КВАЛИМЕТРИЧЕСКИЙ ПОДХОД В ОБРАЗОВАНИИ

УДК 303.094.7

Maslak Anatoly A.

Doctor of Science, Professor, Head of Laboratory for Objective Measurement, Affiliate of Kuban State University, Slavyansk-on-Kuban.
E-mail: ruch_measurement@mail.ru

Osipov Sergey A.

Candidate of Sciences, Associate Professor, Department of Mathematics and Informatics, Affiliate of Kuban State University, Slavyansk-on-Kuban.
E-mail: osipov21@mail.ru

Goncharova Tatyana N.

Postgraduate, Department of Mathematics and Informatics, Affiliate of Kuban State University, Slavyansk-on-Kuban.
E-mail: tngonchrova@mail.ru

INVESTIGATION OF MEASUREMENT PRECISION OF LATENT VARIABLES IN EDUCATION[1]

Abstract. *The objective* of the study is to investigate the measurement accuracy of latent variables depending on a number of dichotomous test items and variation range.

Methods: Investigation is based on the simulation experiments.

Results: The authors make recommendations for selecting a number of dichotomous test items and variation range depending on the required measurement precision of latent variables.

Scientific novelty: The research demonstrates statistical correlation between the measurement precision of latent variables and a number of test items and variation range.

Importance for practice: The research results can be used while developing the questionnaires and tests for measuring the latent variables.

Keywords: latent variable, Rasch model, measurement precision, dichotomous items, simulation experiment.

[1] Статья публикуется в авторской редакции.

Маслак Анатолий Андреевич

доктор технических наук, профессор, заведующий лабораторией объективных измерений филиала Кубанского государственного университета, Славянск-на-Кубани.
E-mail: ruch_measurement@mail.ru

Осипов Сергей Александрович

кандидат технических наук, доцент кафедры математики и информатики филиала Кубанского государственного университета, Славянск-на-Кубани.
E-mail: osipov21@mail.ru

Гончарова Татьяна Николаевна

аспирант кафедры математики и информатики филиала Кубанского государственного университета, Славянск-на-Кубани.
E-mail: tngonchrova@mail.ru

ИССЛЕДОВАНИЕ ТОЧНОСТИ ИЗМЕРЕНИЯ ЛАТЕНТНЫХ ПЕРЕМЕННЫХ В ОБРАЗОВАНИИ

Аннотация. *Цель* работы – исследование точности измерения латентных переменных в зависимости от числа дихотомических тестовых заданий и диапазона их варьирования.

Методика и методы: исследование проведено на основе имитационного моделирования.

Результаты. Разработаны рекомендации по выбору числа дихотомических тестовых заданий и диапазона их варьирования в зависимости от требуемой точности измерения латентных переменных.

Научная новизна. Получена статистическая взаимосвязь между точностью измерения латентной переменной и числом тестовых заданий и диапазона их варьирования.

Практическая значимость. Результаты исследования могут быть использованы при построении тестов и опросников для измерения латентных переменных.

Ключевые слова: латентная переменная, модель Раша, точность измерения, имитационный эксперимент.

Introduction

In education and other social systems, the majority of variables, for example students' proficiency, are latent, i.e. they cannot be measured in such way as, for example weight or length. In the middle of the last century, there appeared a possibility of measuring the latent variables on a linear scale due to the developed theory of latent variables. After the work of Georg Rasch [1], a large number of research papers applying and discussing the Rasch model have been published. It allowed shifting to essentially more advanced level of research in education and other social systems [2–4]. But still there are some open issues. One of them is about the number of test items needed to obtain the required precision of measurement of latent variable [5, 6]. Another issue is the influence of a range of items variation on measurement precision of a latent variable.

The work purpose

Tests and questionnaires play an important role in individual decision-making in areas such as educational testing, personnel selection, and many others. The research is aimed at measuring the precision of a latent variable depending on number of dichotomous items and a range of its variation. The need for this research results from the fact that the measurement cost substantially depends on a number of test items. Therefore, it is important to choose a minimum number of test items to provide the required precision of the latent variable measurement.

Methods

The authors use the paradigm of measuring the latent variables, developed by the Danish mathematician G. Rasch. In this paradigm the estimation of a latent variable, for example students' proficiency, does not depend on difficulty of a set of test items [7, 8]. Besides, students' proficiency and items difficulty are measured on the same linear interval scale in logits. By means of linear operations, the scale of latent variable can be transformed into any other scale. For example, the Federal Centre of Testing of the Ministry of Education and Science of the Russian Federation transform logits of the Unified State Exam into a 100-mark scale.

It is convenient to estimate the precision of measurement by a standard error. So there is a need to establish quantitative dependence of standard measurement error of a latent variable on a number of test items and range of its variation.

Research was based on the simulation experiment. Such method of research is used due to the fact that the model of measurement (Rasch model) is a probabilistic and nonlinear one. Analytical research methods in such situations are ineffective [9, 10].

For generating of a matrix of data the following scheme was used. Students' proficiency varied from –4.0 to +4.0 logits. This range covers the majority of practical Rasch model applications. For convenience of the analysis of measurement precision 17 values of a latent variable (17 levels of proficiency) was used with step.5 logits: the first level equals –4.0, the second level equals –3.5, …, the seventeenth – +4.0 logits. Each of 17 levels was used triple that is in a generated matrix there is 51 lines.

Difficulty of test items varies on intervals [–2; +2] and [–4; +4] logits. There were used 10 set of test items. The first set consists of 10 items, the second consists of 20, …, the tenth of 100 items. In each set items were evenly distributed within above-mentioned intervals.

In terms of design of experiment there was used a four-way block randomized plan with replication having three treatment factors A, B, C and block-factor D [11]:

- Factor A is the range of test items variation; a = 2 levels: (–2.0, 2.0 logits), (–4.0, 4.0 logits).
- Factor B is the student location; b = 17: (–4.0, –3.5, –3.0, …, +4.0 logits).
- Factor C is the number of items set; c = 10: (10, 20, 30, …, 100).
- Block-factor D varied on three levels; d = 3.

The response variable Y is the standard error of measurement of students' proficiency (latent variable).

Data of simulation experiment were generated in accordance with Rasch model for dichotomous items.

$$p_{ij} = \frac{e^{\beta_i - \delta_j}}{1 + e^{\beta_i - \delta_j}}, \tag{1}$$

where p_{ij} – probability of a right answer of i-th student on j-th item, β_i – level of i-th student proficiency (logits), δ_j – difficulty of j-th test item (logits).

Then based on the calculated probabilities (1) data of dichotomous matrix are generated:

$$x_{ij} = \text{Int}\,(p_{ij} - \text{Rnd} + 1), \tag{2}$$

Where Int (Y) – the whole part of number Y, Rnd – a random number evenly distributed on an interval (0; 1).

As an example in Table 1 the generated matrix of data for 30 items which varies in a range from – 4.0 logits to + 4.0 logits is presented.

Table 1

Data of simulation experiment with 30 items

Student	Student Proficiency	Items (30)
1	4.0	111111111111111111111111111101
2	4.0	111111111111111111111111111001
3	4.0	111111111111111111111111111011
4	3.5	111111111111111111111101110010
5	3.5	111111111111111111111101111000
6	3.5	111111111111111111110111111110
7	3.0	111111111111111111111111100010
8	3.0	111111111111111111111111011010
9	3.0	111111111111111011110110111100
10	2.5	111111111110111111111111010000
11	2.5	111111111111111111101011100011
12	2.5	111111111111111111100111100000
13	2.0	111111111111111101100101000010
14	2.0	111111111111110111100111001100
15	2.0	111111111111111111111110010000
16	1.5	111111111111111101111111001000
17	1.5	111111111111011111101010000000
18	1.5	111111111111111111001100000000
19	1.0	111111111111011101010110010000
20	1.0	111111111111110101100100000000
21	1.0	111111110101100011001000100000
22	0.5	111111111111110100001100000000
23	0.5	111111111001101111001000100000
24	0.5	111111111111111111111000000000
25	0.0	111111100111101101000000000000
26	0.0	111111110111100100000000000000
27	0.0	111111111010011001000100000000
28	–0.5	111111101111111100010000000000
29	–0.5	111111110010111000000000100000
30	–0.5	111111111110110100100000000000
31	–1.0	111111011110011101000000000000
32	–1.0	011111111111000010001001000000
33	–1.0	111111011110000100000000000000
34	–1.5	011110111111100000001000000000
35	–1.5	111010101001000000000001000000
36	–1.5	111111111101100000000010000000
37	–2.0	111111110100000000000001000000
38	–2.0	111111100100100001000000000000
39	–2.0	111110100000000100100000100000
40	–2.5	010011010010000000000000000000
41	–2.5	111000000100011000100000000000
42	–2.5	100101000000000000000000000000
43	–3.0	110110100000000100000000000000
44	–3.0	011101000000000100000000000000
45	–3.0	001000010000000010000000000000
46	–3.5	111010100000000000000000000000
47	–3.5	010000000000000000010000000000
48	–3.5	111000000000000000000000000000
49	–4.0	101000010000000000001000000000
50	–4.0	111100000010000000000000000000
51	–4.0	111010000000000000000000000000

Based on the generated data matrix there were obtained estimations of students' proficiency. For these purposes dialogue system «MLV» developed by authors of this paper (Measurement of Latent Variables), developed in Laboratory for Objective Measurements of the Kuban State University was used.

Precision of measurement of students' proficiency is characterized by a standard error of measurement. For *i*-th student the standard error is:

$$SE_i = \frac{1}{\sqrt{\sum_{j=1}^{m} p_{ij}(1-p_{ij})}}, \qquad (3)$$

where p_{ij} – probability of a right answer of *i*-th student on *j*-th test item; m – number of test items. Unlike the classical theory of testing where a measurement error same for all students, in the theory of latent variables these errors are different. For example, if *i*- th student has correctly answered all items the standard error tends to infinity. If the student has incorrectly answered all items the standard error also tends to infinity. The least error is observed for students who correctly answer approximately half of test items. From the formula (3) follows, that on the edges of a scale the standard error has maximum values.

Results

Measurement precision of latent variable obtained from the simulation experiment is described by the following model

$$y_{ijkl\,(m)} = \mu + \alpha_i + \beta_j + \gamma_k + \tau_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl\,(m)}, \quad (4)$$

Where $y_{ijkl\,(m)}$ is the response variable which is standard error of measurement of latent variable;

μ is the overall mean;

α_1, α_2 are the main effects for the levels of factor *A*;

$\beta_1, \beta_2, \ldots, \beta_{17}$ are the main effects for the levels of factor *B*;

$\gamma_1, \gamma_2, \ldots, \gamma_{10}$ are the main effects for the levels of factor *C*;

τ_1, τ_2, τ_3 are the main effects for the levels of factor *D*;

$(\alpha\beta)_{ij}$ are the interactions for the combinations of factors *A* and *B*;

$(\alpha\gamma)_{jk}$ are the interactions for the combinations of factors *A* and *C*

$(\alpha\beta\gamma_{ijk})$ are the interactions for the combinations of factors *A*, *B* and *C*;

$\varepsilon_{ijkl\,(m)}$ are the errors that satisfy the conditions of mean equal to 0, equal variances, normality, and independence.

With the purpose of an illustration Figure 1 displays precision of measurement of latent variable based on 50 items.

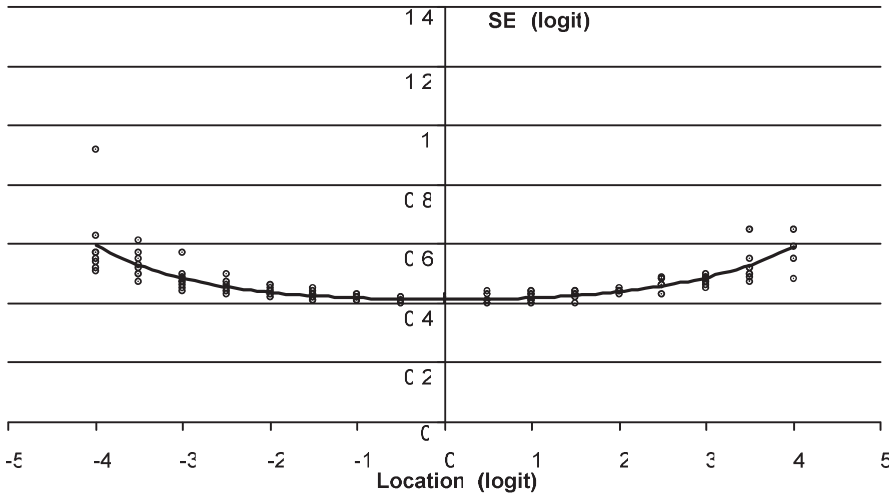

Figure 1. Precision of measurement of a latent variable based on 50 items with (–4.0, +4.0) range of items variation

The statistical analysis of measurement precision of a latent variable

Results of the variance analysis (ANOVA) of a standard measurement error are presented in Table 2.

Table 2

ANOVA of standard error of measurement

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	*F*	Sig.
Factor A	.078	1	.078	18.473	<.001
Factor B	24.393	16	1.525	362.435	<.001
Factor C	111.507	9	12.390	2945.430	<.001
Block-factor D	.039	2	.019	4.586	.010
Interaction AB	4.395	16	.275	65.297	<.001
Interaction AC	3.423	9	.380	90.417	<.001
Interaction BC	1.373	144	.010	2.267	<.001
Interaction ABC	.991	144	.007	1.636	<.001
Error	11.433	2718	.004		
Total	157.631	3059			

All sources of variation are significant. In a certain degree it is due to the great volume of experimental data. The average values of measurement precision of a latent variable depending on the items set are presented in Table 3.

Table 3

Mean Standard Error of Items Set

Set of Items	Mean	Volume	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
10	.993	306	.004	.986	1.000
20	.708	306	.004	.701	.715
30	.604	306	.004	.597	.612
40	.539	306	.004	.532	.547
50	.477	306	.004	.470	.484
60	.443	306	.004	.436	.450
70	.407	306	.004	.399	.414
80	.375	306	.004	.368	.382
90	.362	306	.004	.355	.369
100	.343	306	.004	.336	.350

Important aspect of the investigation is the finding out measurement precision depending on location of persons on a scale (Figure 2).

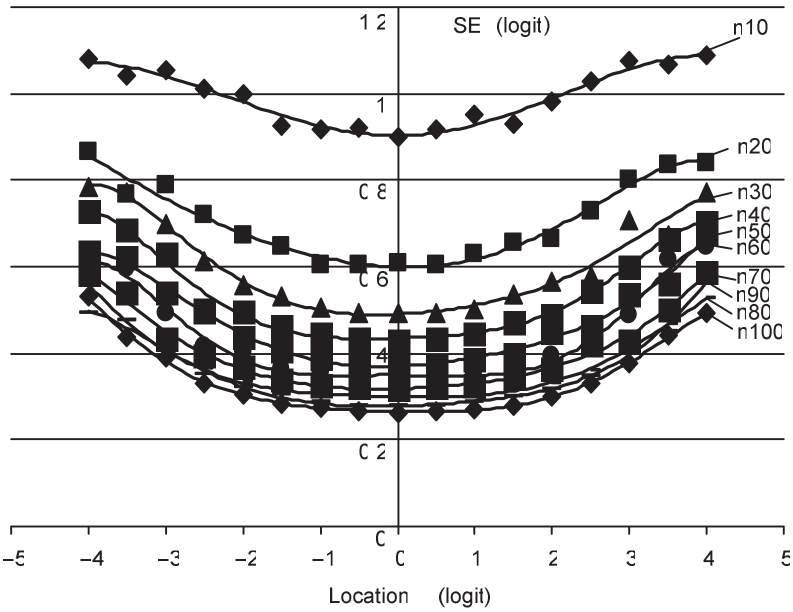

Figure 2. Standard error of measurement of a latent variable depending on students' location on a scale and numbers of test items

Fig. 3. A standard error of measurement of a latent variable depending on students' location on a scale and a range of items variation

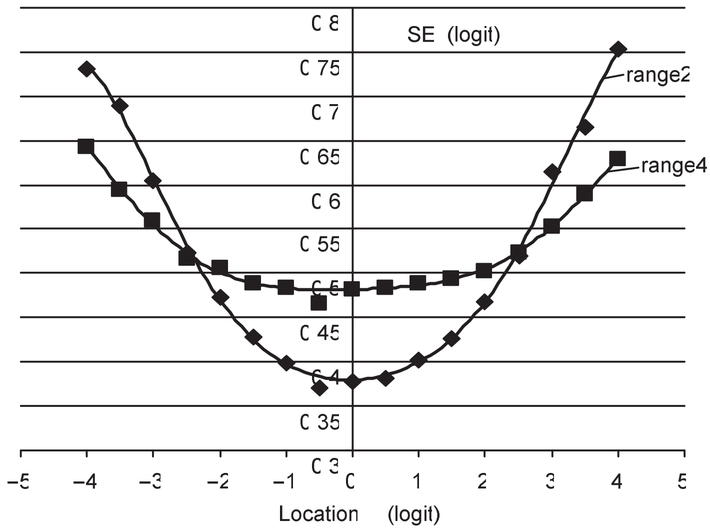

Figure 3. Shows the influence of a range of items variation on measurement precision of a latent variable

On the average at a small interval of items variation measurement precision a little higher, than at wider range (Table 4).

Table 4

Mean standard error of persons depending on range of items variation

Range	Mean	Volume	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
[–2.0, + 2.0]	.520	1530	.002	.517	.523
[–4.0, + 4.0]	.530	1530	.002	.527	.533

Discussion

Students' ability and items difficulty varied in a simulation experiment over a wide range: from –4.0 logits to +4.0 logits. This wide range covers the majority of practical testing.

As a result of the carried out research it is shown, that for achievement of a standard error of measurement in.5 logits there are enough 50 dichotomous items (Table 3). The further increase in number of items slightly increases measurement precision. So, even 100 dichotomous items do not provide measurement precision less than.3 logits (Figure 2).

The range of a variation of test tasks significantly influences measurement precision of latent variable. Besides measurement precision in the middle of a scale is higher than on the edges of a scale (Figure 3).

The results are obtained for the case that latent variable vary from –4.0 to +4.0 logits. It is obvious, that for drawing conclusions concerning other intervals of a variation of a latent variable additional investigation is required.

Another possible way of increasing of measurement precision is replacing dichotomous items by polytomous ones. In the last case it is possible to take into account partially correct variants of the answer.

Conclusion

1. For achieving a standard measurement error of 5 logits, 50 dichotomous items is enough. It is necessary to notice, that students' proficiency and test item difficulty vary in the same interval: from –4.0 to +4.0 logits.

2. Measurement precision can slightly increase when the items number exceeds 50. However, even 100 items do not provide the measurement precision below 3 logits.

3. The measurement precision of students' proficiency (latent variable) is higher in the middle of the scale and lower on its edges.

Acknowledgment

This research was supported by the grant from the Russian Foundation for Basic Research 05–06–80110 «Development of the technique of measurement on an interval scale of latent variables in social and economic systems» (2005–2007), and the grant from the Russian Foundation for Humanities 08–06–00694a «Development of the technique of quality analysis of questionnaires used for measurement of latent variables» (2008–2010).

Статья рекомендована к публикации
д-ром пед. наук, проф. Н. Е. Эргановой

References

1. Rasch G., 1980. Probabilistic models for some intelligence and attainment tests (Expanded edition, with foreword and afterword by Benjamin D. Wright). Chicago: *University of Chicago Press*. P. 199.

2. Maslak A. A. Measurement of latent variables in social systems. Slavyansk-on-Kuban. *Publishing center of KubSU*. 2012. P. 432. (In Russian)

3. Maslak A., Karabatsos G., Anisimova T., Osipov S. Measuring and Comparing Higher Education Quality between Countries Worldwide. *Journal of Applied Measurement*. 2005. V. 6. № 4. P. 432–442.

4. Crocker L. Algina Introduction to Classical and Modern Test Theory. Ohio. *Cengage Learning Mason*. 2008. P. 527.

5. Kruyen P. M. Using Short Tests and Questionnaires for Making Decisions about Individuals: When is Short too Short? *Ridderkerk*. 2012. 161 p.

6. Kruyen P. M., Emons, W. H. M. and Sijtsma K. Test Length and decision quality in personnel selection: When is short too short? *International Journal of Testing*. 2012. № 12. P. 321–344.

7. Letova L. V., Maslak A. A., Osipov S. A. Family of Rasch f models for objective measurement of latent variables. *Informatization of Science and Education*. 2013. № 4 (20). P. 131–141.

8. Humphry S. M., Andrich D. Understanding the unit in the Rasch Model. *Journal of Applied Measurement*. 2008. № 9 (3). P. 249–264.

9. Wilson M. Constructing Measures: An Item Response modeling approach. Mahwah. *Lawrence Erlbaum Associates* Publ. 2005. P. 228.

10. Wolfe E. W., Smith V. Instrument Development Tools and Activities for Measure Validation Using Rasch Models; Part I – Instrument Development Tools. *Journal of Applied Measurement*. 2007. № 8 (1). P. 249–264.

11. Maslak A. A. Fundamentals of Design of Experiment in Management. Slavyansk-on-Kuban. *Publishing center of KubSU*. 2013. № 116.