PLISKA
STUDIA MATHEMATICA

ПЛИСКА
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. **24** (2015), 131–138

PLISKA
STUDIA MATHEMATICA

EQUATING TEST SCORES BASED ON THEIR IRT CALIBRATION*

D. V. Atanasov, D. M. Dimitrov

Common problem arising in the everyday practice of ability evaluation using tests is how one can compare or equate the scores obtained on different tests or different forms of the same test. Under the main assumption that these tests are based on the same unidimensional latent trait, their scores can be compared using the IRT calibration of the items in both tests. In this paper a procedure for test score equating, based on the sequence of tests with common items in each test are considered.

1. Introduction

In everyday assessment practice one of the problems which arises is how different test scores, obtained from different people on different tests, can be compared. For example if a person has a test score T_1 on a test X_1, what would be the test score of the same person, obtained on a diffent test X_2? There exists a number of different approaches, focused on this problem. A classical linear and equipercentile score equating methods can be found in [3]. Score equating, based on the IRT methodology, is presented in [4]. An overview of some topics in this field can be found in [1]. Our approach is based on the existence of a set of common items between the tests. Both tests should be calibrated under the Item

*The Study is partially supported by National Center for Assessment in Higher Education of the Kingdom of Saudi Arabia (KSA)

2010 *Mathematics Subject Classification:* 91E10, 91E45.

Key words: key words IRT, score equating.

Response Theory (IRT) model. This means that for any item i in the tests there exists a set (a_i, b_i, c_i) of known parameters, representing the probability of correct response $(X = 1)$;

$$P_i(X = 1 \mid a_i, b_i, c_i, \theta) = c_i + (1 - c_i)\frac{1}{1 + \exp(Da_i(\theta - b_i))}. \tag{1}$$

Here the value θ represents the person's ability and D is a constant which can be arbitrarily set, but usually it is set to $D = 1.702$, so that P_i fits the normal ogive curve. A detailed description of the IRT methodology used in this paper can be found in [2], [5] and [6].

2. Equating scores of two tests

2.1. Transformation

The first task to be solved is to place the IRT parameters of the items on a common scale. This can be achieved by a simple transformation, based on a set of common items between the tests.

Let us have two tests X_1 and X_2 with N_1 and N_2 items respectively. The test X_2 should be equated to X_1 (we will note this by $X_2 \to X_1$). Let there exist m shared items $i_1, \cdots, i_m$ between X_1 and X_2, $i_k = \{i_k^1, i_k^2\}$, where i_k^1 is the index of the common item in X_1 and i_k^2 in X_2 respectively.

The transformation is based on four constants, representing the average values of the IRT parameters $(a^j_{i_k^j}, b^j_{i_k^j}, c^j_{i_k^j})$, $j = 1, 2$ of the common items i_k^j from test X_j,

$$a^1 = \frac{1}{m}\sum_{k=1}^{m} a^1_{i_k^1}, \quad a^2 = \frac{1}{m}\sum_{k=1}^{m} a^2_{i_k^2},$$

$$b^1 = \frac{1}{m}\sum_{k=1}^{m} b^1_{i_k^1}, \quad b^2 = \frac{1}{m}\sum_{k=1}^{m} b^2_{i_k^2},$$

where $(a^j_{i_k^j}, b^j_{i_k^j}, c^j_{i_k^j})$, $j = 1, 2$ are the IRT parameters of item i_k from test X_j.

Then, the coefficients A and B, which convert parameters of the items from test X_2 on the scale of the test X_1, can be calculated as follows

$$A = a^2/a^1 \tag{2}$$

$$B = b^1 - Ab^2. \tag{3}$$

The transformed parameters $(a_i^{x_1}, b_i^{x_1}, c_i^{x_1})$ of the test item i from test X_2 on the scale of test X_1 can be obtained from the original IRT parameters (a_i, b_i, c_i) as

$$a_i^{x_1} = a_i^2/A, \tag{4}$$

$$b_i^{x_1} = Ab_i^2 + B, \tag{5}$$

$$c_i^{x_1} = c_i^2. \tag{6}$$

Having the item parameters of both tests placed on the same scale, one can calculate the expected test score T_j of a person with latent trait value θ for the test X_j, $j = 1, 2$.

$$T_j(\theta) = \sum_{k=1}^{N_i} P_k(X = 1 \mid a_k^i, b_k^i, c_k^i, \theta). \tag{7}$$

Then, to equate the scores of X_1 and X_2, the next algorithm can be used.

2.2. Equating algorithm

1. For values less than $c_2 = \sum_{k=1}^{N^2} c_k^i$ – uniform step.
2. For values from c_2 to $\min T_2$: linear extrapolation
3. For values from $\min T_2$ to $\max T_2$: test score equating
4. For values greater than $\max T_2$: linear extrapolation

The test equating at step 3 of the algorithm is based on the idea that close values of the test scores should be performed by persons with close values of the latent trait. Thus, if t is a test score value of the test X_2,

$$\imath_d = \max idx\{T_2 < t\}$$

is the greatest index of the ability level, which gives test score less then t and the

$$\imath_u = \min idx\{T_2 > t\}$$

is the minimal index, for score greater than t, then the score value $\hat{T}_1$ of the test X_1 which corresponds to the same abilities is

$$\hat{T}_1 = T_1(\min\{\imath_d, \imath_u\}).$$

Here it is important to have the IRT paramaters (respectively the test scores), placed on the same ability scale (which is given by (4)).

2.3. An Example

Let us consider the following example. In Table 1 the IRT parameters of the items (in the first column) in the test X_1 (columns 2,3,4) and in the test X_2 (columns 5,6,7) are shown. Note that test X_2 has 12, but test X_1 has 10 items. This means that the test score of the test X_2 (from 0 to 12) should be placed on the scale of the test score of X_1, from 0 to 10. There are 3 items, shared between

Table 1: IRT parameters of two tests

Item i	X_1			X_2			X_2 on X_1 scale		
	a_i^1	b_i^1	c_i^1	a_i^2	b_i^2	c_i^2	$a_i^{X_1}$	$b_i^{X_1}$	$c_i^{X_1}$
1	0.59	0.52	0.14	0.34	3.31	0.24	0.56	2.17	0.24
2	1.49	2.12	0.20	0.15	2.53	0.22	0.26	1.69	0.22
3	0.44	0.83	0.19	1.10	1.52	0.23	1.83	1.09	0.23
4	1.02	2.16	0.36	0.18	3.12	0.19	0.30	2.05	0.19
5	0.40	1.56	0.17	0.33	1.37	0.18	0.55	1.00	0.18
6	0.97	1.68	0.21	1.09	2.04	0.32	1.81	1.40	0.32
7	0.52	−2.16	0.23	0.42	−0.37	0.16	0.71	−0.05	0.16
8	0.31	−0.07	0.22	0.41	1.26	0.20	0.68	0.93	0.20
9	0.76	−0.95	0.23	0.99	2.29	0.13	1.64	1.55	0.13
10	0.24	0.97	0.23	0.41	0.04	0.21	0.69	0.20	0.21
11				0.82	−0.93	0.21	1.37	−0.39	0.21
12				0.33	−3.19	0.21	0.56	−1.75	0.21

Table 2: Equationg test X_2 on the scale of X_1

X_2 score	$X_2^{X_1}$	$[X_2^{X_1}]$	θ
0	0	0	NaN
1	0.8786	1	NaN
2	1.7573	2	NaN
3	2.9673	3	-2.8
4	4.0501	4	-1.1
5	4.7697	5	-0.3
6	5.3598	5	0.3
7	5.9012	6	0.8
8	6.3951	6	1.2
9	6.9694	7	1.6
10	7.7819	8	2.1
11	9.0621	9	3.2
12	9.7846	10	4.2

the tests $i_1 = \{8, 10\}, i_2 = \{6, 5\}, i_3 = \{7, 12\}$ (item 8 in test X_1 is the same as item 10 from test X_2 etc.). Therefore the transformation coefficients $A = 0.6029$ and $B = 0.1711$ are calculated from (2) and (3). Then, the parameters of the items from test X_2, transformed on the scale of the test X_1, calculated according to (4) are presented in the last three columns of Table 1.

Applying the equating algorithm for the tests X_1 and X_2, placed on a common scale, the equaated score of the test X_2 can be calculated. The result is presented in Table 2. The first column is the original score of test X_2, the second and the third columns are the scores of X_2, scaled on X_1 and corresponding rounded (to the nearest integer) value. The last column of the table represents the ability of the person for achieving the corresponding test score, where NaN stands for which can not be estimated due to the lack of observed score.

3. Equating scores of sequence of tests

The same approach can be applied on a sequence of tests $X_1, \cdots, X_g$. Two requirements should be satisfied, as in the previous case. The parameters of all of the tests should be transformed on a common scale and there should be a sequence of shared items between them. The proposed algorithm can be performed to any couple of test (X_i, X_{i-1}) using (2) and (3) and a sequence of equating coefficients A_{i-1}, B_{i-1} can be calculated. Such a sequence is presented in Table 3.

Table 3: Equating of a sequence of tests

$X_g \to X_{g-1}$	$X_{g-1} \to X_{g-2}$	$\cdots$	$X_2 \to X_1$
A_{g-1}, B_{g-1}	A_{g-2}, B_{g-2}	$\cdots$	A_1, B_1

Then using a simple algebra one can represent the equating coefficients between X_g and X_1 ($X_g \to X_1$) using equations (8) and (9).

$$A = \prod_{k=1}^{g-2} A_k \tag{8}$$

$$B = B_1 + \sum_{k=2}^{g-2} \left(B_k \prod_{s=1}^{k-1} A_s \right) \tag{9}$$

4. TEQNCA Package

We have developed a software package (named TEQNCA) where the algorithm, presented in current paper, is implemented. The package is written in MATLAB and is available at http://www.ir-statistics.net/index.cgi/software-test-compare in both source code and precompiled version. A screen-shot of the main window of the GUI is presented on Figure 1.

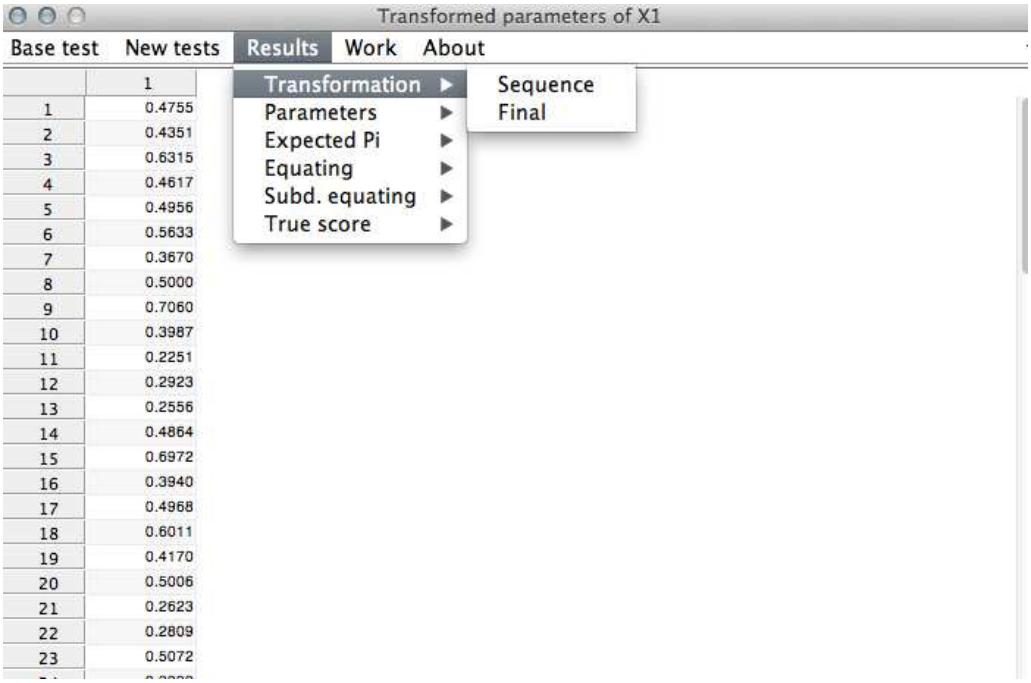

Figure 1: TEQNCA Package screen-shot

REFERENCES

[1] L. L. Cook, D. R. Eignor. IKT Equating Methods Educational Testing Service. 1991.

[2] L. Crocker, J. Algila. Introduction to Classical and Modern Test Theory. Warsworth, 1986

[3] M. J. Kolen. Traditional equating methodology. *Educational Measurement: Issues and Practice*, **7(4)** (1988), 29–36.

[4] F. M. Lord. Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, **14** (1977), 117–138.

[5] M. J. Kolen, R. L. Brennan. Test equating, scaling, and linking: Methods and practice (2nd ed.). New York, NY, Springer, 2004.

[6] D. Li, Y. Jiang, Y., A. A. von Davier, The accuracy and consistency of a series of IRT true score equating. *Journal of Educational Measurement*, **49** (2012), 167–189.

Dimitar Atanasov
New Bulgarian University
21 Montevideo Blvd
1618 Sofia, Bulgaria
e-mail: datanasov@nbu.bg

Dimiter Dimitrov
George Mason University
Fairfax Campus
West Building 2007 4400
University Dr. MS 6D2 Fairfax
VA 22030, USA
e-mail: ddimitro@gmu.edu