Item Analyses and Test Equating Using Rasch-GZ

Jinyu Wu[1], Xue Yang[2], Jinggang Wei[3] and Quan Zhang [4] (ID)

[1] University of Electronic Science and Technology of China,
Zhongshan Institute, Guangdong, China
[2] Zhongshan Overseas Chinese Secondary School, Guangdong, China
[3]City Training Institute (CTI), Guangzhou, China/Tarlac State University, the Philippines
[4]Jiaxing University, Zhejiang, China
qzhang141@rasch-gz.com

Abstract: This article is a user guide to RASCH-GZ, the first Chinese platform for both CCT-based Item Analysis and Rasch-based Test Equating successfully developed over the past three years (2019-2023). It provides Chinese-speaking researchers and scholars with lots of help in promoting the use of Rasch model, contributing to the development and popularity in Rasch-based research for language testing in China. The present article is part of the one-day pre-conference workshop run by the program developers, providing the basic operations, the topics of item analyses and test equating and the interpretation of result files.

Key words: Rasch Model, Test Equating, Linking Items, Item Analysis, Rasch-GZ

1. Introduction

Of all the Rasch-based measurement methods, researchers have been using various kinds of systems such as WINSTEPS, FACET [1], R-, etc. because the Rasch model is powerful and feasible.[2], [3], [4],[5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22].

In China, (GITEST) system [23] was held as the only member of Rasch software family. Developed in 1980's, (GITEST) runs under DOS, calibrating small data matrix of (200 items X 10,000 candidates). (GITEST) was used to conduct the ten-year (1990-1999) equating project of Matriculation English Test (MET) sponsored by the Examination Center under Ministry of Education, China. In this sense, (GITEST) played an important role in MET equating project, making a solid foundation for the standardized tests to be implemented in China. However, with the advent of AI era and the sustained improvement for computer technology as well as application, such a DOS version can no longer meet the current needs. It is motivated by this that the authors have managed to comprehensively updated GITEST system to RASCH-GZ[1].

Starting from PROMS2021, Nanjing, China, pre-conference workshops about introduction to and application of RASCH-GZ have been conducted. The present article is the one-day pre-conference workshop material introducing the basic operation, via illustration, of both item analysis and test equating of (RASCH-GZ).

2. Basic operation[2] of RASCH-GZ system

---

[1]http://www.rasch-gz.com

To avoid complicated technical terms, the authors try to use simple language, diagrams and texts to guide users how to get familiar with RASCH-GZ. The operation of RASCH-GZ falls into six parts. In this section, the first three parts will be introduced.

2.1.Booting and login,

2.2. Data editing and inputting, and

2.3.Item analysis for test items including the evaluation report for a test paper.

2.1. System login

First, to start the RASCH-GZ system, we click on the program icon to obtain the login interface on the screen as shown in Figure 1 below.

Fig. 1. Login interface of Rasch-GZ

To perform system login, we need follow the four steps below:

Step 1:Key in the user name and the password (to be created by the users);

Step 2: To get help, click the "Help"button (if you need);

Step 3: To exit the system, click the "Exit" button;

Step 4:To enter the system, click the "Login" button to get the data entry interface as shown in Figure 2 below.
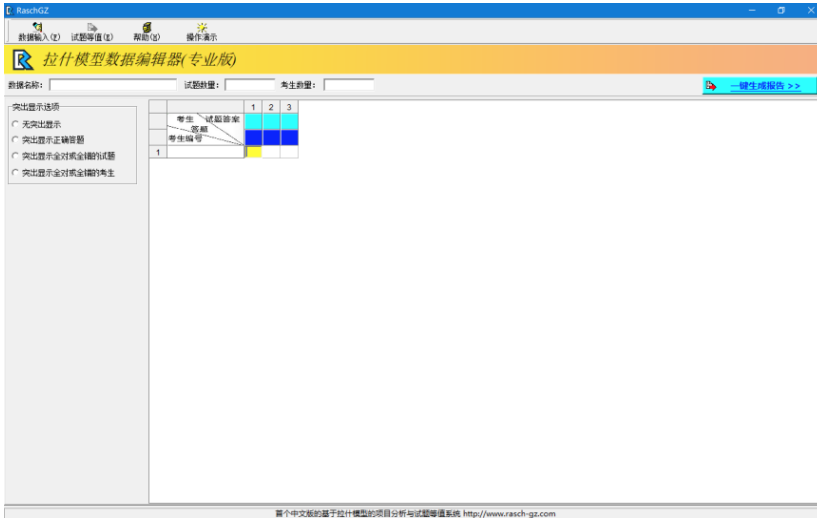

Fig. 2. The Rasch-GZ data entry interface

2.2. Data input and storage

RASCH-GZ data can be input either by hand entry or data file import.

The data type and format. (Rasch-GZ) accepts text file, EXCEL file, CSV and so on. The data type can be of chars like ABCDACDABACD, or integers such as 101010101111100 wherein '1' represents a correct answer, while '0', an incorrect answer. The data matrix is of 30-item by 40-subject is illustrated in Figure 3 as follows.

Data files to be created manually. To create data files manually, we need follow the example of data matrix illustrated in Figure 3 above. Once done, the current data file is saved.Now, Rasch-GZ will pop up a window, indicating that the user has successfully saved the data with the file name of "GD08" defined by the user. And the system automatically adds 'dat', the suffix to "GD08.dat", indicating this is a data file.

Next, with just one click, the user could obtain the report file of analysis.

External data files to be imported. To import external data, the user needs to have a data file edited in Rasch-GZ format (as shown in Figure 3). With a click of mouse, they would obtain an analysis report file they specified.

To import text data, click "Import GITEST Data" from the "Data Input" menu in the upper left corner of the menu bar, and the following dialog box will pop up as seen in Figure 4 below. By now the user simply selects the corresponding GITEST format data file and opens the current data window to import data.

To import EXCEL data, click "Import EXCEL Data" from the "Data Input" menu in the menu bar. The dialog box will appear (as seen Figure 5 below).We select
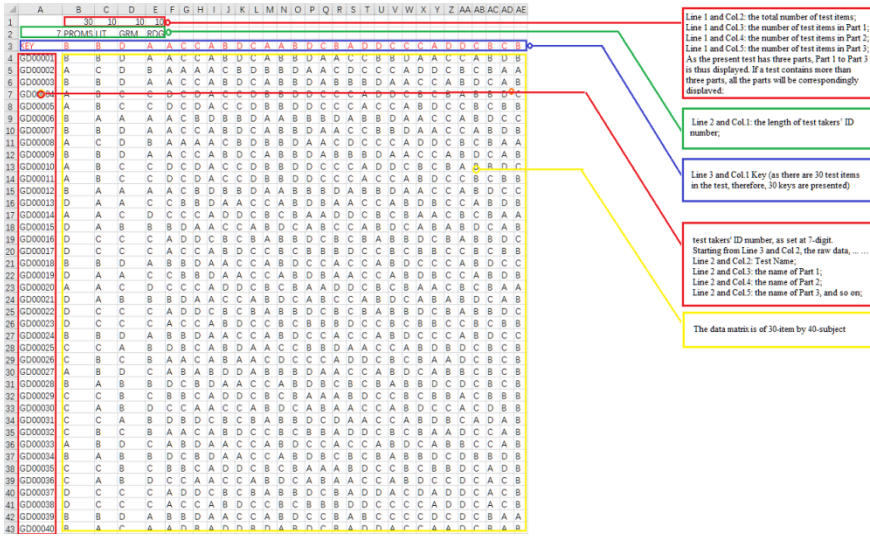
the corresponding EXCEL format data file, and open the current data editing window to import data. Rasch-GZ accepts older versions as well as versions after 2007.

Fig. 3. The data format set in the data file with the interpretation for each part in the data file
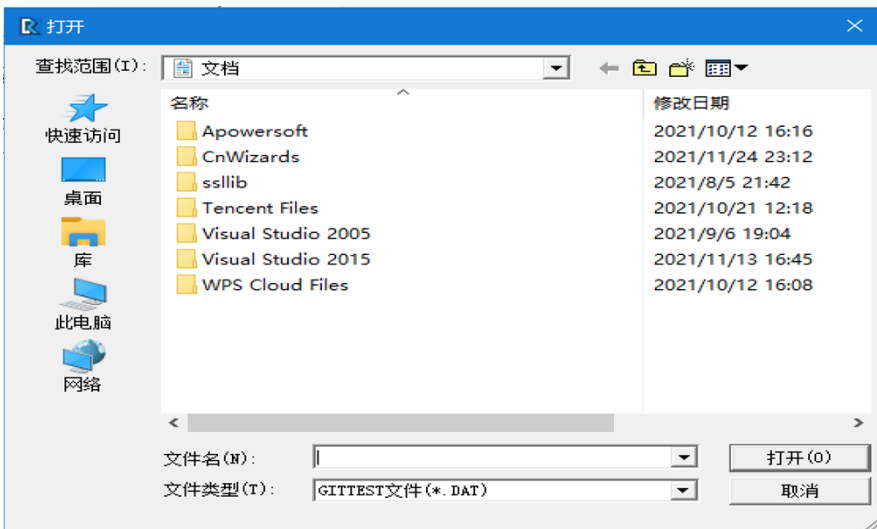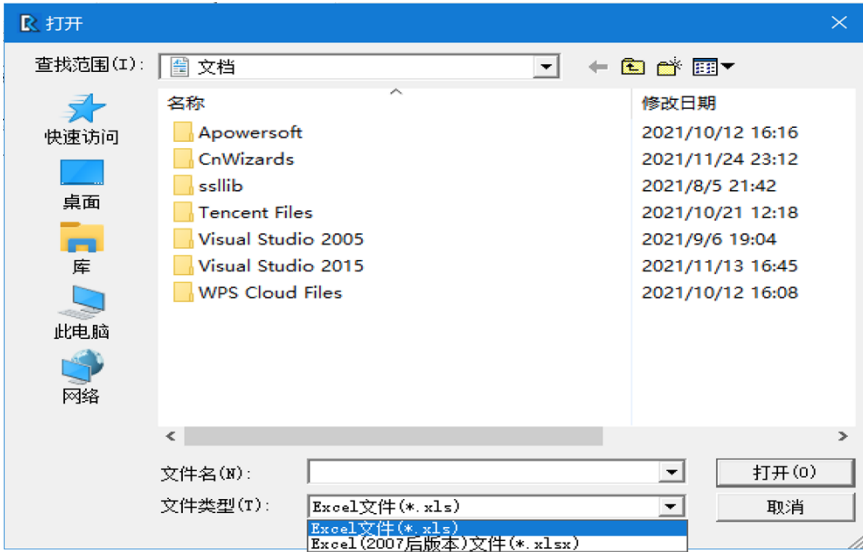

**Fig. 4.** GITEST data file input interface

Figure 5. EXCEL external data file import dialog

2.3. Item analysis reports and Rasch files

Once the data file is imported successfully, or manually input is completed, we could obtain the required data analysis reportsby a single click of mouse. In case the data matrix is much larger, a progressing bar will prompt on the screen while the data is being processed. After it completes, the interface for data analysis report will appear as shown in Figure 6.
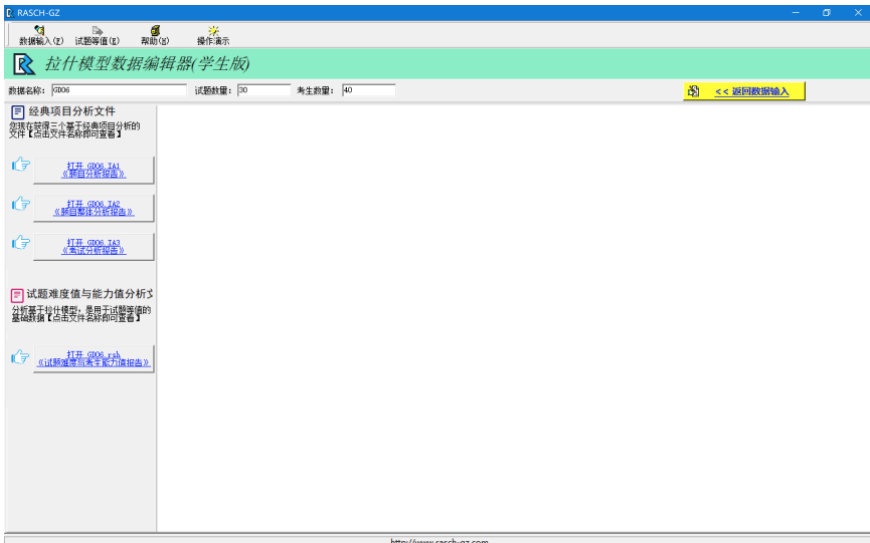
Fig. 6. (RASCH-GZ) interface of data analysis report

The interface description shown in Figure 6 goes as follows:

If you specified the data file name as 'RGZ data 001', there would be four file names with different suffixes in the left column as shown in Figure 6.

(1) RGZ Data 001.**ia1** is "Item Analysis Report for Each test item";

(2) RGZ Data 001.**ia2 is** "TWCT about difficulty-discrimination of all the test items";

(3) RGZ Data 001.**ia3** is "An analysis report about a whole test paper ".These three files are based on the CTT, with the suffixes -.ia1, -.ia2 and -.ia3 respectively.

(4) RGZ data 001.**rsh** is the document containing "test item difficulty and test taker ability" based on Rasch model. In (Rasch-GZ), data files with –rsh ending are indispensable files for conducting test equating.

Sometimes, the user only needs a specific data file according to his own research purpose. In such a case, just select the required data analysis report. See Figure 7 below for details.

Fig. 7. Users' specific data file interface.

In what follows, the first three files: RGZ Data 001.**ia1,** RGZ Data 001.**ia2**, and RGZ Data 001.**ia3** will be briefly interpreted.

RGZ Data 001.ia1. Figure 8 below shows RGZ Data 001.**ia1**wherein item analy sis report for each test item is produced.

Fig. 8. Screenshot of RGZ Data 001.ial

Technically, if we entered 85 test items, (Rasch-GZ) would generate 85 tables for an analysis of each test item. The analysis of the first item is presented in Table 1 below[3].

Table 3. RGZ Data 001.ia1

Te st	Par t	Test D	Part D	Item D	R/ N	A- Scr	B- scr	C- Scr	D- Scr	O- Scr
GD 08	LTN 1	15.6 6	15.5 9	14.7 6	0. 33	11.3 0	12 .16	14 .70	12.7 5	
To tal	Key	A-D	B-D	C-D	D- D	A-N	B- N	C- N	D-N	O- N
10 0	C	– 0.29	– 0.16	0.39	– 0.05	18	23	33	26	0

Based on the above, the authors should say that -ia1 file is very important for professionals to use as a reference to moderate test items after the raw scores obtained from a pre-test conducted.

---

[3]Due to the limited space and avoidance of repetition, it is recommended that readers refer to（2.1）Rasch GZ: The Most Updated Rasch-Based Development in China.

RGZ Data 001.ia2. Figure 9 below shows RGZ Data 001.**ia2** which is a two-way checklist table (TWCT) about difficulty-discrimination of all the test items in a whole test paper.

Fig. 9. Screenshot of RGZ Data 001.ia2

In this TWCT, the horizontal line indicates the P-value (0-1);vertical line shows the discrimination index (between 0.1-1). At the bottom of the table, clear at a glance are the expected quantity, the actual quantity, and the quantity that meets the requirements of P-value of the test items within a test paper respectively in a distribution. This can be summarized in the following three points:

(1) In some cases, though failing to meet the requirements, these calibrated test items are close to the required parameters, and the difficulty is appropriate. Putting the error factors into consideration, the test developers are supposed to retain these calibrated items for tests on small scale;

(2) It is convenient for testing professionals to adjust the item difficulty in a test paper. For example, if easier or more difficult items are observed, which makes the difficulty level of the test paper unbalanced, the solution is to add more difficult items and withdraw easier ones or vice versa. In this way, we could ensure that the overall difficulty of the test paper to meet the requirements, and

(3) The replaced test items (up to requirements) can be properly saved for future use because these items have been well calibrated with good discriminating power.

On the whole, the TWCT generated by Rasch-GZ gives a clear understanding of whether the test design conforms to the requirement the test developers desired. In popular words, the adjustment of overall difficulty level of a test paper is simply selecting calibrated test items with more appropriate parameters to replace those

that fail to meet the requirements. However, for most examination developers, for schools or social training centers, it is infeasible to set up an item bank. This is the ongoing situation across China. Now with Rasch-GZ, things are easy to handle in such a quantitative way.

RGZ Data 001.ia3. RGZ Data 001.ia3 as shown in Table 2 belowis a general analysis report about a whole test paper.

Table 4. An analysis report of a test paper

Test Name:GD06

Test date:2021/09/26

The following report is generated via Rasch-GZ

The average score of this test paper is 9.12, and the correct answer rate is 72.50%, indicating that this is a relatively easy test.

The SD is 3.12, while the expected SD is 12.17. The full range of the highest score and the lowest score is 12, indicating that the distribution of scores is flat. From the whole test paper, the score distribution is skewed to the right. This test (sample) is not reliable, and the standard error of measurement is+2.28.

The factor analysis of the test paper results shows that the whole test paper can explain the knowledge and ability of 20.29% of the examinees

LST and RDG are one factor,

The analysis of the test paper shows that the very easy items account for 54%, the easy items account for 28%, the intermediate items account for 17%, the difficult items account for 1%, and the very difficult items account for 0.00%.

The discrimination of this test paper is -0.54. There are 4 test items that do not meet the requirements.

(The following analysis of each part of the test items with plotting is omitted)

2.4. Other operations

How to open a saved data file. Click "Open" from the "Data Input" menu, and the "Open Data File" window will appear as shown in Figure10 below.

How to open recently used data files. Click "Open the recently used data" from the "Data Input" menu, and the file names of the 5 recently used data files would be displayed in the menu. Click the file name you need to get the data.

How to moderate the property value of the current data. Click "Properties" from the "Data Input" menu and the "Data Properties" window will appear. In Rasch-GZ, modifiable values refer to (1) data name, (2) number of test items, (3) number of test takers, (4) ID length of test takers, and (5) name and number of each section of the test paper. By entering the new corresponding attribute value in the corresponding input box, the user can add or delete operations in the test section. Click the "OK" button to apply the modified properties, click the "Cancel" button to discard the modification.
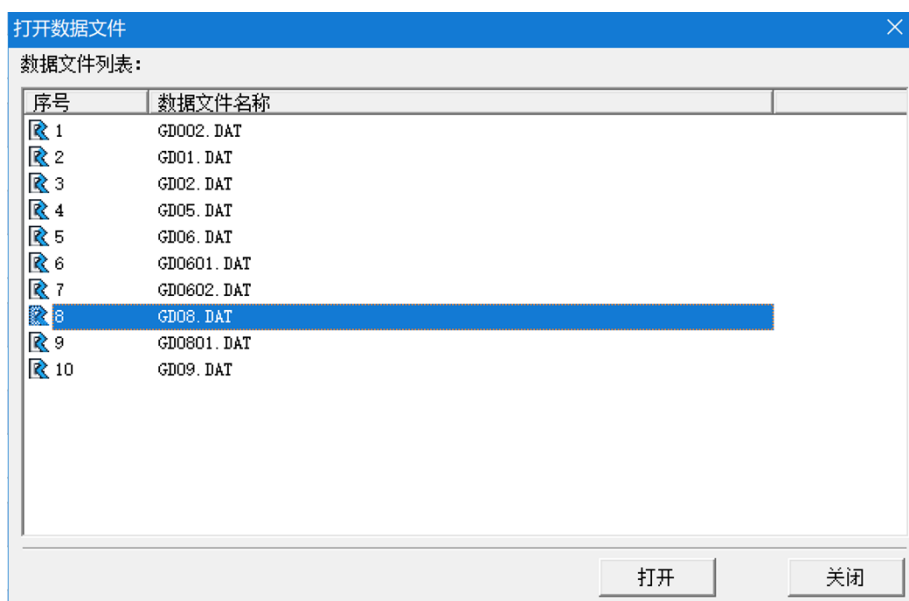
Fig. 10. Rasch-GZInterface for "Open Data File"

Close the current data. Click "Close" from the "Data Input" menu to close the data currently being edited.

Save the current data as. Click "Save As" from the "Data Input" menu, and the "Data Properties" setting window will pop up to modify the new data name. Save the data currently being edited to a new data file.

3. Test Equating[4]

---

[4]This user guider only explains how to conduct test equating. For the theoretical discussion regarding test equating, please refer to (2.2) Rasch GZ: The Most Updated Rasch-Based Development in China.

This section introduces how to use RASCH-GZ to conduct test equating. Prior to it, we need to have simple understanding of the concept, the significance, the conditions and specific methods that test equating should undergo. For the theoretical elaboration, users are suggested to refer to the reference books provided at the end of this paper.

3.1.The Concept of Test Equating
The concept of "test equating" Rasch-GZ developers follow refers to the linking of separate test forms through common items so that scores derived from the tests which were administered separately to different test takers on different occasions, after conversion, will be comparable on the same scale [8], [9], [24], [25], [26], [27], [28], [29].

3.2. The Significance of Test Equating
The significance of test equating is to ensure the fairness regarding the converted scores obtained from different parallel test forms administered to different test takers on different dates. This is a universal issue. In China, large-scale exams with high stakes such as Matriculation English Test (MET), College English Test (CET) are good practices of conducting Rasch-based equating.

3.3. The Data Files required for Test Equating
To conduct test equating using Rasch-GZ, the following data files must be available: At least two data files with –rsh (2.3). One file is set as the basal test, while the other as the one to be equated. The format of these two test papers should be the same, but the number of test items may or may not be exactly the same. A set of items used as the linking items are put in both tests;

3.4. The Way to Conduct Test Equating
The following simple steps will lead you to obtain good equating results:

   Step 1: Click the Equating button in the menu bar. "Rasch Model Test Equating" window pops up. See Figure 11 below.
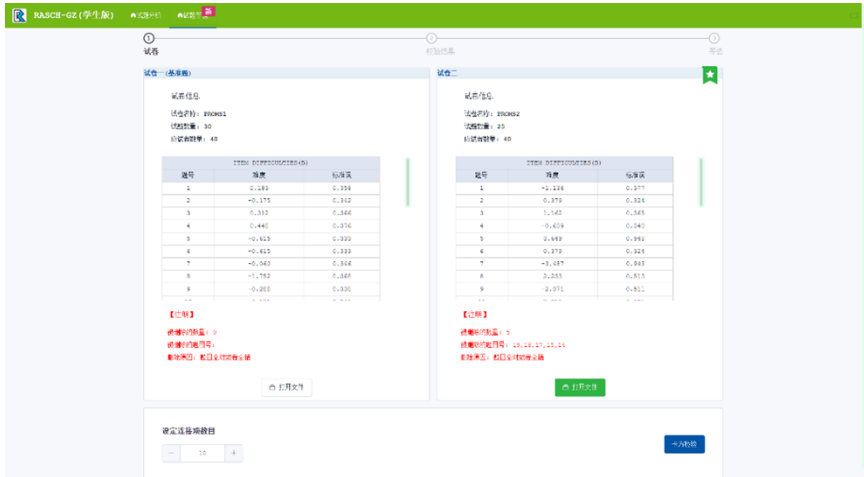
Fig. 11. "Rasch Model Test Equating" interface in (Rasch-GZ)

Step 2: Follow the prompts in the above figure and select the "basal test" button. Rasch—GZ will pop up a list of data files with –rsh suffix. Note that the data files with –rsh ending at this point are the data files containing item difficulties/ability estimates. If no data file with –rsh suffix available, the user is supposed to conduct Rasch analysis first (2.3).

Step 3: Click the "Confirm" button, and the name of the test paper, number of test items, and number of test takers will be displayed in the information box of basal test;

Step 4: Select the information of the data file to be equated, and the system will pop up a list of data files, from which you can select the data file to be equated. Note that both the basal file and the data file to be equated at this point must be the data files with item difficulties /ability estimates ending in "-. rsh", as instructed in Step 2. By now, click the "Confirm" button, and the name of the test paper, the number of items, and the number of test takers will be displayed in the information box for the data file to be equated.

Step 5: Set the number of linking items to be used for equating in the information box. Note that this is a new function thatRasch-GZ offers, i.e. the number of linking items can be flexibly determined by the user themselves. Here users just clicks the "Set the number of Linking Items" button on the toolbar and enter the number of linking items that need to be used for equating.Note that in setting the number of linking items,Rasch GZ has made improvement so that the number of linking items can be adjusted right before test equating, rather than being fixed like the other similar program(s)whereinthe number of linking items is written in the command file.

Step 6: Perform Chi square test on the linking items. Click the "Chi Square Test" button on the toolbar, and the RASCH-GZ system will automatically perform a Chi square test on the linking items set for the equating. Note that Rasch-GZ offers another new automatic screening function, i.e. mark the linking items that do not meet

the conditions as "deleted" and automatically delete the items. Figure 12 shows the Chi Square Test Interface.

Fig. 12. The Interface of Chi Square Test

Step 7: Display the test equating results

Click the "Equating Result" button on the toolbar, and the whole equating process using the qualified linking items (with good Chi square test parameters) will

be done with a single click of mouse. Figure 13 shows the equating results of the two parallel tests forms.

Fig.13. The equating results of the two parallel tests forms

4. Conclusion

This paper presented the whole process regarding both item analysis and test equating using Rasch-GZ. In general, Rasch model can provide good solutions to many problems of objective measurement encountered in social sciences. In specific, the use of Rasch-GZ is feasible for researchers in language testing. The reason is obvious. The raw data of such a discipline are difficult to control and the concept is vague. In this sense, RASCH-GZ can calibrate very well the dichotomous data to generate quantitative parameters, thus greatly promoting the application to Rasch model among Chinese researchers of non-English majors. To conclude, the student version of RASCH-GZ and the user guide offer the easy way for Chinese young researchers to learn how to use Rasch model measurement[5]. Though processing a small data matrix, the student version offers good illustration plus short video examples to enable new comers to get familiar with Rasch-Gz quickly for their study or classroom teaching. If data based on large-scale examinations with high stakes in nature are to be calibrated and test equating is a must, the users are suggested to register online to use the professional version of Rasch-GZ.

List of Abbreviations
COVID-19, *Corona Virus* Disease 2019
CTT, classical testing theory
ETS, Educational Testing Service
MC, multiple choice
SD, standardized deviation
TWCT, two-way checklist table

References

[1] **Linacre**, J. M. (2016). A user's guide to WINSTEPS Rasch-model computer
        programs: program manual 3.92.0. ISBN 0-941938-03-4.

---

[5]The student version of Rasch-GZ can be downloaded, free of charge, from https://www.rasch-gz.com.

[2] **Trevor** G. Bond, Yanzi & Moritz. (2021). Applying the Rasch Model: Fundamental Measurement in the Human Science (4th Ed). Routledge.

[3] **Trevor** G. Bond & Christine M. Fox. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Science (3rd Ed). Routledge.

[4] **Steven** Steiner.(2021)."Better measurement and fewer parameters! The true value of Rasch over IRT", the keynote speech at PROMS2021, Nanjing, China.

[5] **Mok, M. M. C., & Zhang, Q** (Eds.). (2014). Constructing variables. Book of Abstracts Vol. II. Journal of Applied Measurement. ISBN 978-1-934116-10-4. JAM Press. P.O.Box 1283 Maple Grove, MN55311. USA.

[6] **Mok, M. M. C., &Zhang, Q** (Eds.). (2015). Constructing variables. Book of Abstracts Vol. I. Journal of Applied Measurement. ISBN 978-1-934116-11-1. JAM Press. P.O.Box 1283 Maple Grove, MN55311. USA.

[7] **Mok, M. M. C.,& Zhang, Q**. (2018).[Introduction to Rasch Measurement].ISBN 978-1-934116-13-5. JAM Press. P.O.Box 1283 Maple Grove, MN55311. USA.

[8] **Wright**, B. D. & Stone, M. H. (1979). Best test design: Rasch measurement. MESA Press.

[9] **Wright**, B. D. (1992). IRT in the 1990s: Which models work best? *Rasch Measurement Transactions*, 6(1), 196–200.

[10] **Zhang, Q** (2022a). From GITEST to RASCH-GZ: Inheritance and Development of Rasch-based Research in China. http://dx.doi.org/10.20431/2349-0381.09S1001

[12] **Zhang, Q.** (2022b).Rasch Model and Test Equating in China: The most updated development in China. International Education e-Colloquium, KL, Malaysia.

[13] **Zhang, Q.** (2021a). Rasch-GZ, the first Chinese version of Rasch-based item analysis and test equating. Online Symposium. PROMS2021, Nanjing China. Retrieved from http://www.proms.promsociety.org

[14] **Zhang, Q.** (2021b).Rasch Model and Test Equating in China: Rasch Model in Chinese version has come. The 6th International Conference on Language Testing and Assessment. Shanghai, China

[15] **Zhang, Q.** (2019). Rasch Model:Research and Practice in China. In Myint Swe Khine. (Ed.). *International Trends in Educational Assessment*. Brill ǀ Sense. Retrieved from http://catalog.loc.gov.

[16] **Zhang, Q.** (Ed.). (2016). Pacific-Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceeding. Springerdoi: 10.1007/978-981-10-8138-5;

[17] **Zhang, Q.** (Ed.). (2015). Pacific-Rim Objective Measurement Symposium (PROMS) 2015 Conference Proceeding. Springerdoi: 10.1007/978-981-10-1687-5;

[18] **Zhang, Q**& Zhang, T.T.(2014). Rasch Model: Status quo and prospect in China. In Zhang & Yang. (Eds.).Pacific Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceedings Rasch and the Future. (pp.17-24). https://doi.org/10.1007/978-3-662-47490-7

[19] **Zhang, Q**& Yang, H. (Eds.). (2014). Pacific-Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceeding. Springer doi: 10.1007/978-3-662-47490-7

[20] **Zhang, Q**& Yang, H. (Eds.). (2012a). Pacific-Rim Objective Measurement Symposium (PROMS) 2012 Conference Proceeding. Springer doi: 10.1007/978-3-642-37592-7

[21] **Zhang, Q.** (2012). Towards International Practice of Language Testing in China. Keynote speech given at the PROMS2012, Jiaxing, China, August 6-9, 2012.

[22] **Zhang, Q.** (2004). [Item Analysis and Test Equating: Research and Application]. [M].Higher Education Press

[23] **Zhang**, Zhigan. (Ed.). (2021).p.270. [Biography of Outstanding Social Scientists in Guangdong]. Sun-Yat-sen University Press. ISBN 978-7-306-07100-2

[24] **Hambleton**, Swaminathan & Rogers. (1985). *Item Response Theory: Principle and Application*. Academic Publisher.

[25] **Gui, S.C,** Li, W & Zhang, Q. (1993). [The Application of IRT to MET Equating]. In NEEA (Ed.). (pp. 391-393).The 4th Annual Forum on Educational Testing in Beijing, China Peace Press.

[26] **Kolen, M.J.,**& Brennan, R. L. (2004). Test equating, Scaling, and Linking: Methods and Practices. (2nd Ed). Springer Vertag.

[27] **Kolen, M.J.,**& Brennan, R. L. (1995). Test equating: methods and practices. Springer Vertag. New York, Inc.

[28] **Bachman,** L.F. & Adrian Palmer. (2010). Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World. Oxford University Press

[29] **Liu, Yuming. (2020).** *Test Equating, Scaling, and Linking: Methods and Practices* (Kolen, M.J., & Brennan, R. L, Trans.; (3rd Ed). (Original work published 2004)