



## FUZZY CLASSIFICATION OF DICHOTOMOUS TEST ITEMS AND SOCIAL INDICATORS DIFFERENTIATION PROPERTY

Aleksandras KRYLOVAS<sup>1</sup>, Natalja KOSAREVA<sup>1\*</sup>, Julija KARALIŪNAITĖ<sup>1,2</sup>

<sup>1</sup>*Mathematical Modelling Department, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania*

<sup>2</sup>*Information Systems Department, Vilnius College, J. Jasinskio g. 15, LT-01111 Vilnius, Lithuania*

Received 19 May 2017; accepted 07 April 2018

**Abstract.** In many fields of human activities such as economics, sustainable development, construction, human resources management etc., dichotomous tests are employed to measure some observed property, for example knowledge level in a specific field or applicant's eligibility for a job position. Fuzzy classification method for dichotomous test items is proposed in this paper. Depending on the observed property, each test item may well differentiate all testees or only the testees who are strong or weak at that property. Also, the test item may badly differentiate all testees and be inappropriate for that purpose. The method presented in the paper may be applied for small groups of testees with known estimates of the investigated property, for example raw test scores. The proposed method for dichotomous test item classification is based on the fuzzy set theory. Though the tests were originally constructed for knowledge measurement, their mathematical models can be applied for social indicators and wide range of other areas.

**Keywords:** mathematical modelling, fuzzy sets, dichotomous tests, social indicators, least squares method.

**JEL Classification:** C02, C43, C38, C63, C83, C99, D81, I25, O15.

### Introduction

Knowledge testing problem is the object of investigation of Classical Test Theory (CTT) and Item Response Theory (IRT). CTT was originally the main framework for analyzing and developing standardized tests (Lord & Novick, 1968). The main idea of CTT – the observed score of a testee is an estimate of his true knowledge score and some unobservable measurement error. Nevertheless, this approach has the shortcoming – it did not investigate the interrelation between the test takers ability level and test difficulty. Since the beginning of the 1970's IRT in essence replaced CTT and became the main theoretical framework used

---

\*Corresponding author. E-mail: [natalja.kosareva@vgtu.lt](mailto:natalja.kosareva@vgtu.lt)

for knowledge level measurement (Rasch, 1960; Lord, 1980; Hambleton, Swaminathan, & Rogers, 1991).

The following well-known logistic IRT models for binary response are the one (1PL), two (2PL) and three (3PL) parameter logistic IRT models. The 3PL IRT model can be defined using the notion of item characteristic function:

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{a_i(\theta - b_i)}},$$

where  $P_i(\theta)$  is the probability that a given test-taker with ability  $\theta$  answer an item  $i$  correctly,  $a_i$  is the item discrimination,  $b_i$  is the item difficulty (location) and  $c_i$  is the probability that test taker will guess an item. According to item response theory the value of test information function  $I(\theta)$  is the reciprocal of the standard error of estimation  $SE(\theta)$ , so the more information we have at a given ability level  $\theta$ , the less is the error of measurement. Therefore, it makes sense to elaborate such tests adapted for each group of test takers, which could provide more information. Kosareva and Krylovas (2011) proposed to reduce the standard error of measurement not only by increasing the number of examinees in calibration group or/and number of test items proposed to the examinees but also by expanding the set of item characteristic functions  $P_i(\theta)$ .

Various economic indices, social indicators and similar values defined by some agreements can be considered as sums  $\sum_{i=1}^n v_i$  of discrete valuations  $v_i$ . Moreover, the valuations have positive correlation with  $S$ . The easiest way to evaluate them is to make  $n$ -item questionnaire, where  $v_i$  is scoring of the answer to the  $i$ -th question. A test is called dichotomous when  $v_i \in \{0, 1\}$  and the answer may only be right or wrong. A test is called polytomous when  $v_i \in \{0, 1, \dots, m_i - 1\}$  and the answer is measured on  $m_i$  scores scale.

In the papers (Krylovas & Kosareva, 2008; Kosareva & Krylovas, 2011) mathematical models for both dichotomous and polytomous tests were proposed. The spotlight of these works is construction of test item characteristic functions that have required properties. Though the tests were originally constructed for knowledge measurement, their mathematical models can be applied for various areas, such as economic problems, sustainable development, construction, human resources management, etc.

In the paper (Krylovas, Kosareva, & Navickienė, 2013) such IRT models are used to construct social indicators in a case study to evaluate economic development level of world countries. In (Krylovas & Kosareva, 2011) this methodology was applied to examine the Environmental Performance Index.

Item response theory (IRT) models are often applied in educational and psychological testing for evaluating the latent ability of examinees and the parameters of the test items that measure the latent ability. The paper of Toribio and Albert (2011) compares the effectiveness of several discrepancy measures, used within the posterior predictive model check procedure, in detecting misfitted test items. This method was applied to real data coming from a mathematics placement exam.

Although IRT is a traditional technique of educational and psychological measurement, it can be successfully used as an econometric model in the measurement of economic and

social constructs (Raileanu, 2008). Szeles and Fusco (2013) investigated the use of IRT models in the field of deprivation measurement.

Finn, Ben-Porath, and Tellegen (2015) examined the optimal number of response options for psychological questionnaires; they compared the dichotomous (true/false) response format test with a polytomous 4-choice format. The differences between correlations with external criteria were very rarely statistically significant.

Item response theory logistic models, with a particular focus on the one-parameter logistic model, or Rasch model, and their properties and assumptions, discussed in Noventa, Stefanutti, and Vidotto (2014) for both infinite and finite populations. An estimation of the finite population's size, at which the results of an infinite population can be applied, is given.

In the comprehensive study of De Champlain, Boulais, and Dallas (2016), different methods of calibrating multiple choice test item and clinical decision making components for the Medical Council of Canada's Qualifying Examination Part I based on test item response theory were compared. As a result, the dichotomous calibrations provided better fit of the test item response matrix than more complex polytomous calibrations.

Andreis and Ferrari (2014) analyzed multidimensional item response theory models for dichotomous data for evaluating customer satisfaction. Markov Chain Monte Carlo techniques were elaborated for estimation.

A new class of mixture IRT models was developed (Jin & Wang, 2014) to account for performance decline during testing behavior in dichotomous and polytomous test items, by adding a decrement parameter to each latent class.

Posterior predictive model checking for multidimensionality in IRT proposed by Levy, Mislevy, and Sinharay (2009). Procedures described for estimating single-administration classification consistency and accuracy indices for complex assessments using test item response theory comprising dichotomous and polytomous test items proposed by Lee (2010).

Weissman (2013) proposed a method of marginal maximum likelihood (MML) via the expectation-maximization (EM) algorithm, which utilized to estimate the test item parameters in IRT.

Wong (2015) proposed an approach to derive the asymptotic standard errors of item response theory true score equating involving polytomous test items, for equivalent and nonequivalent groups of examinees. Tests modeled using the generalized partial credit model or the graded response model were analyzed in this article.

The article of Kohli, Koran, and Henn (2015) is devoted to classical test theory (CTT) and IRT frameworks considered for the purpose of estimating person and test item parameters. Results of a small Monte Carlo study show the frameworks of IRT and CTT with underlying normal variable assumption to be quite comparable, with neither framework showing an advantage over the other.

The number of articles describe the software application for IRT problem solution. In the study of Huang (2015), various multilevel higher order test item response theory (ML-HIRT) models for simultaneously measuring growth in the second – and first-order latent traits of dichotomous and polytomous test items are proposed. A series of simulations conducted using the WinBUGS software with Markov chain Monte Carlo (MCMC) methods reveal that the parameters could be recovered satisfactorily and that latent trait estimation was reli-

able across measurement times. The article of Ames and Samonte (2015) intends to provide an accessible overview of Bayesian methods in the context of test item response theory to serve as a useful guide in estimating and interpreting test IRT models. A description of the estimation procedure used by SAS PROC MCMC is provided. The article of Paek and Han (2013) reviews a new test item response theory (IRT) model estimation program, IRTPRO 2.1, for Windows. Tay and Drasgow (2012) investigated the effectiveness of the mean adjusted chi(2)/df statistic and developed a new approach for assessing the goodness of fit of a test item response theory model. The study by Chen and Duh (2008) developed a personalized intelligent tutoring system based on the proposed fuzzy test item response theory (FIRT), which could be capable of recommending courseware with suitable difficulty levels for learners according to learner's uncertain/fuzzy feedback responses. To maximize the effectiveness of a decision an integrated approach for determining the weights of students' personal goals using multiple-criteria methods are proposed in Dadelo, Turskis, Zavadskas, Kačerauskas, and Dadelienė (2016). The entropy and the expert judgment methods were combined for determining the criteria weights.

In the article (Krylovas & Kosareva, 2008) authors proposed to describe dichotomous test items with nondecreasing functions  $k(p): [0,1] \rightarrow [0,1]$ , when  $p$  is knowledge or other property level measured by the test. Methodologically it is quite difficult to validate this assumption because this concept is complex and comprising many aspects. It is often treated as an imagined construct and researchers are limited to its measurement procedures. When measuring knowledge it is much easier to get the relative estimate provided by, for example, testee's raw test scores.

Using these relative knowledge assessments we can calculate an estimate of a test item solvability (percentage of correct responses to this test item). This article proposes to separate all test takers into three fuzzy subsets  $W$ ,  $A$ ,  $S$  (weak, average and strong students) and estimate solvability for each subset by a fuzzy number:  $I^W$ ,  $I^A$ ,  $I^S$  (see, for example, Zimmermann, 2001). These numbers (in this article triangular fuzzy numbers  $T_{r_W}, T_{r_A}, T_{r_S}$  are compared to each other, and on that basis classification of test items is performed.

All phenomena in the physical universe have a degree of uncertainty. Under real conditions, vague or imprecise information creates difficulties in assigning a crisp value of a subjective judgment. Lofti Zadeh (1965) introduced fuzzy sets theory. Fuzzy sets theory provides a mathematical framework for the precise and rigorous study of vague conceptual phenomena. Applications of fuzzy sets can be found in engineering, communications, computer science, expert systems, decision theory, operations research, supplier evaluation and selection, energy management (e.g. Ghorabae, Amiri, Zavadskas, & Antucheviciene, 2017; Mardani et al., 2017; Liu, Li, & Antuchevičienė, 2016). Human knowledge by its nature is both lexically imprecise and noncategorical. The knowledge representation by fuzzy set is a kind of intuitive behavior of human advanced thinking in essence. In our knowledge, there are only few scientific papers developing applications of fuzzy sets in IRT. The synergy of IRT and fuzzy set theory is applied in the new approach for generating fuzzy numbers to assess the competitiveness of the tourism industries in Asian countries (Huang & Peng, 2012). The applicability of the basic ideas of fuzzy set theory in the context of medical assessment questionnaires, which are commonly used, for instance, to support the diagnosis of psycho-

logical disorders, was analysed by Vetterlein and Zamansky (2016). The investigation of Lee, Wang, Lin, Yang, and Lin (2016) proposes an online self-learning platform for self-learners to improve their English learning for listening in southern Taiwan. Authors applied a genetic fuzzy markup language (GFML) to propose an item response theory agent for inferring possibility of correctly answering a specific item. A genetic learning mechanism is also adopted to improve the experimental results.

The main purposes of IRT are calibration of test items and evaluation of test takers according to some latent property. The proposed methodology is useful in the first stage of the assessment when a) there is no need to evaluate the test item parameters with sufficient accuracy but we rather want to classify test items for using them in the future (is the particular test item reasonable for weak, average or strong testees?) b) we are just going to classify the test takers rather than accurately assess their knowledge (is the testee weak, average or strong on that property?) The intention is to determine the group of testees for which the particular test item is eligible.

Let us also notice that fuzzy sets theory can be very suitable to evaluate a not very strictly described phenomena or in case of few statistical data. The proposed methodology allows comparing dichotomous items, which evaluate certain social phenomena, as well as selecting appropriate questions and constructing evaluation tests and indicators. This may be the advantage of the proposed methodology compared to the classical statistical procedures but a comprehensive comparison of the methodologies requires further research.

This article is organized as follows. In Section 1, comparison relationships of triangular fuzzy numbers and classification of the dichotomous test item differentiation property are defined. In Section 2, triangular fuzzy number construction method is described. In Section 3, fuzzy subsets  $W, A, S$  are determined by the trapezoidal membership functions. In Section 4, the complete classification procedure of the particular test item by its differentiation property is described. In Section 5, a practical case study from one particular advanced mathematics test is presented. Section 6 is devoted to conclusions and discussion on future research.

### 1. Order relations of the fuzzy triangular numbers

If  $S$  is a fuzzy set, then we will define the solvability of the test item by the fuzzy triangular number  $Tr(L, T, R)$  ( $0 \leq L \leq T \leq R \leq 100$ ). In general  $Tr(L, T, R)$  is a fuzzy number of the set  $x \in [0, 100]$  which has the triangular membership function (Zadeh, 1965):

$$\mu(x) = \begin{cases} \frac{x-L}{T-L}, & \text{for } x \in [L, T] \\ \frac{R-x}{R-T}, & \text{for } x \in (T, R] \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

It is clear that  $0 \leq \mu(x) \leq 1$ . In case it is known which students from the fuzzy set answered the test item correctly, the algorithm for constructing fuzzy triangles (the numbers  $L, T, R$ ) will be discussed in Section 2.

In the special case when  $S$  is a crisp set ( $L = T = R$ ), the solvability of the test item  $k(S)$  is defined as the percentage of correctly answered testees' number  $t(S)$ :

$$k(S) = \frac{t(S)}{|S|} \cdot 100(\%), \tag{2}$$

where  $S$  is the set of all testees and  $|S|$  is the number of elements of  $S$ .

Suppose that  $T_{r_1}(L_1, T_1, R_1)$  and  $T_{r_2}(L_2, T_2, R_2)$  are two triangular fuzzy numbers. We will denote

$$T_{r_1} \preceq T_{r_2}, \text{ when } L_1 \leq L_2 \ \& \ T_1 \leq T_2 \ \& \ R_1 \leq R_2. \tag{3}$$

Note that the relation  $\preceq$  (Zimmermann 2001) is reflexive ( $\forall T_r : T_r \preceq T_r$ ), antisymmetric (if  $T_{r_1} \preceq T_{r_2}$  and  $T_{r_2} \preceq T_{r_1}$ , then  $T_{r_1} = T_{r_2}$ ) and transitive (if  $T_{r_1} \preceq T_{r_2}$  and  $T_{r_2} \preceq T_{r_3}$ , then  $T_{r_1} \preceq T_{r_3}$ ). It means that the relation  $\preceq$  is a non-strict order relation (Anderson, 2004). The relation  $\preceq$  is not total, because there exist triangular numbers for which neither  $T_{r_1} \preceq T_{r_2}$ , nor  $T_{r_2} \preceq T_{r_1}$  is valid. Note that

$$T_{r_1} \prec T_{r_2}, \text{ when } R_1 < L_2. \tag{4}$$

Relation  $\prec$  is also transitive and antisymmetric, but irreflexive ( $\forall Tr: Tr \not\prec Tr$ ), so this is a strict order relation.

We consider three subsets of the tested set: weak students  $W$ , average –  $A$  and strong –  $S$ . For each subset we will define the fuzzy triangular number of the test item solvability  $T_{r_W}, T_{r_A}, T_{r_S}$ . In the Table 1 we suggested classification for the test items' differentiation property.

Table 1. Classification for the test items' differentiation property

The test item differentiates <b>all</b> students well, when:	$T_{r_W} \prec T_{r_A} \prec T_{r_S}$
The test item differentiates <b>strong</b> students well, when:	$T_{r_W} \preceq T_{r_A} \prec T_{r_S}$
The test item differentiates <b>weak</b> students well, when:	$T_{r_W} \prec T_{r_A} \preceq T_{r_S}$
The test item differentiates students <b>badly</b> , when:	$T_{r_W} \preceq T_{r_A} \preceq T_{r_S}$
The test item is <b>inappropriate</b> in all other cases.	

## 2. Procedure of making the triangular fuzzy numbers

Suppose, that  $A = \{a \in S : 0 \leq \mu_A(a) \leq 1\}$  is a fuzzy set (the fuzzy subset of the testees' set  $S$ ) and  $\mu_A(a)$  – its membership function. Also suppose that  $A$  is a normal set, i.e.  $\sup_{a \in S} \mu(a) = 1$ .

The  $\alpha$  – cuts ( $\alpha$  level sets) of the set  $A$  are called (crisp) sets  $A_\alpha = \{a \in S : \mu_A(a) \geq \alpha\}$ . From the fuzzy sets theory the following formula is known (Zimmermann, 2001):

$$A = \bigcup_{\alpha \geq 0} \alpha A_\alpha, \tag{5}$$

where  $\alpha A_\alpha = \{a \in A_\alpha : 0 \leq \mu_A(a) \leq 1\}$ . For example, if  $A = \{(1, 0.5), (2, 0.6), (3, 0.7), (4, 1.0), (5, 0.8)\}$ , then  $A_{0.5} = S = \{1, 2, 3, 4, 5\}$ ,  $A_{0.6} = \{2, 3, 4, 5\}$ ,  $A_{0.7} = \{3, 4, 5\}$ ,  $A_{0.8} = \{4, 5\}$ ,  $A_{1.0} = \{4\}$ ,

$$\begin{aligned}
 A = \bigcup_{\alpha \geq 0} \alpha A_\alpha = & \{(1,0.5), (2,0.5), (3,0.5), (4,0.5), (5,0.5)\} \cup \\
 & \{(2,0.6), (3,0.6), (4,0.6), (5,0.6)\} \cup \{(3,0.7), (4,0.7), (5,0.7)\} \cup \\
 & \{(4,0.8), (5,0.8)\} \cup \{(4,1.0)\} = \{(1,0.5), (2, \max\{0.5, 0.6\}), \\
 & (3, \max\{0.5, 0.6, 0.7\}), (4, \max\{0.5, 0.6, 0.7, 0.8, 1.0\}), (5, \max\{0.5, 0.6, 0.7, 0.8\})\} = A.
 \end{aligned}$$

We'll take the nonempty cuts  $A_{\alpha_1}, A_{\alpha_2}, \dots, A_{\alpha_n}$  of the set  $A$ . There is a finite number of such cuts, because  $S$  is a finite set. For each cut  $A_{\alpha_i}$  we calculate the values

$$p_{\alpha_i} = \frac{t_{\alpha_i}}{|A_{\alpha_i}|} \cdot 100(\%), \tag{6}$$

where  $t_{\alpha_i}$  – number of the testees from the set  $A_{\alpha_i}$  who correctly answered the test item,  $|A_{\alpha_i}|$  – number of elements of the set  $A_{\alpha_i}$  (we apply formula (2)). We get number  $T$  for the triangular fuzzy number  $Tr(L, T, R)$  when  $\alpha_i = 1.0 : T = p_{1.0}$ . Note that the set  $A$  is normal, thus  $A_{1.0} \neq \emptyset$  (is a nonempty set).  $A_{1.0}$  is the subset of all the testees from the set  $A$  who definitely belong to this set.

The values  $L$  and  $R$  are calculated by the least squares method (e.g. Bretscher, 1995). By  $\alpha_1^-, \alpha_2^-, \dots, \alpha_{n^-}^-$  such  $\alpha_i < 1$  values will be denoted for which  $p_{\alpha_i}^- < T$ , i.e.  $\alpha$ -cuts numbers for which the percentage of the correctly answered testees is lower than in the set  $A_{1.0}$ . Correspondingly denote  $\alpha_1^+, \alpha_2^+, \dots, \alpha_{n^+}^+$ , when  $p_{\alpha_i}^+ > T$ , i.e.  $\alpha$ -cuts numbers for which this percentage is higher than in  $A_{1.0}$ ;  $n^- + n^+ = n - 1$ .  $p_{\alpha_i}^-$  and  $p_{\alpha_i}^+$  are obtained from empirical data. Theoretical values  $p_{\alpha_i}$  assessed from formula (1). If  $\alpha_i \in [L, T]$ , we obtain  $p_{\alpha_i} = (T - L) \cdot \alpha_i + L$ , if  $\alpha_i \in [T, R]$ , then we have  $p_{\alpha_i} = R - (R - T) \cdot \alpha_i$ . Define the objective functions for the parameters  $L$  and  $R$  as sum of squares of differences between theoretical and empirical data:

$$f(L) = \sum_{i=1}^{n^-} (p_{\alpha_i}^- - (T - L) \cdot \alpha_i^- - L)^2, \quad g(R) = \sum_{i=1}^{n^+} (p_{\alpha_i}^+ - R + (R - T) \cdot \alpha_i^+)^2.$$

The functions  $f(L), g(R)$  are then the sums of squares of the values  $d_i^\pm$  shown in Figure 1. We minimize values of the functions  $f(L)$  and  $g(R)$

$$L = \underset{L}{\operatorname{argmin}} f(L); \quad R = \underset{R}{\operatorname{argmin}} g(R)$$

by calculating functions'  $f(L)$  and  $g(R)$  derivatives with respect to  $L$  and  $R$  respectively and solving the equations  $f'(L) = 0$  and  $g'(R) = 0$ :

$$L = \frac{\sum_{i=1}^{n^-} (p_{\alpha_i}^- - T \cdot \alpha_i^-) \cdot (1 - \alpha_i^-)}{\sum_{i=1}^{n^-} (1 - \alpha_i^-)^2}, \quad R = \frac{\sum_{i=1}^{n^+} (p_{\alpha_i}^+ - T \cdot \alpha_i^+) \cdot (1 - \alpha_i^+)}{\sum_{i=1}^{n^+} (1 - \alpha_i^+)^2}. \tag{7}$$

Example. Suppose that the fuzzy set of testees is:

$A = \{(a, 1.0), (b, 0.5), (c, 1.0), (d, 0.3), (e, 0.1), (f, 0.3)\}$ . Suppose that only  $a$  and  $b$  have answered the test item correctly. Then the set  $A$  has four different  $\alpha$ -cuts. The corresponding  $p_\alpha$  values from formula (6) are:

$$A_{1.0} = \{a, c\}, p_{1.0} = \frac{1}{2} \cdot 100 = 50;$$

$$A_{0.5} = \{a, b, c\}, p_{0.5} = \frac{2}{3} \cdot 100 = 67;$$

$$A_{0.3} = \{a, b, c, d, f\}, p_{0.3} = \frac{2}{5} \cdot 100 = 40;$$

$$A_{0.1} = \{a, b, c, d, e, f\}, p_{0.1} = \frac{2}{6} \cdot 100 = 33.$$

Thus  $n^- = 2$ ,  $\alpha_1^- = 0.1$ ,  $p_1^- = 33$ ,  $\alpha_2^- = 0.3$ ,  $p_2^- = 40$ ,  $n^+ = 1$ ,  $\alpha_1^+ = 0.5$ ,  $p_1^+ = 67$ . We find the parameters of the fuzzy triangular number  $Tr(L, T, R)$ :

$$T = 50,$$

$$L = \frac{(33 - 50 \cdot 0.1) \cdot (1.0 - 0.1) + (40 - 50 \cdot 0.3) \cdot (1.0 - 0.3)}{(1.0 - 0.1)^2 + (1.0 - 0.3)^2} = 32.85,$$

$$R = \frac{(67 - 50 \cdot 0.5) \cdot (1.0 - 0.5)}{(1.0 - 0.5)^2} = 84,$$

Finally, we obtain triangular fuzzy number describing this test item:  $(32.85, 50, 84)$ .

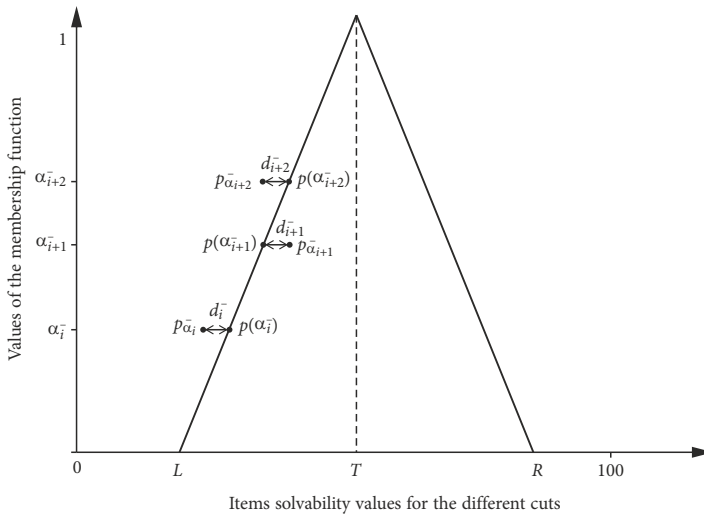


Figure 1. The values  $L$  are obtained by the least squares method minimizing the sum  $\sum_{i=1}^{n^-} (d_i^-)^2$  by  $L$ , where  $d_i^- = p_{\alpha_i^-} - p(\alpha_i^-)$  and  $p(\alpha_i^-) = (T - L) \cdot \alpha_i^- + L$ . Similarly, the values  $R$  are obtained



### 3. Procedure of finding fuzzy subsets of the testees set

The fuzzy subsets  $W, A, S$  (weak, average and strong students) of the testees set are defined by the trapezoidal membership functions

$$\mu_{(a,b,c,d)}(t) = \begin{cases} \frac{t-a}{b-a}, & \text{when } t \in [a, b] \\ \frac{d-t}{d-c}, & \text{when } t \in [c, d] \\ 1, & \text{when } t \in (b, c) \\ 0, & \text{otherwise,} \end{cases}$$

where  $a \leq b \leq c \leq d$ . Suppose that knowledge (or other considered property) of all tested students is rated by some scores  $b_1, b_2, \dots, b_n$ . Denote the lowest and the highest  $b_i$  values as *min* and *max*. Let us take four numbers

$$\min < \alpha < \beta < \gamma < \delta < \max \tag{8}$$

and let us define the membership functions of the subsets  $W, A, S$  as:

$$\mu_{(\min, \min, \alpha, t_1)}^W(t), \mu_{(t_2, \beta, \gamma, t_3)}^A(t), \mu_{(t_4, \delta, \max, \max)}^S(t). \tag{9}$$

Then for parameters  $t_i$  the following inequalities hold

$$\alpha \leq t_1 \leq t_2 \leq \beta; \gamma \leq t_3 \leq t_4 \leq \delta \tag{10}$$

and the meaning of the parameters  $\alpha, \beta, \gamma, \delta$  is the following: when  $b_i \leq \alpha$  then the student for sure is weak, when  $\beta \leq b_i \leq \gamma$  – average, when  $b_i \geq \delta$  – strong. Membership functions of the subsets  $W, A, S$  are depicted in Figure 2.

We construct subsets  $W, A, S$  many times by re-selecting different values for the parameters  $t_i$  which fulfill (10) restrictions. The number of possible  $t_1, t_2, t_3, t_4$  combinations is finite. From the obtained fuzzy triangular numbers  $Tr(L_i, T_i, R_i)$  we will make optimistic triangles (by taking averages of  $L_i, T_i, R_i$  when all  $t_i$  values have been reselected):

$$L_{opt} = \frac{1}{N} \sum_{i=1}^N L_i, T_{opt} = \frac{1}{N} \sum_{i=1}^N T_i, R_{opt} = \frac{1}{N} \sum_{i=1}^N R_i, \tag{11}$$

here  $N$  is the number of all fuzzy triangles. For the pessimistic triangles, we have wider intervals, when we take  $\min L, \max R$  and average of  $T$  values, when all  $t_i$  values have been

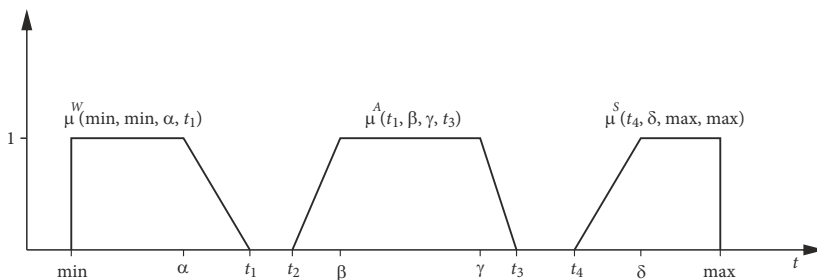


Figure 2. Membership functions of the weak, average and strong students

reselected:

$$L_{pes} = \min_i L_i, T_{pes} = \frac{1}{N} \sum_{i=1}^N T_i, R_{pes} = \max_i R_i. \tag{12}$$

Thus, while classifying the test items we pay attention not only to the average (optimistic), but also to the pessimistic fuzzy assessments. Note that the optimistic case always gives shorter intervals  $[L, R]$ , while pessimistic case – wider intervals.

#### 4. Procedure of establishing test items differentiation property

Next, we present procedure for classification of test items' differentiation property according to definitions proposed in Section 1. Note that the test is created from dichotomous items. Let us analyse one particular test item, our purpose is establishing its differentiation property.

1. Input: a) total test scores for each student, b) responses to this particular test item for each student (+ or -).
2. For the test item establish parameters  $\alpha, \beta, \gamma, \delta$ , which must satisfy the conditions (8).
3. Select value for parameter  $t_1$  satisfying the first condition in (10).
4. Calculate the membership function  $\mu_{(min, min, \alpha, t_1)}^W(t)$  for the subset  $W$  of weak students; calculate  $A_{1,0}$  and  $T = p_{1,0}$  by formula (6).
5. Construct all different  $\alpha$ -cuts and calculate corresponding  $p_\alpha$  values from the formula (6).
6. Calculate  $L$  and  $R$  values from formulas (7). Fix fuzzy triangle  $(L, T, R)$ .
7. Choose another value for the parameter  $t_1$  satisfying the first condition in (10). If such value exists, go to step 4, otherwise go to step 8.
8. Calculate optimistic and pessimistic fuzzy triangles for this test item for the subset of weak students  $W$  from all  $(L, T, R)$  values obtained in step 5 by formulas (11)–(12).
9. Repeat steps 3–8 for the subsets of average ( $A$ ) and strong ( $S$ ) students<sup>1</sup> and obtain optimistic and pessimistic fuzzy triangles for these subsets.
10. Output: draw optimistic and pessimistic fuzzy triangles for this test item for the  $W, A, S$  subsets, classify the test item's differentiation property.

Example. Suppose that we obtained total scores of 20 items test for 5 testees. Furthermore, we know that 3 of them responded correctly to the particular test item (+) and 2 answered incorrectly (-):  $s_1(2,+), s_2(4,+), s_3(3,-), s_4(3,-), s_5(5,+)$ <sup>2</sup>. Let us construct optimistic and pessimistic fuzzy triangles describing this test item. From the test results, we see that all testees are probably weak students. The membership function of the subset of weak students  $W$  is  $\mu_{(min, min, \alpha, t_1)}^W(t)$ , where  $min = 2, \alpha = 3$ . By changing  $\alpha$  and  $t_1$  values ( $3 \leq \alpha \leq t_1 \leq 5$ ) we get 6 different trapezoids (fuzzy set's membership functions):  $\mu_1^W = \mu_{(2,2,3,3)}^W(t), \mu_2^W = \mu_{(2,2,3,4)}^W(t), \mu_3^W = \mu_{(2,2,3,5)}^W(t), \mu_4^W = \mu_{(2,2,4,4)}^W(t), \mu_5^W = \mu_{(2,2,4,5)}^W(t), \mu_6^W = \mu_{(2,2,5,5)}^W(t)$ . For each trapezoid, construct a fuzzy triangle describing the test item.

<sup>1</sup> For the subset of average students, procedure reselects values for the parameters  $t_2$  and  $t_3$ , for the subset of strong students – values for the parameter  $t_4$ .

<sup>2</sup>  $s_i(j, z)$  where  $j$  is the total score and  $z$  is the indicator for the answer to this particular test item of the student  $i$ .

1.  $\alpha = t_1 = 3$ .  $A_1 = \{(s_1, 1.0), (s_3, 1.0), (s_4, 1.0)\}$ . Fuzzy set of weak students is a crisp set in this case, so  $L_1 = T_1 = R_1 = \frac{1}{3} \cdot 100 = 33$ .
2.  $\alpha = 3, t_1 = 4$ . Fuzzy set of testees is  $A_2 = \{(s_1, 1.0), (s_2, 0.0), (s_3, 1.0), (s_4, 1.0)\}$   
 $L_2 = T_2 = \frac{1}{3} \cdot 100 = 33$ . The set  $A_2$  has 2 different  $\alpha$ -cuts and corresponding  $p_\alpha$  values  
 $A_{0.0} = \{s_1, s_2, s_3, s_4\}$  and  $A_{1.0} = \{s_1, s_3, s_4\}$ .  $\alpha_1^+ = 0.0, p_{0.0}^+ = \frac{2}{4} \cdot 100 = 50$ .  
 $R_2 = \frac{(50 - 33 \cdot 0) \cdot (1 - 0)}{(1 - 0)^2} = 50$ .
3.  $\alpha = 3, t_1 = 5$ .  $A_3 = \{(s_1, 1.0), (s_2, 0.5), (s_3, 1.0), (s_4, 1.0), (s_5, 0.0)\}$ .  $L_3 = T_3 = 33$ . The set  $A_3$  has 3  $\alpha$ -cuts and corresponding  $p_\alpha$  values  $A_{0.0} = \{s_1, s_2, s_3, s_4, s_5\}$ ,  
 $A_{0.5} = \{s_1, s_2, s_3, s_4\}$  and  $A_{1.0} = \{s_1, s_3, s_4\}$ .  $\alpha_1^+ = 0.0, p_{0.0}^+ = 60, \alpha_2^+ = 0.5, p_{0.5}^+ = 50$ .  
 $R_3 = \frac{(50 - 33 \cdot 0.5) \cdot (1 - 0.5) + (60 - 33 \cdot 0.0) \cdot (1 - 0.0)}{(1 - 0.5)^2 + (1 - 0.0)^2} = 61.4$ .
4.  $\alpha = 4, t_1 = 4$ .  $A_4$  is also a crisp set,  $A_4 = \{(s_1, 1.0), (s_2, 1.0), (s_3, 1.0), (s_4, 1.0)\}$ .  
 $L_4 = T_4 = R_4 = \frac{2}{4} \cdot 100 = 50$ .
5.  $\alpha = 4, t_1 = 5$ .  $A_5 = \{(s_1, 1.0), (s_2, 1.0), (s_3, 1.0), (s_4, 1.0), (s_5, 0.0)\}$ . The set  $A_5$  has 2 different  $\alpha$ -cuts and corresponding  $p_\alpha$  values  $A_{0.0} = \{s_1, s_2, s_3, s_4, s_5\}$  and  
 $A_{1.0} = \{s_1, s_2, s_3, s_4\}$ .  $L_5 = T_5 = \frac{2}{4} \cdot 100 = 50$ .  $\alpha_1^+ = 0.0, p_{0.0}^+ = 60$ .  $R_5 = \frac{(60 - 33 \cdot 0)}{(1 - 0)^2} = 60$ .
6.  $\alpha = 5, t_1 = 5$ .  $A_6$  is a crisp set,  $A_6 = \{(s_1, 1.0), (s_2, 1.0), (s_3, 1.0), (s_4, 1.0), (s_5, 1.0)\}$ .  
 $L_6 = T_6 = R_6 = \frac{3}{5} \cdot 100 = 60$ .

So, we have 6 fuzzy triangles:

$$(33, 33, 33); (33, 33, 50); (33, 33, 61.4); (50, 50, 50); (50, 50, 60); (60, 60, 60).$$

Construct optimistic and pessimistic triangles according to formulas (11)–(12):  
 $(L_{opt}, T_{opt}, R_{opt}) = (43.17, 43.17, 52.4), (L_{pes}, T_{pes}, R_{pes}) = (33, 43.17, 61.4)$ .

### 5. Analysis of the test items of one particular test

In this section, illustrative examples of advanced mathematics test items for technical university students are presented. The test was given to 106 students from Vilnius Gediminas Technical University faculty of Civil Engineering. The test includes items on differentiation rules, limits and continuity, application to graphing, etc. Students' knowledge was evaluated by the results of the test consisting of 20 test items, each valued by one score. We present the fragment of obtained results in the Table 2.

In the first column, there are students' numbers, in the first row – test items' numbers. Test items considered in the article are shaded. In the last column, there are total numbers of correctly answered test items for each student – scores  $b_i$ . Full information on the considered test items is presented in the Appendix A. The results of each group given separately. There were 4 groups with 29, 26, 28 and 23 students correspondingly in each group.

Table 2. A part of the obtained results

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	+	+	+	-	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	6
2	+	+	+	-	-	+	+	+	+	-	+	-	+	+	+	-	-	-	-	+	12
3	+	+	-	+	+	+	+	+	+	-	+	+	+	+	-	-	-	-	-	-	12
4	+	+	+	+	+	+	+	+	-	+	-	+	+	-	-	-	-	+	-	+	13

Students' results are shown in the histogram (Figure 3). Parameters (8) were as follows:  $\min = 1, \max = 20, \alpha = 9, \beta = 12, \gamma = 14, \delta = 17$ . So, we have 24 weak students for which we are sure ( $1 \leq b_i \leq 9$ ); 30 – average ( $12 \leq b_i \leq 14$ ); 19 – strong ( $17 \leq b_i \leq 20$ ). For the rest of the students we cannot confidently state that they are strong, average or weak. There are  $33 = 106 - (24+30+19)$  such students. We varied with assigning them to one of the three subsets, changing parameters  $t_i$  in (10), constructing trapezoids (9), which numbers are correspondingly 10, 100 and 10.

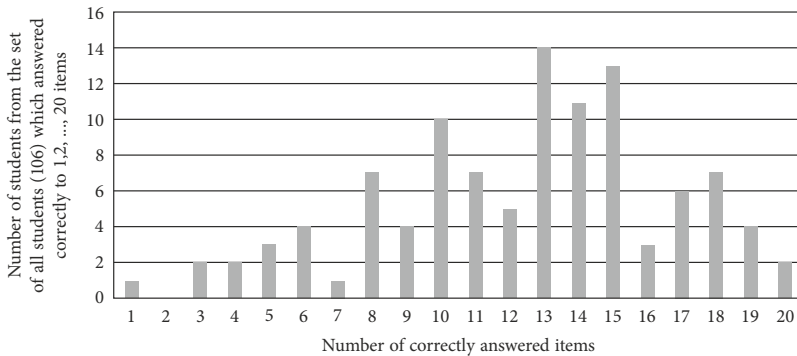


Figure 3. Histogram shows number of students who gave correct answers to 1, 2, ..., 20 test items

Let's examine item No. 6. Subset  $W$  of weak students can be constructed with 10 different values of  $t_1$ . Accordingly we obtain 10 fuzzy triangles  $(L_1, T_1, R_1), \dots, (L_{10}, T_{10}, R_{10})$ . The values of  $(L_{opt}, T_{opt}, R_{opt})$  for the optimistic case are calculated by formulas (11) and for the pessimistic case by formulas (12), where  $N = 10$ .

The example of the test item 6 is given in the Table 3.

Table 3. The example of test item 6

6	$\lim_{x \rightarrow 0} x^{14} \ln x =$	1) 14;	2) 0;
		3) $\infty$ ;	4) $\ln 14$ ;
		5) limit does not exist;	6) 1.

The solvability of the test item is 74% (i.e. 74% of all students correctly answered to that test item), and for each group the solvability is correspondingly 76, 65, 75 and 78. The fuzzy triangular numbers for weak, average and strong students  $T_{r_W}, T_{r_A}, T_{r_S}$  for separate groups and for all students are shown in the Figures 4–8.

Notice that only for one (2<sup>nd</sup>) group test item 6 badly differentiates students in the pessimistic case as pessimistic triangles intersect (relations  $T_{r_W} \preceq T_{r_A} \preceq T_{r_S}$ ). For all other groups separately and for all groups together the test item well differentiates weak students both in optimistic and pessimistic cases (relations  $T_{r_W} \prec T_{r_A} \preceq T_{r_S}$ ).

The other test items will be analyzed for all groups together. The example of the test item 10 proposed in the Table 4.

Table 4. The example of test item 10

10	Find $y'$ , if $y = (\sin x)^{x^7}$	<ol style="list-style-type: none"> <li>1) <math>x(\sin x)^{x^6} (\sin x \ln \sin x + x \cos x)</math>;</li> <li>2) <math>x^6 (\sin x)^{x^7-1} (7 \sin x \ln \sin x + x \cos x)</math>;</li> <li>3) <math>x^4 (\sin x)^{x^6-1} (\sin x \ln \sin x + x \cos x)</math>;</li> <li>4) <math>x^6 (\sin x)^{x^7-1} (7 \sin x \ln \sin x - x \cos x)</math>;</li> <li>5) <math>x^6 (\sin x)^{x^6} (6 \sin x \ln \sin x + x \cos x)</math>;</li> <li>6) <math>x(\sin x)^{x^7-1} (7 \sin x + x^5 \cos x)</math>.</li> </ol>
----	-------------------------------------	---

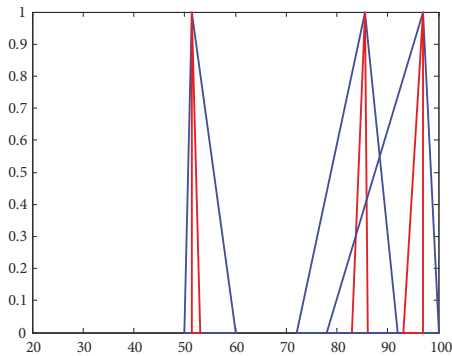


Figure 4. Fuzzy triangles – test item 6, group 1 (blue triangles – pessimistic, red – optimistic cases)

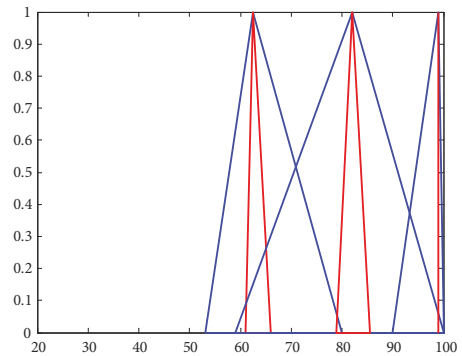


Figure 5. Fuzzy triangles – test item 6, group 2 (blue triangles – pessimistic, red – optimistic cases)

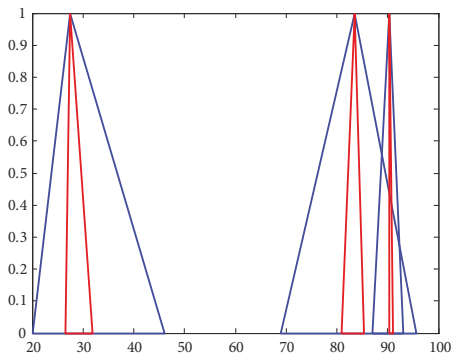


Figure 6. Fuzzy triangles – test item 6, group 3 (blue triangles – pessimistic, red – optimistic cases)

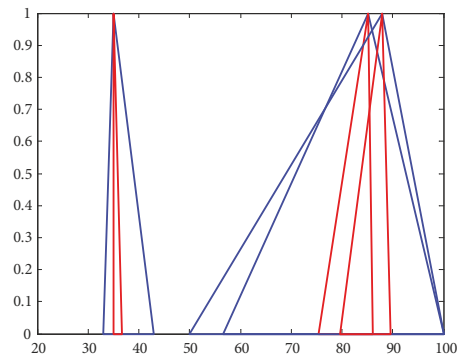


Figure 7. Fuzzy triangles – test item 6, group 4 (blue triangles – pessimistic, red – optimistic cases)

The solvability of the test item 10 is 57(%). From the Figure 9 we can see, that the test item 10, as the test item 6, well differentiates weak students, and in the optimistic case – all students

$$(T_{r_W} \prec T_{r_A} \prec T_{r_S}).$$

Test items 14 and 15 combined in one block and given as one in the Table 5.

Table 5. The example of test items 14 and 15

Given the curve $y = 2x^3 + 3x^2 - 2x - 4$ at the point $x = -1$ its:	
14	tangent line is: 1) $y = -\frac{1}{3}x - \frac{2}{3}$ ;    2) $y = -x$ ;    3) $y = -2x - 3$ ; 4) $y = 2x - 2$ ;    5) $y = -x - 2$ ;    6) $y = x$ .
15	normal line is: 1) $y = x - 2$ ;    2) $y = -x - 2$ ;    3) $y = 3x - 1$ ; 4) $y = 2x + 2$ ;    5) $y = \frac{1}{2}x - \frac{1}{2}$ ;    6) $y = x + 1$ .

The solvability of the test items 14 and 15 correspondingly is 53 and 30(%). From the Figures 10, 11 we can see, that the test item 14 well differentiates strong students in the optimistic case and the test item 15 – all students in both pessimistic and optimistic cases.

The next block of two test items 18 and 19 is given in the Table 6.

The solvability of the test items: 70 and 53(%). Test item 18 well differentiates weak students and in the optimistic case – all students (Figure 12), and the item 19 – all students in all cases (Figure 13).

Table 6. The example of test items 18 and 19

The asymptote of the function $y = \frac{-8x^2 + 7x}{15x - 27}$ as $x \rightarrow \infty$ is a line $y = lx + d$	
18	$l =$ 1) $-\frac{27}{7}$ ;    2) $-\frac{7}{27}$ ;    3) $\frac{7}{27}$ ;    4) $\frac{27}{7}$ ; 5) $-\frac{15}{8}$ ;    6) $-\frac{8}{15}$ ;    7) $\frac{15}{8}$ ;    8) $\frac{8}{15}$ .
19	$d =$ 1) $-\frac{75}{37}$ ;    2) $\frac{248}{209}$ ;    3) $\frac{75}{37}$ ;    4) $\frac{209}{248}$ ; 5) $-\frac{248}{209}$ ;    6) $\frac{37}{75}$ ;    7) $-\frac{37}{75}$ ;    8) $-\frac{209}{248}$ .

The example of the last test item 20 given in the Table 7. This item has the solvability 32(%) and well differentiates all students (Figure 14).

Table 7. The example of test item 20

20	$\lim_{z \rightarrow 2} \frac{z^2 + 5z - 14}{\operatorname{tg}(17z - 34)} =$ <p>1) <math>\frac{9}{17}</math>; 2) <math>\frac{25}{17}</math>; 3) <math>-\frac{9}{17}</math>; 4) <math>-\frac{25}{17}</math>; 5) <math>-\frac{90}{17}</math>; 6) 0; 7) <math>\infty</math>.</p>
----	---

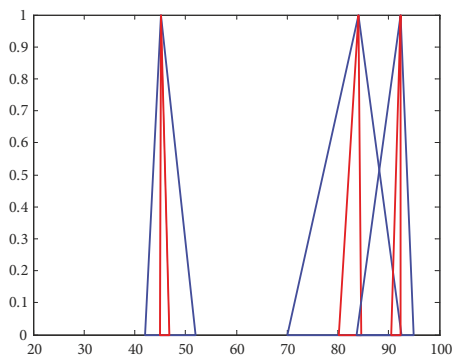


Figure 8. Fuzzy triangles – test item 6, all groups (blue triangles – pessimistic, red – optimistic cases)

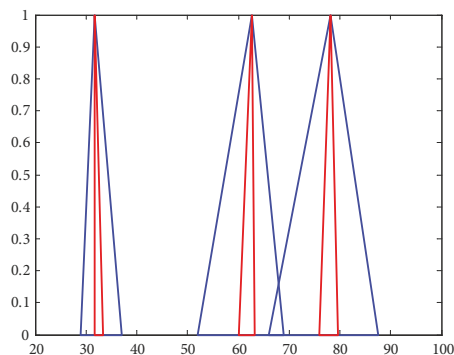


Figure 9. Fuzzy triangles – test item 10, all groups (blue triangles – pessimistic, red – optimistic cases)

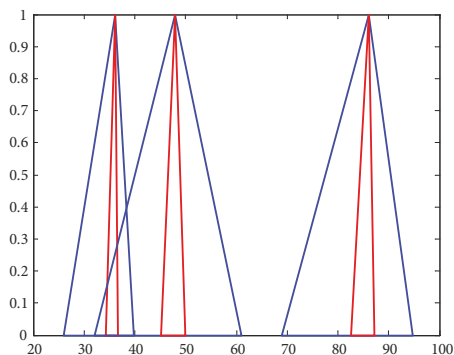


Figure 10. Fuzzy triangles – test item 14, all groups (blue triangles – pessimistic, red – optimistic cases)

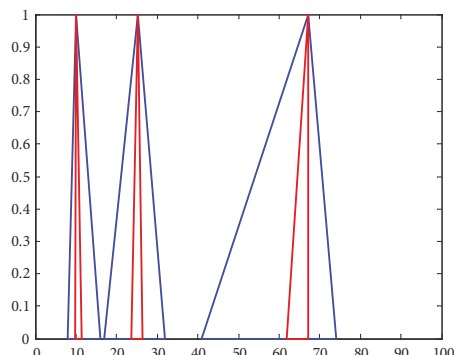


Figure 11. Fuzzy triangles – test item 15, all groups (blue triangles – pessimistic, red – optimistic cases)

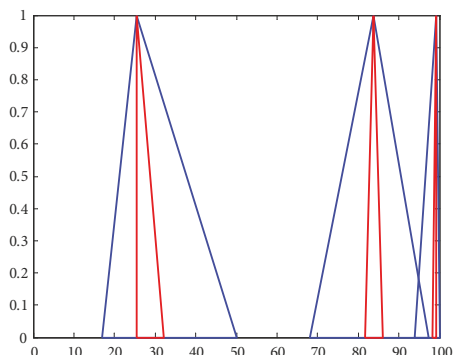


Figure 12. Fuzzy triangles – test item 18, all groups (blue triangles – pessimistic, red – optimistic cases)

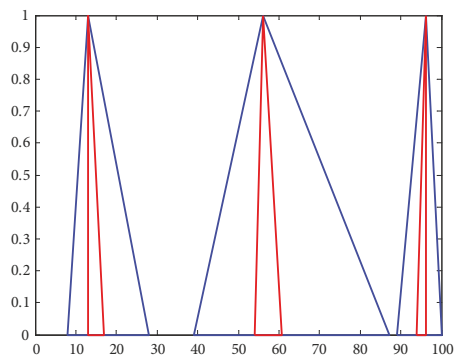


Figure 13. Fuzzy triangles – test item 19, all groups (blue triangles – pessimistic, red – optimistic cases)

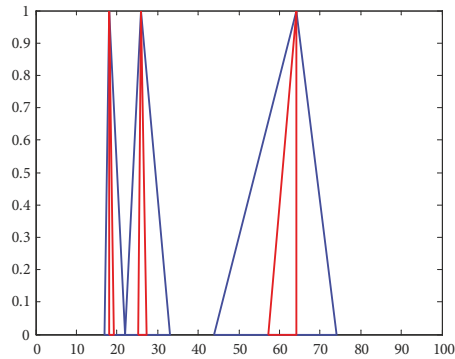


Figure 14. Fuzzy triangles – test item 20, all groups (blue triangles – pessimistic, red – optimistic cases)

## Conclusions and future research

The purpose of the classical approach is obtaining test item's parameter estimates and testees' assessments using a calibration procedure, marginal maximum likelihood and other methods. The purpose of this article is to classify test items in order to find out for which group of students they are most suitable. Therefore, the test constructed from that items will be optimal for respective group of students. The proposed methodology is useful for establishing the test item differentiation property. Considered method does not require strict evaluation of testees' knowledge. It is enough to have only relative achievement scores, for example raw test scores. This methodology allows us to get a stable test item classification for not big (20–30 students) groups of testees, however for deriving statistically reliable conclusions we need to do more detailed research, for example, carry out Monte Carlo type experiments. This will be our future research object. It is worth mentioning that in this survey all test items are considered independent from one another, though test items blocks (for example 14–15, 18–19) essentially make one unit and it would be correct to study them together (Krylovas & Kosareva, 2011), but such survey aspects are non trivial and have not been investigated enough.

The advantage of the applying this particular approach is reduction of the measurement error. According to item response theory the value of test information function is the reciprocal of the standard error of estimation, so the more information we have at a given ability level, the less is the error of measurement. Tests based on perfectly fitted items provide more information. Therefore, it makes sense to elaborate such tests adapted for each group of test takers.

## References

- Ames, A. J., & Samonte, K. (2015). Using SAS PROC MCMC for Item Response Theory Models. *Educational and Psychological Measurement*, 75(4), 585-609. <https://doi.org/10.1177/0013164414551411>
- Anderson, J. A. (2004). *Discrete mathematics mathematics with combinatorics* (2<sup>nd</sup> ed.). University of South Carolina – Spartanburg, Prentice Hall, New Jersey.



- Andreis, F., & Ferrari, P. A. (2014). Multidimensional item response theory models for dichotomous data in customer satisfaction evaluation. *Journal of Applied Statistics*, 41(9), 2044-2055. <https://doi.org/10.1080/02664763.2014.907395>
- Bretschler, O. (1995). *Linear algebra with applications* (3<sup>rd</sup> ed.). Upper Saddle River, NJ: Prentice Hall.
- Chen, C.-M., & Duh, L.-J. (2008). Personalized web-based tutoring system based on fuzzy item response theory. *Expert Systems with Applications*, 34(4), 2298-2315. <https://doi.org/10.1016/j.eswa.2007.03.010>
- Dadelo, S., Turskis, Z., Zavadskas, E. K., Kačerauskas, T., & Dadelienė, R. (2016). Is the evaluation of the students' values possible? An integrated approach to determining the weights of students' personal goals using multiple-criteria methods. *EURASIA Journal of Mathematics, Science and Technology Education*, 12(11), 2771-2781. <https://doi.org/10.12973/eurasia.2016.02303a>
- De Champlain, A. F., Boulais, A.-P., & Dallas, A. (2016). Calibrating the Medical Council of Canada's Qualifying Examination Part I using an integrated item response theory framework: a comparison of models and designs. *Journal of Educational Evaluation for Health Professions*, 13(6). <https://doi.org/10.3352/jeehp.2016.13.6>
- Finn, J. A., Ben-Porath, Y. S., & Tellegen, A. (2015). Dichotomous Versus Polytomous response options in psychopathology assessment: Method or meaningful variance?. *Psychological Assessment*, 27(1), 184-193. <https://doi.org/10.1037/pas0000044>
- Ghorabae, M. K., Amiri, M., Zavadskas, E. K., & Antucheviciene, J. (2017). Supplier evaluation and selection in fuzzy environments: A review of MADM approaches. *Economic Research-Ekonomiska Istraživanja* 30(1), 1073-1118. <https://doi.org/10.1080/1331677X.2017.1314828>
- Hambleton, R. K., Swaminathan, H., & Rogers, J. H. (1991). *Fundamentals of item response theory*. New York: Sage Publications.
- Huang, H.-Y. (2015). A Multilevel higher order item response theory model for measuring latent growth in longitudinal data. *Applied Psychological Measurement*, 39(5), 362-372. <https://doi.org/10.1177/0146621614568112>
- Huang, J.-H., & Peng, K.-H. (2012). Fuzzy Rasch model in TOPSIS: A new approach for generating fuzzy numbers to assess the competitiveness of the tourism industries in Asian countries. *Tourism Management* 33(2), 456-465. <https://doi.org/10.1016/j.tourman.2011.05.006>
- Jin, K.-Y., & Wang, W.-C. (2014). Item response theory models for performance decline during testing. *Journal of Educational Measurement*, 51(2), 178-200. <https://doi.org/10.1111/jedm.12041>
- Kohli, N., Koran, J., & Henn, L. (2015). Relationships among classical test theory and item response theory frameworks via factor analytic models. *Educational and Psychological Measurement*, 75(3), 389-405. <https://doi.org/10.1177/0013164414559071>
- Kosareva, N., & Krylovas, A. (2011). A numerical experiment on mathematical model of forecasting the results of knowledge testing. *Technological and Economic Development of Economy*, 17(1), 42-61. <https://doi.org/10.3846/13928619.2011.553994>
- Krylovas, A., & Kosareva, N. (2008). Mathematical modelling of forecasting the results of knowledge testing. *Technological and Economic Development of Economy*, 14(3), 388-401. <https://doi.org/10.3846/1392-8619.2008.14.388-401>
- Krylovas, A., & Kosareva, N. (2011). Item response theory applications for social phenomena modeling. *Societal Studies*, 3(1), 77-93.
- Krylovas, A., Kosareva, N., & Navickienė, O. (2013). Economic and social phenomena indicators design Methodology based on averaging values of dichotomous operators. In *3<sup>rd</sup> International Scientific Conference Whither Our Economies*, 24–25 October 2013, Vilnius, Lithuania.
- Lee, C.-S., Wang, M.-H., Lin, K.-H., Yang, S.-C., & Lin, T.-T. (2016). Genetic Fuzzy markup language-based item response theory agent for online self-learning platform construction. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 24–29 July 2016, Vancouver, Canada. <https://doi.org/10.1109/FUZZ-IEEE.2016.7737806>

- Lee, W.-C. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement*, 47(1), 1-17.  
<https://doi.org/10.1111/j.1745-3984.2009.00096.x>
- Levy, R., Mislevy, R. J., & Sinharay, S. (2009). Posterior predictive model checking for multidimensionality in item response theory. *Applied Psychological Measurement*, 33(7), 519-537.  
<https://doi.org/10.1177/0146621608329504>
- Liu, P., Li, Y., & Antuchevičienė, J. (2016). Multi-criteria decision-making method based on intuitionistic trapezoidal fuzzy prioritised OWA operator. *Technological and Economic Development of Economy*, 22(3), 453-469. <https://doi.org/10.3846/20294913.2016.1171262>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley Publishing Company.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Mardani, A., Zavadskas, E. K., Khalifah, Z., Zakuan, N., Jusoh, A., Md Nor, K., & Khoshnoudic, M. (2017). A review of multi-criteria decision-making applications to solve energy management problems: Two decades from 1995 to 2015. *Renewable and Sustainable Energy Reviews* 71: 216-256.  
<https://doi.org/10.1016/j.rser.2016.12.053>
- Noventa, S., Stefanutti, L., & Vidotto, G. (2014). An analysis of item response theory and Rasch models based on the most probable distribution method. *Psychometrica*, 79(3), 377-402.  
<https://doi.org/10.1007/s11336-013-9348-y>
- Paek, I., & Han, K. T. (2013). IRTPRO 2.1 for Windows (Item response theory for patient-reported outcomes). *Applied Psychological Measurement*, 37(3), 242-252.  
<https://doi.org/10.1177/0146621612468223>
- Raileanu, M. S. (2008). Exploring the extension of item response theory models to the economic and social measurement. *Proceedings of the 12<sup>th</sup> WSEAS International Conference on COMPUTERS*, 23-25 July 2008, Heraklion, Greece.
- Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research. Expanded edition: 1980, Chicago: The University of Chicago Press. 199 p.
- Szeles, M. R., & Fusco, A. (2013). Item response theory and the measurement of deprivation: evidence from Luxembourg data. *Quality & Quantity*, 47(3), 1545-1560.  
<https://doi.org/10.1007/s11135-011-9607-x>
- Tay, L., & Drasgow, F. (2012). Adjusting the Adjusted chi(2)/df Ratio Statistic for Dichotomous Item Response Theory Analyses: Does the Model Fit?. *Educational and Psychological Measurement*, 72(3), 510-527. <https://doi.org/10.1177/00131644111416976>
- Toribio, S. G., & Albert, J. H. (2011). Discrepancy measures for item fit analysis in item response theory. *Journal of Statistical Computation and Simulation*, 81(10): 1345-1360.  
<https://doi.org/10.1080/00949655.2010.4851311>
- Vetterlein, T., & Zamansky, A. (2016). Reasoning with graded information: The case of diagnostic rating scales in healthcare. *Fuzzy Sets and Systems*, 298, 207-221. <https://doi.org/10.1016/j.fss.2015.11.002>
- Weissman, A. (2013). Optimizing information using the EM algorithm in item response theory. *Annals of Operations Research*, 206(1), 627-646. <https://doi.org/10.1007/s10479-012-1204-4>
- Wong, C. C. (2015). Asymptotic standard errors for item response theory true score equating of polytomous items. *Journal of Educational Measurement*, 52(1), 106-120.  
<https://doi.org/10.1111/jedm.12065>
- Zadeh, L. A. (1965) Fuzzy sets. *Information and Control* 8(3), 338-353.  
[https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zimmermann, H. J. (2001). *Fuzzy set theory – and its applications* (4<sup>th</sup> ed.). Kluwer Academic Publisher. <https://doi.org/10.1007/978-94-010-0646-0>

APPENDIX A

Table 8. Test results for the group 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1						+				+				+	-			+	+	+	18
2						-				+				-	-			+	+	-	9
3						-				-				-	-			-	-	-	5
4						-				+				-	-			+	+	-	15
5						+				+				-	-			+	-	-	11
6						+				+				+	+			+	-	-	13
7						-				-				+	-			+	+	-	12
8						+				+				+	+			-	-	-	5
9						+				+				-	-			-	-	-	13
10						-				-				+	-			-	-	-	10
11						+				+				-	-			+	+	-	14
12						+				+				+ -	+			+	+	+	20
13						+				+				-	-			+	+	-	13
14						+				+				+	-			+	+	-	16
15						+				-				+	-			+	+	+	17
16						+				-				-	-			-	-	-	8
17						+				-				-	+			+	+	+	13
18						+				-				-	-			-	-	-	10
19						+				-				-	-			-	-	+	9
20						-				+				-	-			-	-	-	8
21						+				-				+	+			+	+	-	15
22						+				+				-	-			+	-	-	12
23						+				+				-	-			+	-	-	13
24						+				-				+	-			-	-	+	8
25						+				+				-	-			+	+	+	15
26						-				-				-	-			-	-	-	6
27						+				+				+	-			+	+	-	15
28						+				-				-	-			+	+	+	14
29						+				+				-	-			+	+	+	17

Table 9. Test results for the group 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1						+				+				+	+			+	+	-	15
2						+				-				-	-			+	+	-	15
3						+				-				-	-			-	-	+	7
4						+				-				-	-			-	-	-	10
5						-				-				-	-			-	-	-	3
6						+				+				-	-			-	-	-	6
7						+				+				-	-			-	-	-	10
8						-				-				-	-			+	-	+	11
9						+				+				+	+			+	+	+	15
10						+				-				+	-			+	+	-	14
11						-				-				+	-			+	+	+	14
12						+				+				-	-			+	-	-	12
13						+				+				-	-			+	+	-	13
14						+				-				+	+			-	-	+	10
15						+				+				+	+			+	+	+	18
16						+				+				+	+			-	-	-	13
17						+				-				+	-			+	+	-	15
18						+				+				+	-			-	-	-	9
19						-				+				-	-			+	-	+	10
20						-				+				-	-			-	-	-	5
21						+				+				+	-			+	-	-	14
22						+				+				+	+			+	+	+	20
23						+				+				+	-			-	+	-	15

Table 10. Test results for the group 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1						-				-				+	-			-	-	-	6
2						+				-				+	+			-	-	+	12
3						+				-				+	-			-	-	-	13
4						+				-				-	-			+	-	+	13
5						-				+				-	-			+	-	-	14
6						+				+				-	-			+	+	+	15
7						+				+				+	-			+	+	+	15
8						+				-				+	-			+	+	+	17
9						+				+				+	+			+	+	-	18
10						+				+				+	+			+	+	+	18
11						-				+				+	+			+	+	+	19
12						+				+				+	+			-	-	+	13
13						+				+				+	+			+	+	+	18
14						+				+				+	+			+	+	-	17
15						+				-				+	+			+	+	-	13
16						-				-				-	-			+	+	-	11
17						+				+				+	+			+	+	-	18
18						+				+				-	-			+	+	-	14
19						+				+				-	-			+	+	-	14
20						+				+				+	+			+	-	-	16
21						-				-				-	-			-	-	-	3
22						-				-				+	-			-	-	-	1
23						+				+				-	-			+	-	-	10
24						+				-				+	+			+	+	+	19
25						+				+				+	+			+	+	+	19
26						-				-				-	-			+	+	-	17
27						-				-				-	-			-	-	-	9
28						+				+				+	+			+	-	+	4

Table 11. Test results for the group 4

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1						+				+				+	-			+	-	-	14
2						+				-				-	-			+	-	-	11
3						+				+				-	-			+	+	+	11
4						+				-				+	-			+	+	-	14
5						+				+				-	-			+	-	-	15
6						-				+				-	-			+	+	-	11
7						+				+				-	-			+	-	-	13
8						-				-				-	-			-	-	+	6
9						+				+				+	-			+	-	+	13
10						+				-				+	-			-	-	-	4
11						+				+				+	+			+	+	+	19
12						-				-				+	-			+	-	-	8
13						+				+				+	+			+	+	-	18
14						+				-				+	+			+	+	+	17
15						+				-				+	+			+	+	+	15
16						-				-				+	+			+	+	-	10
17						-				+				+	+			+	+	-	16
18						-				-				-	-			+	+	-	11
19						-				-				+	-			+	+	+	8
20						+				+				+	+			+	+	-	12
21						+				-				-	-			-	-	-	8
22						+				-				+	+			+	+	+	13
23						+				+				-	-			+	+	-	14
24						+				+				-	-			-	-	-	10
25						-				-				-	-			-	-	+	8
26						-				-				+	-			+	+	+	10