LOOK ACADEMIC PUBLISHERS
OPEN ACCESS

INTERNATIONAL JOURNAL OF ENVIRONMENTAL & SCIENCE EDUCATION
2016, VOL. 11, NO. 18, 11899-11905

Structure of the Complexity Parameter of the Test Physic Problem in Rasch Model

Vyacheslav T. Volov[a] and Alexander A. Gilev[b]

[a]Modern University for the Humanities, Samara, RISSIA; [b]Samara State Technical University, Samara, RUSSIA

ABSTRACT

In today's item response theory (IRT) the response to the test item is considered as a probability event depending on the student's ability and difficulty of items. It is noted that in the scientific literature there is very little agreement about how to determine factors affecting the item difficulty. It is suggested that the difficulty of the item increases with the number of key elements, which are the data elements used in solving the item. Based on the statistical analysis of solutions of specially designed test items it has been concluded that there is a linear dependence of the item difficulty in the Rasch model on the number of key elements of the solution. The result allows taking an unbiased look at the difficulty of test items at the stage of their development, as well as predicts the outcome of testing.

KEYWORDS
Rasch model, the item difficulty parameter, cognitive operations

ARTICLE HISTORY
Received 20 July 2016
Revised 28 October 2016
Accepted 19 November 2016

Introduction

Lately, such form of control as testing is implemented in the practice of teaching physics in high schools. It is used for objective assessments of the professional activity of the Department and University, as well as the ongoing and in-session performance appraisal of students. The basis of modern test theory is Item Response Theory (IRT). In Russia it is known from the works by Yu.M. Neumann & V.A. Khlebnikov (2000), M.B. Chelyshkova (2002), V.S. Kim (2007), etc. as the theory of modeling and parameterization of pedagogical tests. Response to the test item is considered as a probability event dependent on two latent, i.e. not subjected to direct measurement, variables, namely, the level of the examinee's ability and level of the item difficulty. The probability of the

CORRESPONDENCE Vyacheslav T. Volov ✉ vtvolov@mail.ru

correct test item execution can be described by the function of success, the simplest model of which was proposed by G. Rasch (Neiman & Khlebnikov, 2000):

$$P_{ij} = \frac{1}{1 + \exp[-1{,}7 \cdot (\theta_i - \beta_j)]} \qquad (1)$$

In the equation (1) parameters θi (the level of the i student's i ability) and βj (the j-item difficulty) are measured in logits. The Rasch function defined on the segment [0, 1] and called the logistic function is equal to the probability that the student i with the ability level of θi logits will perform the j-item of βj logit difficulty. A scale factor of 1.7 in the equation (1) is introduced to bring the Rasch model into coincidence with the model of Fergusson, where the probability of a correct response to the item is expressed by the integral of the normal distribution (Gilev, 2011a; Kim, 2007; Gilev, 2016). For all examinees the j-item difficulty βj is an objective feature of the item independent of students' ability. The difficulty parameter β is determined only at the end of testing by time-consuming calculations with the accuracy up to the arbitrary constant value. However, its structure and methods of control remain unknown. There are also no criteria to assess numerical values of β and the difficulty of tests under development. A preliminary assessment of the test difficulty is quite subjective and based only on the experience and intuition of the developer.

The goal of research

In the scientific literature there is no a single approach to the definition of the item difficulty (Volov, 2007; Volov & Kaptsov, 2009; Dzhalmuhambetov & Stefanova, 2009). G.A. Ball (1990) considers an algorithmic method based on the assessment of the number of operations required for the response to the item. The method to determine the difficulty generated from the analysis of the physical problem structure seems to be simpler. The difficulty is determined by the number of phenomena, processes and physical quantities whose values need to be specified. I.Ya. Lerner (2000) believes that the item difficulty depends on the amount of data in the condition which requires consideration and mutual correlation. This intuitively clear statement allows coming out with the assumption that with the increase of the solution key elements the item solution difficulty described in the Rasch model by the difficulty parameter β increases. The more difficult the item, the greater number of key elements should be involved in its solution and it should be described by the greater parameter of difficulty β. The goal of the research is to define the dependence of the item difficulty parameter β on the number of key elements of the solution (Andreev, 2000).

Research methods

A set of tests based on the unit "Electrostatics" in the course of general physics was developed to define the item difficulty parameter ß and its dependence on the number of key elements. The test was designed to investigate the operational aspect of the solution, that's why the entire "knowledge" component was in the item text to eliminate the influence of the "unfamiliarity" factor (Gilev, 2011b; Gilev, 2011c; Gilev, 2007). The items were syllogisms containing different numbers of general and particular judgments, as well as of the source data used to make conclusions. The solution time of twenty test items was 20-23 minutes (Gilev, 2008). Each correct answer was estimated at 1 point, incorrect – 0 points. As many as 76 first-year students of civil engineering and computer informatics took part in testing.

Item 1. There are three identical flat capacitors. Q1, Q2, Q3 are charges distributed on the plates of capacitors. U1, U2, U3 are their potential differences and E1, E2, E3 are the field intensity in the inner space. If

1. E3 > E1 and U2 < U3, on what capacitor is the largest potential difference?

2. U3 >U2, and E1 > E3, on what capacitor is the largest charge?

Item 2. There area three flat capacitors permanently connected to the source of EMF (electromotive force). C1, C2, C3 are their capacity. Q1, Q2, Q3 are charges distributed on the plates of capacitors; d1, d2, d3 are the distance between the plates, E1, E2, E3 are the field intensity in the inner space. If

3. d1 is twice larger than d3 and d2 is four times less than d3, how many times is E1 less than E3?

4. E3 > E1 and E2 < E3, what capacitor has the least distance between plates?

5. C3 > C2 and C3 < C1, what capacitor has the largest charge?

Results and Discussion

Upon completion of testing the binary matrix of test results (Kim, 2007; Crocker & Algina, 2010) was formed and a measure of each item tj difficulty, as the ratio of the number of students incorrectly responded the j question to the total number of test participants, was defined. The parameter tj is, in fact, a probability of the wrong solution of the corresponding item. The range of its changes is the segment [0.1]. For very difficult questions the parameter tj tends to the largest value equal to 1, for very simple ones to zero. In practice we usually use the parameter gi= 1-tj equal to the probability of the correct solution of the item i. From the initial set of items we removed the simplest and most difficult ones with the following values of the difficulty parameter, i.e. g=1 and g=0. All the proposed questions on the value of the difficulty parameter were divided into three groups: simple $0.8<g<1$, questions of intermediate difficulty with a parameter g from the range $0.4<g<0.8$ and difficult items had a sufficiently small value of the parameter $g<0.4$. The score of each student b in the segment [0.1] as the relative portion of correctly solved items, and the average score bCP of the entire group under the test were also determined. These are numerical parameters for the distribution of test results: range Δb=bmax – bmin =0.51; mathematical expectation bCP =0.64; dispersion D=0.036; root-mean-square deviation σ=0.19; median bm=0.67; asymmetry s=-0.24; kurtosis ε=-1.1. The distribution of the probability density of solving test items by students on the score scale from the segment [0.1] approximately corresponded to the Gauss function (fig.1).

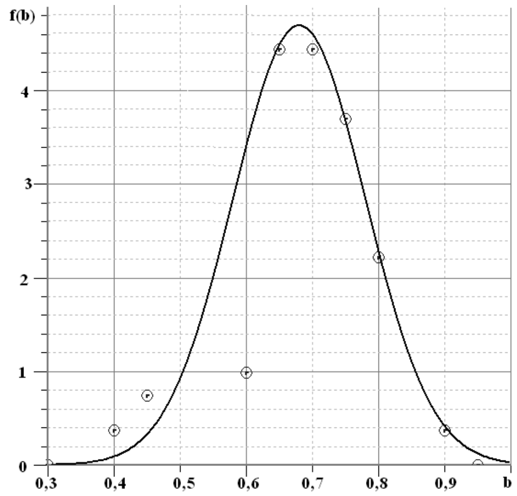

Figure 1. The distribution of the probability density f(b) on the score scale b.

From now on key elements of the solution will mean the elements of the data to be necessary used in solving the item. They may be in the condition or the item question, set explicitly, e.g. as values of physical quantities, and implicitly, as statements or dependences, which are the elements of knowledge. The necessary cognitive operations to perform the test items are the analysis, comparison and logical conclusions based on them. Each test item contains the minimum number of m data elements (specific values or approval of the functional dependence of magnitudes) required to form the solution and response the item question. The list of these elements and the m value for item samples are shown in Table 1. The table contains in the symbolic form the minimum list of key elements for the solution corresponding to the optimal sequence of conclusions. Sometimes students' real solutions contained unnecessary additional elements. The number of these elements was great and it exceeded the minimum by one-two units.

Table 1. The list and amount of key elements of the solution

Item number, j	Key elements for the solution formation	The amount of key elements of the solution, m_j
1.	E3, E1, $E = \frac{U}{d}$, d=const, U1, U2, U3	7
2.	E1, E3, $E = \frac{U}{d}$, d=const, $Q = C \cdot U$, C=const, U1, U2, U3	9
3.	d1/d3, d2/d3, $E = \frac{U}{d}$, U=const, E1/E3	5
4.	E1, E2, E3, $E = \frac{U}{d}$, U=const, $d_{min} \sim \frac{1}{E_{max}}$	6

5.	C1,C2,C3, $C = \frac{Q}{U}$, Q=CU, U=const, Q_{max}~C_{max}	7

The analysis of the structure of the presented solutions shows that the item difficulty depends on the number of m used in the solution of information elements. As their number increases the operation of analysis primarily becomes difficult and requires devoting energies to it. This increases the number of elementary operations needed to obtain the final result. Within the time-limited conditions of carrying out the test it leads to the reduction in the number of students who have performed the test successfully, or to the increase of the item difficulty parameter (Popov, Ustin & Molchanov, 2009; Osipov & Marshalova, 2013; Popov & Ibragimov, 2007)

Information contained in the binary matrix of students' primary responses Aij (0 or 1) is sufficient to determine the values of latent variables θj and βi characterizing the level of the student's j achievements and the difficulty of the item i. In the IRT theory students' assessments Sj (j=1…N, N is the number of students) and assessments of item difficulty Wi (i=1…M, M is a number of items) are considered as random values.

$$\begin{cases} S_j = \sum_{i=1}^{M} A_{ij} \\ W_i = \sum_{j=1}^{N} A_{ij} \end{cases}, \qquad (2)$$

They can be implemented with different combinations of summands which are binary evaluations Aij (0 or1) of student's responses j to the test item i. Assessments Sj and Wi should then be average values og the corresponding sums

$$\begin{cases} S_j = \left(\sum_{i=1}^{M} A_{ij}\right)_{сред} = \sum_{i=1}^{M} (A_{ij})_{сред} \\ W_i = \left(\sum_{j=1}^{N} A_{ij}\right)_{сред} = \sum_{j=1}^{N} (A_{ij})_{сред} \end{cases} \qquad (3)$$

The average assessment of the j student's response to the test item i is equal to the corresponding probability Pij of the correct response specified in the Rasch model by the function of the success and the following ratio:

$$(A_{ij})_{сред} = P_{ij} = \frac{1}{1 + \exp[-1,702 \cdot (\theta_j - \beta_i)]} \qquad (4)$$

After a series of transformations from relations (3) and (4) we will obtain the system of nonlinear equations (N+M)required to calculate variables θ1, θ2,…, θN and и β1, β2,…, βM characterizing the level of students' achievements and test item difficulty (Neiman & Khlebnikov, 2000):

$$\begin{cases} b_j - \frac{1}{M} \cdot \sum_{i=1}^{M} \{1 + \exp[-1{,}702 \cdot (\theta_j - \beta_i)]\}^{-1} = 0 \\ g_i - \frac{1}{N} \cdot \sum_{j=1}^{N} \{1 + \exp[-1{,}702 \cdot (\theta_j - \beta_i)]\}^{-1} = 0 \end{cases} \quad (i=1,\dots,M;\ j=1,\dots,N) \quad (5)$$

To solve this system of equations by the method of successive approximations, standard tools of the mathematical package MathCAD were used. The values of the variables obtained through a numerical approximate solution of the system (5) in the sequence described in the work by M.B. Chelyshkova (2002) were taken as a first approximation to accelerate the convergence. To validate the found root values the numerical solution of the system of equations based on another algorithm (described in work 7) was carried out. The difference of the results obtained after normalization of the average level of the item difficulty βCP=0 did not exceed 0.001 logit. The results of calculations of test item difficulty parameter are given in Table 2.

Table 2. Values of test item difficulty parameter

Item number, j	1	2	3	4	5
m	7	9	5	6	7
β_j (logits)	0,44	4,81	-2,70	-1,02	0,44

The calculation of the item difficulty parameter values and the number of the solution key element **m** correlation revealed the presence of linear dependence (Pearson correlation coefficient k=0.9):

$$\beta(m) = A \cdot (m - B) \quad (6)$$

Test items from other sections of General physics (mechanics, molecular physics, etc.) gave similar results under the testing 47 senior pupils (10-11 forms) and 49 first and second year students of the University. Binary matrixes of test responses were processed in the sequence considered above. We calculated the item difficulty parameter and the number of the solution key elements. For all groups of test-takers the values of these variables are linearly dependent. The coefficient of proportionality A in the relation (6) varied from 0.6 to 1.4 depending on the age of test-takers and test items. Nevertheless, the Pearson correlation was statistically significant and ranged from 0.6 to 0.97. Please note that the scale of item difficulty is interval. The values of β are defined up to an arbitrary constant. However, under normalization of the average level of the item difficulty β_{CP}=0 the β parameter value was changed slightly in the range from 5.8 to 6.4 for all groups of test-takers.

Conclusion

The difficulty parameter of the item as its objective characteristic is linearly dependent on the number of information key elements required to form the solution. The obtained result allows us to realistically asses the difficulty of test items at the stage of their development and to predict the outcome of the test.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Vyacheslav Teodorovich Volov - Corresponding Member of Russian Academy of Education, Dr. Sc. in Education ,Dr. Sc. in Phys.& Math., Professor, Director of Samara Affiliate of the Modern University for the Humanities, Samara, Russia.

Alexander Alexandrovich Gilev - PhD in Phys.& Math., Professor of Department of General and Applied Physics and Chemistry of Architecture and Civil Engineering Institute of Samara State Technical University, Samara, Russia.

References

Andreev, V.I. (2000). *Education: Training for creative self-development*. Kazan: Center for Innovative Technologies.

Ball, G.A. (1990). *The theory of academic pursuits: Psychology and education aspect*. Moscow: Pedagogic.

Chelyshkova, M.B. (2002). *Theory and practice of designing educational tests.* Moscow: Logos.

Crocker, L. & Algina, J. (2010). *Introduction to classical and modern test theory*. Moscow: Logos.

Dzhalmuhambetov, A.U. & Stefanova, G.P. (2009). *Tasks - problems, tasks assessment on physics, methods of their solution: the manual*. Astrakhan: Izd Gos.ped.un.

Gilev, A.A. (2007). Cognitive analysis of the process of training solutions of physical problems. *Physics in Higher Education, 2,* 26-37.

Gilev, A.A. (2008). *Workshop on the decision of physical problems in a technical college: the manual*. St. Petersburg: Lan 'Publisher.

Gilev, A.A. (2011c). *Cognitive competences: development and diagnostics in the physics course of the higher technical school.* Samara: SGASU.

Gilev, A.A. (2011b). Level measurement logical inference. *Proceedings of the SFU. Jurisprudence, 8,* 175-180.

Gilev, A.A. (2011a). Quality diagnosis of cognitive processing verbal information in solving physical problems. *Physics in Higher Education, 17,* 16-27.

Gilev, A.A. (2016). *Methodical system of cognitive competence development of students in teaching physics: monograph*. Samara: SGASU.

Kim, V.S. (2007). *Testing of educational achievements.* Ussuriisk: University press UGPI.

Lerner, I.Ya. (2000). *Problem-based learning, The theory and methods of learning physics in school: General quesrions.* In S.E. Kamenetski, N.S. Purysheva (Eds.). Moscow: Publishing center "Academy".

Neiman, Yu.M. & Khlebnikov, V.A. (2000). *Introduction to the theory of modeling and parameterization of education tests.* Moscow: Nauka.

Osipov, P.N. & Marshalova, I.N. (2013). Formation of research competence of the future engineers in design activity. *Herald of Kazan Technological University, 16,* 194-197.

Popov, L.M. & Ibragimov, E.N. (2007). Psychology intellectual development of the student-activity. *Scientific notes of the Kazan State University, 149,* 5-19.

Popov, L.M., Ustin, P.N. & Molchanov, E.I. (2009). Orientation of students on the intellectual and ethical self-development and self-development. *Education, 6(16),* 3-8.

Stenner, A.J., Wright, B.D. & Linacre, J.M. (1994). From P-values and raw score statistics to logits. Direct access: http://www.rasch.org/rmt/rmt81.htm.

Volov, V.T. & Kaptsov, A.V. (2009). Methods of diagnosis and management of the dynamics of socio-psychological characteristics of the person based on the information and thermodynamic approach. *Bulletin of the University of RAO, 9,* 60-64.

Volov, V.T. (2007). On the principles of the system of education. *Philosophy of Education, 4(21),* 4-11.