International Journal of Humanities Social Sciences and Education (IJHSSE)
Volume 9, Special Issue 1, May 2022, PP 25-35
ISSN 2349-0373 (Print) & ISSN 2349-0381 (Online)
https://doi.org/10.20431/2349-0381.09S1004
www.arcjournals.org

RASCH Model and Test Equating in China ---- A Comparison and Contrast of WINSTEPS and GITEST

Wu Jinyu

University of Electronic Science and Technology of China, Zhongshan Institute, China

Corresponding Author:** ***Wu Jinyu, *PhD, University of Electronic Science and Technology of China, Zhongshan Institute, China. Dr. Wu has been actively involved in language testing since her PhD studies in 2015. 12316603@ qq.com*

Abstract: *This paper addresses the basic concept, definition and practice regarding test equating. In China, test equating plays a central role for large-scale examinations with high stakes. It is still held as the prerequisite condition for item banking in computerized as well as in Internet-based testing. For the purpose of comparison and contrast, both Winsteps and GITEST are used to calibrate the same group of data of 40 Chinese students of non-English major collected from parallel tests through linking items ad hoc administered in a university in Guangdong Province, China. The item difficulty parameters thus obtained turned out to be 99.8% correlated. Comparison and contrast of these two types of software are elaborated. The paper concludes that Winsteps and GITEST are equally good for conducting test equating.*

Keywords: *Test equating, Rasch Model, CAT, Winsteps, GITEST, PROMS, RASCH-GZ*

1. Research Background

Over the past decades, especially since the Pacific-Rim Objective Measurement Symposium (PROMS) was first held in Malaysia in 2005, the academic community has paid increasingly attention to the application of Rasch model to the research of objective measurement. PROMS entered China for the first time in 2012. As the PROMS organizers in China, the authors believe that among various kinds of Rasch-based computer software available for test equating, both GITEST and Winsteps are great software programs to consider. They offer a wide range of application of Rasch model to practical testing problems, assumes binary scoring of item responses and gives stable and accurate estimates of item parameters and scale scores for both long and short tests and classroom exercises. They are best representing respectively the current applications based on the Rasch model in China and outside China. This paper attempts to present, among many of their features, the significant aspect of Winsteps: equating for parallel tests based on a group of minimum yet representative data and comparison and contrast with GITEST.

2. Test Equating and Its Significance

Though Winsteps is widely used for objective measurement of various purposes, test equating is seldom applied. Equating is a statistical process that is used to adjust scores on test forms so that scores on the forms can be used interchangeably (Kolen & Brennan, 2004). Many testing programs use multiple forms of the same test. Such as college admission, in which serious decisions are made about people who might have taken the test at different administrations during a year or in different years, the primary reason for having multiple forms of a test is to maintain security and fairness. However, even though test developers attempt to construct test forms that are as similar as possible with each other in content and statistical specifications, the forms typically differ somewhat in difficulty. The comparability of tests scores across different tests measuring the same ability is an issue of considerable importance to test developers, measurement specialists, and test takers alike (Hambleton, Swaminathan & Rogers, 1991). Take the Matriculation English Test (MET) in China for example, which is the most prestigious and competitive examination of high stake administered annually to approximately 10 million candidates in China, and 60% or so of the participants can be enrolled.

Its item difficulties and test security must be put well under control and thus great importance is attached to it. If the same MET paper is administered repeatedly to different candidates nationwide annually to admit students for college studies, or if the same test paper is used repeatedly to different students before annual graduation for achievement evaluation, there is no way yet of protecting test security after its administration. On the other hand, it would not be feasible to administer two separate tests at once to the same group of candidates so as to compare the item difficulties of the tests. In this sense, equating plays a central role.

3. TEST EQUATING AND ITS CONCEPT

The concept of "equating" discussed in the present paper therefore refers to 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 (Hambleton & Swaminathan, 1985, cited in Gui & Li, 1989). The idea is better illustrated as follows:

Group A test takers took Test A, which has L items with n anchor items;

Group B test takers took Test B, which has L items with n anchor items.

This is interpreted in language testing as two parallel test forms being written, each consisting of "n" anchor items and are administered to two different groups of samples drawn from the same population at either the same or different time. What is intended to achieve is to equate the metric of all the L"s items of the two tests and put them on the same scale.(Zhang & Hu, 2000; Zhang, 2004). To accomplish this, we use Test A as the basal test calibration and choose, from this basal test, n items (n<L) as linking items and put these linking items in Test B. The following array shows the idea wherein Item 27 through Item 42 in both tests are used as linking items. Totally, 16 items in each test.

Test A 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 **27 28 29 30 31 32 33 35 36 37 38 39 40 41 42**

Test B **27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42** 43 44 45 46 47 48 49 50 51 52 53 54 57 58 59 60 61 62 63

This is considered as the typical examples in terms of "equating of parallel tests". In today's testing practice, equating plays a central role and is held as the prerequisite condition for Computerized Adaptive Testing (CAT), item banking and for online testing in the forthcoming Intern-based testing as well. Through equating, the changes of item difficulties in the test forms can be observed and equated, and the corresponding ability estimates across different occasions are thus adjusted. As equating is a complicated process requiring enormous data processing, and manual calculation is by no means feasible, Rasch-based computer software like Winsteps and GITEST offers us an effective tool. In what follows, we present a pair of representative yet real data to demonstrate the complete procedure of how equating is complemented by both GITEST and Winsteps. (Zhang & Hu, 2000; Zhang, 2004)

4. EQUATING BY GITEST: A SIMPLIFIED EXAMPLE

4.1. GITEST Program

GITEST is a Rasch-based system first developed by Ph.D program of applied linguistics headed by Prof. Gui Shichun(1986, 1990, 1993) of Guangzhou Institute of Foreign Languages, China as early as in 1980s. It was written in BASIC according to Rasch Model which is good at performing the following functions:

- It assumes binary (right-wrong) scoring;
- Designed for applications of both CTT and Rasch to practical testing problems;
- Maximum likelihood (ML);
- Tests of fit for individual items;
- Analysis of multiple subtests in one pass;
- Item analysis and test paper evaluation and report;
- Feedback for teaching and testing improvement；

- Linking of 2 test forms through common items ;
- 200 items by10,000 candidates (maximum sample size) in a single run;

4.2. GITEST Data Editing

The data editing for GITEST is simple. The rows of data matrix are the test takers‟ ID followed by all the dichotomous responses presented by each test taker, while each column contains one answer to the corresponding test item. GITEST accepts two types of data responses: integer or char. Like all the other Rasch and IRT programs processing data of dichotomous in nature, if integer data are used, '1' represents right answer and „0‟, wrong answer. If char data are input, a line of key answers should be provided and put at the first line of the data matrix as shown in Table 1 and 2 in the following. Though written in Basic, GITEST can process the data metrics up to 10,000 persons by 200 items with a single run. This is the only Rasch-based software ever used to process data for Ten- Year Equating Project of Matriculation English Test (MET) funded by National Education examination Authority (NEEA) under Ministry of Education from 1990-1999 in China. (Gui, Li & Zhang, 1993)

Table1. *GITEST Integer Data Matrix*

TestA0001 1110101111111110101010101001000001111111011

TestA0002 111101010111010101010101001011001111111101

TestA0003 111010111111111010101010100100111111111011

TestA0004 11110101010000010101111100100100111111101

TestA0005 11101011111111101011111010010000011111111011

… … … … … …

TestA0079 0000000101010111110101010010001011111111101

TestA0080 1110111111111110101010101001111111111111011

Table2. *GITEST Char Data Matrix*

Key AABCBDCBDACDBADCBBCBCDABAABAAAACDBADCDAAAD

TestA0001 DCCCBDCBDACAAAACBBCBCDABAABACAAADBADCDAACC

TestA0002AACCDDCBDACDBAACDBCBCDABAABACAACDBADCDADCC

TestA0003ACCCBDCBDACDBAACBBCBCDABAABACAACDBADCDAADC

TestA0004ACCCBDCBDACDBAACBBCBCDABAABACAACDBADCDADCC

TestA0005ACDCBDDBDDCDBAACBDCBCDABDBBACBBBBBADCDABCC

… … … … …

TestA0079ABCCBDCBDACDBAACBBCBCDABAABACAACDBADCBAACC

TestA0080ACCCBDCBDACBBBACCCCBCDABAABACAACDBADCDAACC

4.3. GITEST Equating

With a single run, GITEST can yield the measure of the 16 linking items in both Test A and Test B thus obtained and listed in Table 3 below.

Table3. *GITEST: Linking Item Difficulties in logits of Test A and B*

ITEM	Test A	Test B
0001	0.335	0.055
0002	-0.237	-0.978
0003	-0.073	-0.669
0004	0.154	0.118
0005	-0.018	0.118
0006	0.154	0.736
0007	-0.073	-0.429
0008	-0.237	-0.068

0009	-0.981	-1.174
0010	1.156	1.472
0011	-0.073	-0.852
0012	-0.027778	-0.608
0013	0.462	0.311
0014	0.213	-0.068
0015	-0.449	-0.189
0016	-0.555	-0.669
MEAN	-0.016	-0.181

What we are interested in here is the means of the 16 linking items in the two tests. As observed at the bottom of the table, the two means of the same linking items in Test A and Test B turned out to be different: -0.016 (logits) in Test A and -0.181 (logits) in Test B. The question is then raised: Why did the difficulties of the same 16 items turn out to be different?

Our interpretation is that the test items to which these common items are linked respectively in Test A and B are different. If -0.016 - (-0.181), the difference obtained from the mean minus the mean is 0.165 logit, indicating the test items in Test A are a little bit easier than those in Test B. That is why the means of the 16 linking items in Test A turn out to be more difficult than those in Test B. In other words, test items in Test A are 0.165 easier in logit than those in Test B. "In such an example, the linking items are the hard items in **EASY** test but the easy items in the **HARD** test" (Wright & Stone, 1979; Zhang & Hu, 2000; Zhang, 2004). And the measure of the other items in both Test A and Test B obtained are listed in Table 4 below.

Table4. *GITEST Equated Item Difficulties*

ITEM	Test A	Test B
0017	0.528	0.378
0018	0.273	0.661
0019	0.528	-0.369
0020	0.596	0.896
0021	-0.29	-0.548
0022	0.596	-0.669
0023	-0.237	-0.791
0024	-0.449	0.98
0025	0.667	0.118
0026	-0.073	1.258
0027	-1.445	-0.488
0028	-0.927	-0.309
0029	0.213	-0.309
0030	-0.29	0.055
0031	0.596	0.516
0032	0.596	-0.309
0033	-0.018	-0.488
0034	-0.344	1.068
0035	0.335	0.118
0036	0.154	0.661
0037	-0.555	0.055
0038	-0.073	-0.852
0039	0.895	-0.791
0040	0.096	1.068
0041	-1.092	0.98
0042	0.977	

5. EQUATING BY WINSTEPS: A SIMPLIFIED EXAMPLE

5.1. Winsteps Porgram

Winsteps is a Rasch-based program developed by John M. Linacre in 1984, which constructs Rasch measures from simple rectangular data sets, usually of persons and items. It is good at performing more functions than GIETST does. For example, Winsteps can process up to 9,999,999 persons by 60,000 items with rating scales up to 255 categories for each item. At the same time, Winsteps provides a familiar "pull-down" user interface, intended to provide the user with maximum speed and flexibility. (Linacre, 2016)

5.2. Winsteps Data Matrix

To input data into Winsteps system for equating, we need create specific data matrix. We open an Excel spreadsheet, of which the first row is the variable names, each row is one person (subject, case), and each column contains one variable. Table 5 below briefly shows the idea.

Table5. *Excel spreadsheet used for Winsteps data matrix for Test Equating*

TestA0001 DCCC BDCBDACAAAACBBCBCDABAA BACAAADBADCDAACC
TestA0002 AACCDDCBDACDBAACDBCBCDABAA BACAACDBADCDADCC
TestA0003 ACCCBDCBDACDBAACBBCBCDABAA BACAACDBADCDAADC
TestA0004 ACCCBDCBDACDBAACBBCBCDABAA BACAACDBADCDADCC
TestA0005 ACDCBDDBDDCDBAACBDCBCDABDBBACBBBBBADCDABCC
… … … …
TestA0079 ACCCBDCBDACDBAACBBCBCDABAABACAACDBADCDADCC
TestA0080 ACDCBDDBDDCDBAACBDCBCDABDBBACBBBBBADCDABCC

TestB0001 BACBBBBBADCDABCC ACBDCBAAABCDACBDCBCDCBACBD
TestB0002 BACBBBBBADCDABCC ACBDCBAAABCDACBDCBCDCBACBD
TestB0003 BACBBBBBADCDABCC ACBDCBAAABCDACBDCBCDCBACBD
TestB0004 BACBBBBBADCDABCC ACBDCBAAABCDACBDCBCDCBACBD
TestB0005 BACBBBBBADCDABCC ACBDCBAAABCDACBDCBCDCBACBD
… … … …
TestB0072 BACBBBBBADCDABCC ACBDCBAAABCDACBDCBCDCBACBD
TestB0073 BACBBBBBADCDABCC ACBDCBAAABCDACBDCBCDCBACBD

5.3. Winstep Equating

Test Equating via linking items are straightforward with Winsteps, but do require prudent care. The more thought is put into test construction and data collection, the easier the equating will be. Such test equating proceed by Winsteps goes thus: Winsteps initially analyzes the linking items from the two tests, i.e. Test A and Test B and then analyzes each test separately. In Winsteps, the item parameter values can be anchored using command IAFILE=. Anchoring facilitates equating test forms and building item banks.

With this, a single run of Winsteps would let us obtain the item measures for all the items and construct the scale. In this step, separate analyses for each test were conducted with the 16 quality linking items anchored at the value that had been calibrated in the above step to the general item measures for all of the items. See Table 6 below.

TABLE6. *16 Linking Item STATISTICS: MEASURE ORDER*

2 CATS WINSTEPS 3.92.1

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MODEL MEASURE	S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	EXACT MATCH EXP.	OBS%	EXP%
10	24	153	1.55	.24	1.00	.0	.91	-.3	.35	.33	83.6	85.0
6	46	153	.59	.19	1.05	.6	1.05	.5	.31	.36	71.1	73.0
13	47	153	.56	.19	.94	-.7	.93	-.6	.42	.36	74.3	72.5
1	53	153	.35	.18	.97	-.4	.94	-.6	.40	.36	70.4	70.0
4	55	153	.28	.18	1.14	2.0	1.15	1.5	.21	.36	63.8	69.1
14	57	153	.22	.18	.94	-.9	.90	-1.0	.43	.36	71.1	68.4
5	58	153	.19	.18	1.11	1.6	1.19	2.0	.23	.36	65.1	68.1
8	65	153	-.03	.18	1.10	1.6	1.14	1.7	.25	.36	61.2	65.9
7	68	153	-.13	.17	1.05	.9	1.05	.7	.31	.36	60.5	65.5
15	71	153	-.22	.17	.84	-3.0	.79	-3.0	.53	.36	73.0	65.0
3	72	153	-.25	.17	1.08	1.4	1.08	1.0	.28	.36	64.5	64.9
11	75	153	-.34	.17	.95	-.9	.99	-.2	.40	.36	69.7	64.7
2	80	153	-.49	.17	.91	-1.6	.88	-1.6	.45	.35	66.4	64.5
16	81	153	-.52	.17	.91	-1.6	.87	-1.7	.45	.35	63.2	64.5
12	89	153	-.76	.18	.94	-1.0	.90	-1.2	.41	.34	67.8	65.6
9	97	153	-1.01	.18	1.09	1.4	1.13	1.3	.23	.33	61.8	67.8
MEAN	64.9	153.0	.00	.18	1.00	.0	.99	-.1			68.0	68.4

where the item difficulties were tailed in the order of decrease with Item 10, the hardest one (1.55 logits), for which merely 24 out of 135 test takers got the correct answer, and Item 9, the easiest one (-1.01 logits), for which 97 out of 135 got the correct answer, where TOTAL COUNT indicates totally 153 test takers from two groups taking respectively Test A and Test B tried these 16 linking items, and both INFIT and OUTFIT of the items were accepted. Table 7 below shows the linking item difficulties in logits of both Test A and B produced by Winsteps, indicating high correlationship with those of GITEST.

Similar to the previous steps, the first round of the analysis was undertaken to identify the under fit persons whose OUTFIT or INFIT MNSQ were larger than 2.0, and the second round of the analysis, without the under fit persons identified in the first round of the analysis, was used to calibrate the difficulty estimates for all of the items. In Winsteps, any items showing misfit to the Rasch model, i.e., the OUTFIT or INFIT MNSQ was larger than 2.0, were removed from the scale. No items were identified by this criterion and removed. Furthermore, any items with extremely high or low difficulty were investigated by experts specialized in English to determine whether they were appropriate for inclusion in the assessment. Consequently, no items were removed because their difficulties were all appropriate for the corresponding grades of the sampled test takers. The remaining items comprised the item pool of the two tests. (Linacre, 2016) The item measures for Test A and B of both pre-and post-equating are presented in the following Table 8, 9, 10 and 11 respectively.

Table7. *WINSTEPS: Linking Item Difficulties in logit of Test A and Test B*

ITEM	Test A	Test B
0001	0.33	0.06
0002	-0.24	-0.98
0003	-0.1	-0.7
0004	0.15	0.12
0005	-0.02	0.12
0006	0.15	0.74
0007	-0.07	-0.43
0008	-0.24	-0.07
0009	-0.98	-1.17
0010*	0.31	1.48
0011	-0.07	-0.85
0012	-1.04	-0.61
0013	0.46	0.31
0014	0.21	-0.07
0015	-0.45	-0.19
0016	-0.55	-0.67
MEAN	-0.13	-0.18
CORR with GITEST	0.8	1

Where all the values are observed highly correlated with those yielded via GITEST EXECPT Item 10 which in some way obviously affected the correlation.

Table8. *The Item Measures for Test A (Pre-Equating)*

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT	MATCH	Item
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	21	80	.53	.26	.96	-.3	.95	-.3	.28	.20	73.8	73.7	Q1
2	25	80	.27	.25	.98	-.2	.95	-.4	.27	.21	66.3	69.0	Q2
3	21	80	.53	.26	1.04	.4	1.03	.3	.12	.20	73.8	73.7	Q3
4	20	80	.60	.26	.99	.0	.98	-.1	.22	.20	75.0	75.0	Q4
5	35	80	-.29	.23	.95	-.8	.96	-.6	.31	.22	65.0	60.4	Q5
6	20	80	.60	.26	1.05	.5	1.07	.5	.08	.20	75.0	75.0	Q6
7	34	80	-.24	.23	1.04	.7	1.03	.5	.15	.22	58.8	61.0	Q7
8	38	80	-.45	.23	.99	-.1	.99	-.1	.24	.22	56.3	59.3	Q8
9	19	80	.67	.27	1.05	.4	1.08	.6	.08	.19	76.3	76.2	Q9
10	31	80	-.07	.24	1.06	.8	1.06	.7	.10	.22	58.8	63.2	Q10

11	56	80	-1.44	.25	.97	-.3	.95	-.3	.28	.21	72.5	70.2	Q11
12	47	80	-.93	.23	.99	-.1	1.02	.2	.22	.22	67.5	61.7	Q12
13	26	80	.21	.24	1.13	1.4	1.16	1.5	-.06	.21	62.5	67.9	Q13
14	35	80	-.29	.23	.98	-.3	.97	-.4	.27	.22	57.5	60.4	Q14
15	20	80	.60	.26	1.00	.1	1.01	.1	.18	.20	75.0	75.0	Q15
16	20	80	.60	.26	1.00	.0	1.00	.0	.20	.20	75.0	75.0	Q16
17	30	80	-.02	.24	.99	-.1	.99	-.1	.24	.22	62.5	64.1	Q17
18	36	80	-.34	.23	.97	-.5	.96	-.6	.29	.22	61.3	59.9	Q18
19	24	80	.33	.25	.95	-.5	.93	-.6	.33	.21	67.5	70.1	Q19
20	27	80	.15	.24	.98	-.2	.98	-.2	.25	.21	67.5	66.9	Q20
21	40	80	-.55	.23	.97	-.6	.96	-.7	.29	.23	57.5	59.1	Q21
22	31	80	-.07	.24	1.00	.0	1.02	.3	.21	.22	66.3	63.2	Q22
23	16	80	.89	.28	.98	-.1	.92	-.4	.25	.18	80.0	80.0	Q23
24	28	80	.10	.24	1.02	.3	1.03	.3	.17	.22	61.3	65.9	Q24
25	50	80	-1.09	.24	1.01	.1	.99	-.1	.21	.22	58.8	64.0	Q25
26	15	80	.98	.29	.97	-.1	.90	-.4	.27	.18	81.3	81.2	Q26
27	24	80	.33	.25	1.00	.0	.99	.0	.21	.21	70.0	70.1	Q27
28	34	80	-.24	.23	.98	-.3	.99	-.1	.26	.22	63.8	61.0	Q28
29	31	80	-.07	.24	1.06	.8	1.05	.6	.11	.22	61.3	63.2	Q29
30	27	80	.15	.24	1.07	.8	1.10	1.0	.06	.21	67.5	66.9	Q30
31	30	80	-.02	.24	1.07	1.0	1.09	1.1	.06	.22	60.0	64.1	Q31
32	27	80	.15	.24	1.08	1.0	1.07	.7	.05	.21	62.5	66.9	Q32
33	31	80	-.07	.24	.93	-1.0	.92	-1.0	.36	.22	66.3	63.2	Q33
34	34	80	-.24	.23	1.08	1.4	1.12	1.7	.04	.22	56.3	61.0	Q34
35	48	80	-.98	.23	1.03	.4	1.03	.5	.16	.22	58.8	62.4	Q35
36	13	80	1.16	.31	.96	-.1	.92	-.2	.26	.17	83.8	83.7	Q36
37	31	80	-.07	.24	.94	-.9	.95	-.5	.34	.22	71.3	63.2	Q37
38	49	80	-1.04	.24	.97	-.4	.97	-.4	.28	.22	70.0	63.2	Q38
39	22	80	.46	.26	.95	-.4	.92	-.5	.31	.20	71.3	72.5	Q39
40	26	80	.21	.24	.98	-.2	.97	-.2	.25	.21	70.0	67.9	Q40
41	38	80	-.45	.23	.94	-1.1	.94	-1.0	.34	.22	63.8	59.3	Q41
42	40	80	-.55	.23	.94	-1.3	.92	-1.5	.37	.23	57.5	59.1	Q42
MEAN	30.2	80.0	.00	.25	1.00	.0	1.00	.0			66.8	67.1	
P.SD	9.9	.0	.57	.02	0.05	.6	.06	.6			7.2	6.5	

Table9. *The Item Measures for Test B (Pre-Equating)*

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT	MATCH	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item
1	29	73	.06	.25	1.03	.4	1.04	.4	.24	.29	57.5	64.1	Q27
2	46	73	-.98	.25	.84	-1.9	.81	-1.7	.51	.27	69.9	65.7	Q28
3	41	73	-.67	.25	.95	-.7	.93	-.7	.36	.28	69.9	63.1	Q29
4	28	73	.12	.25	1.13	1.5	1.15	1.4	.10	.29	58.9	64.8	Q30
5	28	73	.12	.25	1.09	1.0	1.10	1.0	.16	.29	64.4	64.8	Q31
6	19	73	.74	.28	.97	-.1	.93	-.3	.33	.28	74.0	74.9	Q32
7	37	73	-.43	.24	.98	-.3	.96	-.4	.32	.29	58.9	62.0	Q33
8	31	73	-.07	.25	1.05	.8	1.07	.8	.20	.29	57.5	62.9	Q34
9	49	73	-1.17	.26	.90	-1.0	.88	-.8	.41	.26	71.2	68.6	Q35
10	11	73	1.48	.34	.96	-.1	.97	.0	.29	.25	86.3	85.5	Q36
11	44	73	-.85	.25	.92	-1.0	.89	-1.0	.40	.28	67.1	64.4	Q37
12	40	73	-.61	.25	.93	-1.1	.90	-1.1	.40	.28	65.8	62.7	Q38
13	25	73	.31	.26	.98	-.2	.98	-.1	.31	.29	68.5	67.8	Q39
14	31	73	-.07	.25	.97	-.4	.97	-.3	.33	.29	71.2	62.9	Q40
15	33	73	-.19	.25	.89	-1.7	.86	-1.6	.46	.29	65.8	62.2	Q41
16	41	73	-.67	.25	.88	-1.7	.86	-1.5	.46	.28	72.6	63.1	Q42
17	24	73	.38	.26	.88	-1.2	.83	-1.4	.47	.28	71.2	68.9	Q43
18	20	73	.66	.27	1.07	.6	1.11	.7	.17	.28	74.0	73.7	Q44
19	36	73	-.37	.24	1.14	2.1	1.14	1.6	.09	.29	49.3	61.9	Q45
20	17	73	.90	.29	1.07	.5	1.06	.4	.17	.27	78.1	77.6	Q46
21	39	73	-.55	.25	1.11	1.6	1.13	1.4	.12	.28	58.9	62.4	Q47

22	41	73	-.67	.25	1.11	1.6	1.12	1.2	.12	.28	56.2	63.1	Q48
23	43	73	-.79	.25	1.05	.7	1.34	2.9	.14	.28	64.4	63.9	Q49
24	16	73	.98	.29	1.05	.4	.98	.0	.22	.27	76.7	78.9	Q50
25	28	73	.12	.25	.92	-.9	.88	-1.1	.41	.29	64.4	64.8	Q51
26	13	73	1.26	.32	1.14	.7	1.12	.5	.08	.26	80.8	82.9	Q52
27	38	73	-.49	.24	1.02	.4	1.01	.2	.25	.29	60.3	62.2	Q53
28	35	73	-.31	.24	1.00	.1	.99	-.1	.29	.29	61.6	61.9	Q54
29	35	73	-.31	.24	.92	-1.2	.89	-1.2	.41	.29	64.4	61.9	Q55
30	29	73	.06	.25	1.02	.3	1.02	.2	.26	.29	60.3	64.1	Q56
31	22	73	.52	.27	.91	-.8	.87	-.8	.42	.28	75.3	71.3	Q57
32	35	73	-.31	.24	.95	-.8	.92	-.9	.37	.29	67.1	61.9	Q58
33	38	73	-.49	.24	1.01	.1	.99	-.1	.28	.29	54.8	62.2	Q59
34	15	73	1.07	.30	.89	-.6	.88	-.5	.42	.27	80.8	80.2	Q60
35	28	73	.12	.25	1.03	.4	1.07	.7	.23	.29	61.6	64.8	Q61
36	20	73	.66	.27	1.13	1.0	1.20	1.2	.07	.28	68.5	73.7	Q62
37	29	73	.06	.25	1.07	.9	1.09	.9	.18	.29	60.3	64.1	Q63
38	44	73	-.85	.25	1.05	.7	1.06	.6	.20	.28	58.9	64.4	Q64
39	43	73	-.79	.25	.96	-.5	.93	-.6	.35	.28	67.1	63.9	Q65
40	15	73	1.07	.30	1.04	.3	1.25	1.1	.14	.27	80.8	80.2	Q66
41	16	73	.98	.29	.96	-.2	.91	-.4	.34	.27	79.5	78.9	Q67
MEAN	30.5	73.0	.00	.26	1.00	.0	1.00	.0			67.2	67.6	
P.SD	10.2	.0	.68	.02	.08	1.0	.12	1.0			8.3	6.8	

Table10. *The Item Measures for Test A (Post-Equating)*

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-AL CORR.	EXP%	EXACT OBS%	MATCH EXP%	DISPLACE	Item
36	13	80	1.55A	.34	1.21	.9	1.17	.7	.26	.15	83.8	87.6	-.32	Q36
26	15	80	1.05	.29	.97	-.1	.90	-.4	.27	.18	81.3	81.3	.00	Q26
23	16	80	.97	.28	.98	-.1	.92	-.4	.25	.18	80.0	80.0	.00	Q23
9	19	80	.74	.27	1.05	.4	1.09	.6	.08	.19	76.3	76.3	.00	Q9
4	20	80	.67	.26	.99	.0	.98	.0	.22	.20	75.0	75.0	.00	Q4
6	20	80	.67	.26	1.06	.5	1.07	.5	.08	.20	75.0	75.0	.00	Q6
15	20	80	.67	.26	1.01	.1	1.01	.1	.18	.20	75.0	75.0	.00	Q15
16	20	80	.67	.26	1.00	.1	1.00	.0	.20	.20	75.0	75.0	.00	Q16
1	21	80	.60	.26	.96	-.2	.95	-.3	.28	.20	75.0	73.8	.00	Q1
3	21	80	.60	.26	1.04	.4	1.04	.3	.12	.20	72.5	73.8	.00	Q3
32	27	80	.59A	.26	1.26	2.0	1.27	1.8	.05	.20	65.0	73.6	-.36	Q32
39	22	80	.56A	.26	.97	-.3	.93	-.4	.31	.20	71.3	73.0	-.03	Q39
19	24	80	.41	.25	.95	-.5	.93	-.6	.33	.21	67.5	70.2	.00	Q19
27	24	80	.35A	.25	.98	-.2	.97	-.2	.21	.21	72.5	69.2	.06	Q27
2	25	80	.35	.25	.98	-.2	.95	-.4	.27	.21	66.3	69.1	.00	Q2
13	26	80	.29	.24	1.13	1.4	1.16	1.5	-.06	.21	62.5	68.0	.00	Q13
30	27	80	.28A	.24	1.08	.9	1.12	1.1	.06	.21	67.5	67.9	-.05	Q30
20	27	80	.23	.24	.98	-.2	.98	-.2	.25	.21	67.5	67.0	.00	Q20
40	26	80	.22A	.24	.96	-.4	.96	-.4	.25	.21	71.3	66.9	.06	Q40
31	30	80	.19A	.24	1.11	1.3	1.15	1.5	.06	.22	61.3	66.4	-.14	Q31
24	28	80	.17	.24	1.02	.3	1.03	.3	.17	.22	61.3	66.0	.00	Q24
17	30	80	.06	.24	.99	-.1	.99	-.1	.24	.22	62.5	64.2	.00	Q17
10	31	80	.00	.24	1.06	.9	1.06	.7	.10	.22	58.8	63.3	.00	Q10
22	31	80	.00	.24	1.00	.0	1.02	.3	.21	.22	66.3	63.3	.00	Q22
34	34	80	-.03A	.23	1.11	1.6	1.16	2.0	.04	.22	58.8	62.8	-.14	Q34
33	31	80	-.13A	.23	.91	-1.4	.91	-1.4	.36	.22	67.5	61.4	.13	Q33
7	34	80	-.16	.23	1.04	.7	1.03	.5	.15	.22	58.8	61.1	.00	Q7
5	35	80	-.22	.23	.95	-.8	.96	-.6	.32	.22	65.0	60.5	.00	Q5
14	35	80	-.22	.23	.98	-.3	.98	-.4	.27	.22	57.5	60.5	.00	Q14
41	38	80	-.22A	.23	.96	-.7	.96	-.6	.34	.22	58.8	60.5	-.16	Q41
29	31	80	-.25A	.23	1.03	.6	1.02	.4	.11	.23	55.0	60.2	.25	Q29
18	36	80	-.27	.23	.97	-.5	.97	-.6	.29	.23	61.3	60.0	.00	Q18
37	31	80	-.34A	.23	.92	-1.5	.93	-1.3	.34	.23	65.0	59.4	.34	Q37
8	38	80	-.38	.23	.99	-.1	.99	-.1	.24	.23	56.3	59.3	.00	Q8
21	40	80	-.48	.23	.97	-.6	.96	-.7	.29	.23	57.5	59.2	.00	Q21
28	34	80	-.49A	.23	.99	-.3	.99	-.1	.26	.23	62.5	59.2	.32	Q28
42	40	80	-.52A	.23	.94	-1.3	.92	-1.4	.37	.23	57.5	59.2	.04	Q42
38	49	80	-.76A	.23	.95	-.9	.95	-.8	.28	.22	61.3	60.7	-.21	Q38
12	47	80	-.85	.23	.99	-.1	1.02	.3	.22	.22	67.5	61.7	.00	Q12

35	48	80	-1.01A	.24	1.05	.7	1.06	.7	.16	.22	58.8	63.9	.10	Q35
25	50	80	-1.02	.24	1.01	.1	.99	-.1	.21	.22	58.8	64.0	.00	Q25
11	56	80	-1.37	.25	.97	-.3	.95	-.3	.28	.21	72.5	70.2	.00	Q11
MEAN	30.2	80.0	.07	.25	1.01	.0	1.01	.0			64.4	67.2	.00	
P.SD	939	.0	.60	.02	.07	.8	.08	.8			7.4	7.0	.12	

Table11. *The Item Measures for Test B (Post-Equating)*

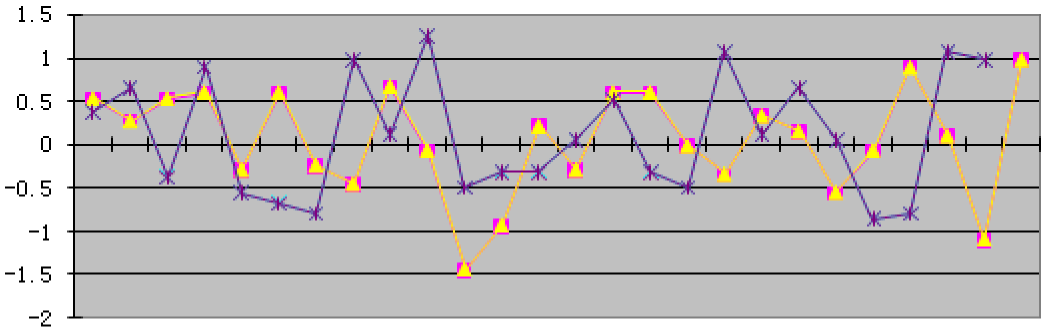
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT	MATCH	DISPLACE	Item
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP%	OBS%	EXP%		
10	11	73	1.55A	.33	.90	-.4	.90	-.3	.29	.25	86.3	84.2	.11	Q36
26	13	73	1.44	.32	1.13	.7	1.11	.5	.08	.26	80.8	82.8	.00	Q52
34	15	73	1.25	.30	.88	-.6	.88	-.5	.42	.26	80.8	80.2	.00	Q60
40	15	73	1.25	.30	1.04	.3	1.24	1.1	.14	.26	80.8	80.2	.00	Q66
24	16	73	1.16	.29	1.05	.4	.97	-.1	.22	.27	76.7	78.9	.00	Q50
41	16	73	1.16	.29	.96	-.2	.91	-.4	.34	.27	79.5	78.9	.00	Q67
20	17	73	1.08	.29	1.07	.5	1.06	.4	.17	.27	78.1	77.5	.00	Q46
18	20	73	.84	.27	1.07	.5	1.11	.7	.17	.28	74.0	73.7	.00	Q44
36	20	73	.84	.27	1.13	1.0	1.19	1.2	.07	.28	68.5	73.7	.00	Q62
31	22	73	.70	.27	.91	-.8	.87	-.9	.42	.28	75.3	71.2	.00	Q57
6	19	73	.59A	.26	.88	-1.2	.84	-1.2	.33	.28	75.3	69.4	.33	Q32
13	25	73	.56A	.26	1.00	.0	1.01	.1	.31	.28	69.9	68.9	-.07	Q39
17	24	73	.56	.26	.88	-1.2	.83	-1.4	.47	.28	71.2	68.9	.00	Q43
1	29	73	.35A	.25	1.05	.7	1.07	.6	.24	.29	58.9	65.4	-.11	Q27
25	28	73	.30	.25	.92	-.9	.88	-1.1	.41	.29	64.4	64.8	.00	Q51
35	28	73	.30	.25	1.03	.4	1.07	.7	.23	.29	61.6	64.8	.00	Q61
4	28	73	.28A	.25	1.12	1.5	1.14	1.4	.10	.29	58.9	64.6	.02	Q30
30	29	73	.24	.25	1.02	.3	1.02	.2	.26	.29	60.3	64.0	.00	Q56
37	29	73	.24	.25	1.07	.9	1.09	.9	.18	.29	60.3	64.0	.00	Q63
14	31	73	.22A	.25	.99	-.1	.99	-.1	.33	.29	68.5	63.9	-.10	Q40
5	28	73	.19A	.25	1.07	.9	1.08	.8	.16	.29	67.1	63.6	.11	Q31
8	31	73	-.03A	.25	1.05	.7	1.05	.6	.20	.29	57.5	62.0	.14	Q34
28	35	73	-.13	.24	1.00	.0	.98	-.1	.29	.29	61.6	61.9	.00	Q54
29	35	73	-.13	.24	.92	-1.2	.89	-1.2	.41	.29	64.4	61.9	.00	Q55
32	35	73	-.13	.24	.95	-.8	.92	-.9	.37	.29	67.1	61.9	.00	Q58
7	37	73	-.13A	.24	.98	-.2	.96	-.4	.32	.29	61.6	61.9	-.12	Q33
19	36	73	-.19	.24	1.13	2.0	1.14	1.6	.09	.29	49.3	61.8	.00	Q45
15	33	73	-.22A	.24	.89	-1.8	.86	-1.6	.46	.29	67.1	61.8	.21	Q41
3	41	73	-.25A	.24	.95	-.8	.93	-.8	.36	.28	64.4	61.9	-.24	Q29
27	38	73	-.31	.24	1.02	.4	1.01	.2	.25	.28	60.3	62.1	.00	Q53
33	38	73	-.31	.24	1.01	.1	.99	-.1	.28	.28	54.8	62.1	.00	Q59
11	44	73	-.34A	.24	.91	-1.4	.90	-1.2	.39	.28	68.5	62.2	-.33	Q37
21	39	73	-.37	.24	1.11	1.6	1.13	1.4	.12	.28	58.9	62.4	.00	Q47
22	41	73	-.49	.25	1.11	1.6	1.12	1.2	.12	.28	56.2	63.0	.00	Q48
2	46	73	-.49A	.25	.82	-2.8	.79	-2.3	.51	.28	79.5	63.0	-.31	Q28
16	41	73	-.52A	.25	.89	-1.7	.86	-1.4	.46	.28	72.6	63.2	.03	Q42
23	43	73	-.61	.25	1.05	.7	1.33	2.9	.14	.28	64.4	63.8	.00	Q49
39	43	73	-.61	.25	.96	-.6	.93	-.6	.35	.28	67.1	63.8	.00	Q65
38	44	73	-.67	.25	1.05	.7	1.06	.6	.20	.27	58.9	64.3	.00	Q64
12	40	73	-.76A	.25	1.00	.0	.98	-.1	.40	.27	64.4	65.3	.33	Q38
9	49	73	-1.01A	.26	.91	-.9	.88	-.8	.41	.26	71.2	68.8	.02	Q35
MEAN	30.5	73.0	.18	.26	1.00	.0	1.00	.0			67.5	67.4	.00	
P.SD	10.2	.0	.66	.02	.08	1.0	.12	1.0			8.4	6.7	.12	

In Table 10 and Table 11 above, items asterisked with „A" indicating they are Anchored Items, i.e. used as linking items between the two tests; therefore, all the test items whose difficulties are rescaled in the similar fashion as discussed in 3.3 above and are comparable on the same scale. This shows us that the equating results obtained from GITEST and Winsteps are the same: of the two tests, Test A is easier as can be observed in Table 12 below. And a careful examination of the parameters obtained further reinforces the assumption proposed by Wright & Stone (1979), i.e. "the linking items are the hard items in EASY test but the easy items in the HARD test". This also shows us that these two types of software are much of the same in terms of equating and are genuinely Rasch-based.

Table12. *Comparison of Equated Test Items Produced by GITEST and Winsteps.*

Item	TestA-GITEST	TestA - Winsteps	TestB -GITEST	Test B- Winsteps
0017	0.528	0.06	0.378	0.38
0018	0.273	0.27	0.661	0.66
0019	0.528	0.53	-0.369	-0.37
0020	0.596	0.6	0.896	0.9
0021	-0.29	-0.29	-0.548	-0.55
0022	0.596	0.6	-0.669	-0.67
0023	-0.237	0.97	-0.791	-0.79
0024	-0.449	-0.45	0.98	0.98
0025	0.667	0.67	0.118	0.12
0026	-0.073	1.05	1.258	1.26
0027	-1.445	-1.44	-0.488	-0.49
0028	-0.927	-0.93	-0.309	-0.31
0029	0.213	0.21	-0.309	-0.31
0030	-0.29	-0.29	0.055	0.06
0031	0.596	0.6	0.516	0.52
0032	0.596	0.6	-0.309	-0.31
0033	-0.018	-0.02	-0.488	-0.49
0034	-0.344	-0.34	1.068	1.07
0035	0.335	0.33	0.118	0.12
0036	0.154	0.155	0.661	0.66
0037	-0.555	-0.55	0.055	0.06
0038	-0.073	-0.07	-0.852	-0.85
0039	0.895	0.89	-0.791	-0.79
0040	0.096	0.1	1.068	1.07
0041	-1.092	-1.09	0.98	0.98
0042	0.98	0.977		

Corr:
TEST A: 0.9855
TEST B: 0.9999



FigureI. *Item Difficulties of both Test A and Test B obtained from GITEST and Winsteps*

6. Conclusion and Development

From the above analyses and discussion, we could come to the conclusion that Winsteps and GITEST are different but alike and their properties can be summarized as what follows:

At the first glance, Winsteps and GITEST seem so different because data matrix for GITEST to process is simply a smaller text file, while for Winsteps, an Excel worksheet doc is needed and the data matrix that can be processed is much bigger. What's more, WINSTEPS can perform more statistical analyses and plotting. In contrast, GITEST handles classic test analyses and Rasch only. If Winsteps is international and paid to use it, GITEST is local but free. Yet, these two types of software, apparently different in some way, possess an affinity with each other. They are both Rasch-based and work well for test equating via anchored or linking items. Both are capable of reporting error messages. On the whole, GITEST and WINSTEPS: each has its own merits and one cannot be replaced by the other. Their utility largely depends on the user's need and purpose.

It is because of these reasons that efforts have been ad hoc made during this COVID-19 pandemic period since 2019 to have successfully updated GITEST which can run online, process sample size of unlimited number of items by unlimited number of subjects, produce desired plotting, testing reports and provide online technical support as WINSTEPS does.

REFERENCES

Benjamin, D. Wright & Mark H. Stone. (1979). Best test design Rasch measurement. MESA.

Gui, S.C. Li, W. & Zhang,Q. (1993). The Application of IRT to Test Equating for MET in China. PP.39`1-393. In NEEA. (Ed.), the 4th China National Education Test Conference Proceeding, China Peace Press. ISBN 7-80101-089-2/G.64

Gui,S.C.(1986). Standardized examination: theory, practice and method. Guangdong Higher Education Press.

Hambleton, R., Swaminathan, H., Rogers, H. (1991). Fundamentals of item response theory. Sage Publications, Inc.

Kolen, M. J., & Brennan, R.L. (2004). Test equating, scaling and linking. Springer Verlag New York, Inc. ISBN: 9780387400860

Linacre, J.M. (2016). A user’s guide to WINSTEPS Rasch-model computer programs: Program manual 3.92.0. ISBN 0-941938-03-4.

Mok, M.M.C & Zhang, Q. (2018).Introduction to Rasch measurement [Translation]. JAM Press. ISBN 978-1-934116-13-5

Mok, M.M.C & Zhang, Q. (2015). Constructing Variables. Book of Abstracts Vol.I. Journal of Applied Measurement. ISBN 978-1-934116-11-1

Mok, M.M.C. & Zhang, Q. (2014). Constructing Variables. Book of Abstracts Vol.II. Journal of Applied Measurement. ISBN 978-934116-10-4

Zhang, Q. et al (2014). A Rasch-based approach for comparison of English listening comprehension between CET and GEPT. PP.115-130. In Zhang, Q. & Yang, H. (Eds.). Pacific Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceedings. Springer. ISBN 978-3-662-47489-1.

doi: 10.1007/978-3-662-47490-7.

Zhang, Q. (2004a). BILOG and PARSCALE: different but alike. In Banerjee, J & Wall, D. (Eds.). Language Testing Update (LTU) Issue 35- Spring 2004. Centre for Research in Language Education. Lancaster University Lancaster, England. International Language Testing Association Official Newsletter.

Zhang, Q. (2004b). Item analysis and test equating for language testing: research and practice. Higher Education Press.

Zhang, Q & Hu, X.X. (2000). PC-BILOG and its Application in China: Equating and Item-banking. In Jose Lai & Tam, Pauline Po-yee. (Eds.) Crosslink in English Language Teaching. Volume 1, 2000. English Language Teaching Unit, the Chinese University of Hong Kong. ISSN 1605-7511

Citation: *Wu Jinyu. "RASCH Model and Test Equating in China ---- A Comparison and Contrast of WINSTEPS and GITEST" International Journal of Humanities Social Sciences and Education (IJHSSE), vol 9, no. s1, 2022, pp. 25-35. DOI: https://doi.org/10.20431/2349- 0381.09S1004.*