineering. From 1989, he worked at the Teikyo University Department of Science and Engineering as an assistant professor and eventually as an engineering professor. Chair professor in Graduate Institute of Educational Measurement, National Taichung University, Taiwan now. His research interests include approximation, strategy system engineering, information communication network technology, agent, *kansei* information processing, grey system theory and engineering application. A regular of IEICE member.