

# Rankings of Students Based on Experts' Assessment of Levels of Verification of Learning Outcomes by Test Items

Aleksandra Mreła<sup>1</sup> and Oleksandr Sokolov<sup>2</sup>

<sup>1</sup> Kujawy and Pomorze University in Bydgoszcz, Technical Department, Toruńska str. 55-57, Bydgoszcz 85-023, Poland  
a.mrela@kpsw.edu.pl

<sup>2</sup> Nicolaus Copernicus University in Toruń, Faculty of Physics, Astronomy and Informatics, Grudziądzka str. 5, 87-100 Toruń, Poland  
osokolov@is.umk.pl

**Abstract.** The paper presents methods of preparing students' rankings based on the results of final secondary school examination test in mathematics in Poland. The currently used method is based on the percentage of earned points and does not take into account levels of acquirement of learning outcomes by students. The data used in this article contains results of students who earned the same number of points. Presented methods of preparing rankings are based on the experts' assessment of levels of verification of learning outcomes by items and methods of fuzzification of this assessment. According to the applied method the rankings show some difference.

**Keywords:** Ranking of students, Fuzzification, Quality assurance.

## 1 Introduction

Development of information technologies has contributed significantly to the teaching methods and students' assessment. Tutorial programs, interactive tests, different monitoring studies and state programs for automated evaluation of knowledge and skills have been introduced recently.

Testing has been applied widely in distance education and during implementation of the Bologna Process for student's self-education. Automated testing application has been expanded to the manufacturing, where personnel management is transformed into a continuous process of training (of course, with the subsequent testing and assessment of trainees). The distinct feature of such systems is that the role of a teacher in the process of learning and assessment is much narrower, and the results have been evaluated automatically which has been caused by the requirement for simultaneous estimation of a large number of trainees, and by the ideology of automated learning itself –

self-consistent learning and independent evaluation. One of the major tasks is the comparability of the results of different tests, ranging of students level of knowledge, formation of the final scoring for the test sets. Use of so-called "raw" scores, i.e., totals for the successful implementation of items resulting from the test might be applied to the very limited extent (if testing is limited to the identifying of the level of knowledge on particular topic and cannot be integrated with other results). Effectiveness of the test score depends not only on the quality of the test, but also on the methods of comparison and interpretation of primary (raw) score of test group.

Therefore one can assume as important to analyze the existing methods of comparison and integration of scores of various tests, study the quality of the students' group assessment, understanding the diversity of evaluation points as a quality criterion for estimating methods.

In 1999 the Polish Government decided to take part in the Bologna Process [1] and because of that all Polish educational institution at the beginning of the process of designing curricula define learning outcomes which students have to acquire during their studies and their teachers have to verify their acquirement by the students.

The mathematics curriculum in Polish secondary schools was approved by the Polish Ministry of Education [2]. This curriculum considers 5 learning outcomes which students are taught during three-year studies. At the end of the studies students take the written final secondary school examination.

The examination is divided into two parts, the first one is consisted of 25 items of multiple choice for which students can earn 1 point for answering correctly and the second part is consisted of 9 items for solving which students can earn more than 1 point. In this paper the data, we use to discuss the methods of preparing rankings, refers only to results of the first part of this examination.

The data comprised with results of the final secondary school examination in mathematics in one of Polish secondary schools in 2015. This examination was written by 149 students. For solving the first part of this examination students can earn up to 25 points. This paper presents discussion on the group of 18 students who earned the number of points equal to 18, so they all achieved the same position in the ranking of students based on the number of earned points or the average mean.

Sometimes we might encounter situations when there is a need to distinguish between these students, for example we would like to choose 5 out of these 18 students. So this ranking does not help us choose 5 better students unless we decide to apply more criteria.

In this paper we will discuss methods of preparing rankings [6,7] of students taking into consideration the results of the test in mathematics and experts' assessment of levels of verification of learning outcomes by test items and we discuss the results of 18 students who answering the first 25 items of the final secondary school examination in mathematics earned total score 18 points in one of Polish high schools.

In order to calculate levels of acquiring learning outcomes by students we will use the theory of fuzzy sets which was introduced by L.A. Zadeh [12] in 1965. In 1975 he generalized the concept of type 1 fuzzy sets and introduced type 2 fuzzy sets [13].

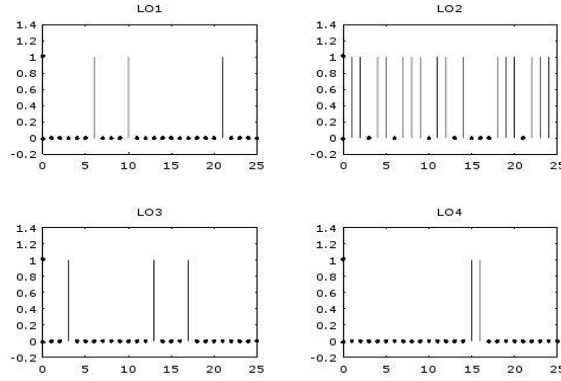
## 2 Crisp Relations

To build the relations between learning outcomes and items we will use the description of four learning outcomes LO1 – LO4 written in [2]:

- LO1 – Student interprets mathematical texts. After solving the tasks, student interprets the achieved result.
- LO2 – Student uses simple, well known mathematical objects.
- LO3 – Student chooses a mathematical object to the simple situation and estimates the pertinence of model critically.
- LO4 – Student applies strategy which results clearly from the content of the task.

On the basis of principles of assessment published in [3], where for each learning outcome the experts indicate items prepared for verifying its acquirement by students, the crisp relation  $R_1$  between learning outcomes and items was prepared, so the value of this relation  $R_1(j,k)$  for learning outcome  $j$ , where  $j = 1, \dots, 4$ , and item  $k$ , where  $k = 1, \dots, 25$ , is equal to 1 if the experts decided that item  $k$  can verify the acquirement of learning outcome  $j$ , or it is equal to 0, otherwise. Note that  $LO_i \cap LO_j = 0$  for each  $i \neq j$  and each item.

Figure 1 presents the membership functions of the relation between learning outcomes and items (bars show the value equal to 1).



**Fig. 1.** The membership functions of the relation between learning outcomes LO1-LO4 and items 1-25.

The value  $R_2(k, i)$  of relation  $R_2$  between item  $k$ , where  $k = 1, \dots, 25$ , and student  $i$ , where  $i = 1, \dots, 149$ , is equal to 1 if the student answered this item correctly or it is equal to 0, otherwise.

To calculate values of relation  $R_3$  between learning outcomes and students S-T-composition is used.

Let us recall the definition of S-T composition [11]. Let  $R = \{(x, y), \mu_R(x, y)\}$  and  $P = \{(y, z), \mu_P(y, z)\}$  be two relations with the membership functions  $\mu_R$  and  $\mu_P$ . The S-T composition of these relations  $R \subset X \times Y$  and  $P \subset Y \times Z$  is a relation  $R \circ P \subset X \times Z$  with the membership function defined as follows:

$$\mu_{R \circ P}(x, z) = S_{y \in Y}(T(\mu_R(x, y), \mu_P(y, z))) .$$

In [9] there is shown that for educational purposes the algebraic T-norm and S-norm are better than the most popular T-norm minimum and S-norm maximum, so we will use S-T composition with algebraic T-norm and S-norm.

**Table 1.** Levels of acquirement of learning outcomes by students.

Students	Learning outcomes			
	LO1	LO2	LO3	LO4
S1 – S9; S11 – S18	1	1	1	1
S10	1	1	1	0

In this case all these students acquired learning outcomes LO1 – LO4 on the level 1 except student S10 whose level of acquirement of leaning outcome LO4 was 0. Thus we cannot distinguish between these students with the exception of S10 and we cannot prepare the ranking.

### 3 Item Response Theory

According to the Item Response Theory (IRT) mathematical ability (understanding mathematics and skills in solving tasks) is a latent trait which cannot be measured. This theory describes the method of measuring this ability on the basis of results of the test which was solved by the group of students [4, 8, 10].

According to the algorithm [10], the mathematical abilities described in learning outcomes LO1-LO4 (all together) and then in these learning outcomes taken into account separately were calculated and put to Table 2.

**Table 2.** Levels of mathematical abilities according to IRT.

Students	Learning outcomes				
	LO1-LO4	LO1	LO2	LO3	LO4
S1	3.15	2.64	1.01	0.78	0
S2	-1.15	2.64	0.66	0.78	1.83
S3	3.15	-0.89	1.41	2.35	0
S4	0.96	0.87	1.41	-0.75	1.83
S5	3.15	0.87	1.01	0.78	1.83
S6	0.96	2.64	1.41	-0.75	0
S7	3.15	-0.89	1.9	0.78	0
S8	0.96	0.87	0.66	2.35	1.83
S9	-1.15	2.64	1.01	-0.75	1.83
S10	0.96	2.64	1.41	0.78	-1.82
S11	3.15	0.87	1.9	-0.75	0
S12	3.15	-0.89	1.41	2.35	0
S13	0.96	-0.89	2.56	-0.75	0
S14	0.96	0.87	1.41	-0.75	1.83
S15	0.96	2.64	1.01	-0.75	1.83
S16	0.96	0.87	1.41	0.78	0
S17	3.15	2.64	0.66	0.78	1.83
S18	0.96	2.64	1.41	-0.75	0

On the basis of these values, the rankings of students were prepared and they are presented in Table 3.

**Table 3.** The rankings of students.

Position	The basis for the ranking				
	LO1-LO4	LO1	LO2	LO3	LO4
1	S1, S3, S5, S7, S11, S12, S17	S1, S2, S6, S9, S10, S15, S17, S18	S13	S3, S8, S12	S2, S4, S5, S8, S9, S14, S15, S17
2	S4, S6, S8, S10, S13, S14, S15, S16, S18	S4, S5, S8, S11, S14, S16	S7, S11	S1, S2, S5, S7, S10, S16, S17	S1, S3, S6, S7, S11, S12, S13, S16, S18
3	S2, S9	S3, S7, S12, S13	S3, S4, S6, S10, S12, S14, S16, S18	S4, S6, S9, S11, S13, 14, S15, S18	S10
4	_____	_____	S1, S5, S9, S15	_____	_____
5	_____	_____	S2, S8, S17	_____	_____

The main problem with levels of mathematical abilities is too few values (too many students achieved the same level of this ability) in order to prepare rankings and differentiate students. For example, when we take into account all learning outcomes (LO1-LO4) there are only 3 positions in the ranking, on the first position there are 7 students (S1, S3, S5, S7, S11, S12, S17), on the second one there are 9 students (S4, S6, S8, S10, S13, S14, S15, S16, S18) and on the third one there are 2 students (S2, S9).

The similar situation we encounter if we take as the basis of the ranking learning outcomes LO1, LO3 and LO4 (there are only 3 positions) and only for learning outcome LO2 there are 6 positions.

If we have more requirements according to the ranking, e.g. assume that learning outcome LO1 is the most important, then LO2, LO3 and the least important learning outcome is LO4, then we can distinguish students and using the lexicographical order we can prepare the ranking of the students presented in Table 4.

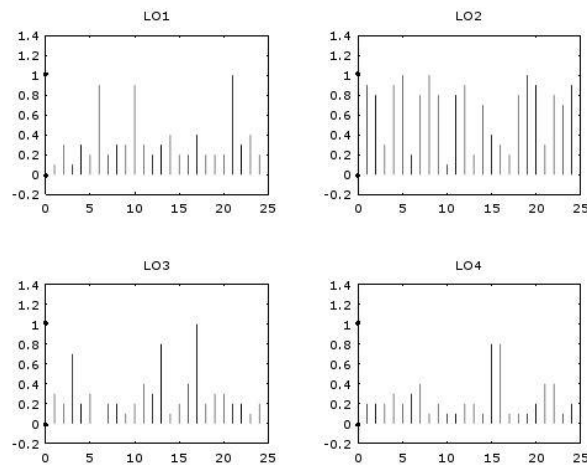
**Table 4.** Rankings of students.

Position	Student	Position	Student
1	S10	8	S4, S14
2	S6, S18	9	S5
3	S1	10	S8
4	S9, S15	11	S13
5	S2, S17	12	S7
6	S11	13	S3, S12
7	S16	14	–

Now we can differentiate students and prepare the ranking. Remembering that the most important learning outcome was LO1, the students who were on the first position in the ranking based on this learning outcome are on the first 5 positions, so we differentiated them. It is interesting that student S10 who did not acquire learning outcome LO4 takes the first position in the ranking. Students who took the third position in the ranking based on learning outcome LO1 take positions 11-13.

#### 4 First Type Fuzzification

Now, we will fuzzify the relation between learning outcomes and items by letting the experts who define levels of verifying learning outcomes by items to use values from the interval  $[0,1]$ . The membership functions of the relation between the given learning outcome and items are presented in Fig. 2.



**Fig. 2.** The membership functions of the relation between learning outcomes LO1-LO4 and items.

Now using the crisp relation between students and items and type 1 fuzzy relation between learning outcomes and items we can calculate, using the S-T composition, type 1 fuzzy relation between learning outcomes and students which values denote levels of acquirement learning outcomes by students. The values of this relationship are presented in Table 5.

**Table 5.** Levels of acquirement of learning outcomes by students.

Students	Learning outcomes			
	LO1	LO2	LO3	LO4
S1	0.79	0.68	0.67	0.66
S2	0.83	0.73	0.74	0.79
S3	0.63	0.71	0.73	0.62
S4	0.71	0.73	0.59	0.77
S5	0.72	0.74	0.68	0.79
S6	0.77	0.68	0.55	0.63
S7	0.56	0.69	0.58	0.60
S8	0.75	0.72	0.78	0.77
S9	0.80	0.72	0.61	0.78
S10	0.79	0.63	0.64	0.41
S11	0.67	0.69	0.52	0.63
S12	0.60	0.71	0.72	0.62
S13	0.55	0.66	0.52	0.57
S14	0.72	0.73	0.59	0.78
S15	0.80	0.72	0.61	0.78
S16	0.72	0.66	0.69	0.60
S17	0.81	0.74	0.70	0.79
S18	0.78	0.67	0.57	0.64

Now we can prepare rankings on the basis of levels of acquirement of each learning outcome LO1 – LO4 (separately) by the students. However, since they should acquire all learning outcomes, we prepare the ranking based on all of them using the lexicographical order. Assume that learning outcome LO1 is the most important, then LO2, LO3 and finally LO4. After using this information we can prepare ranking of students in Table 6.

**Table 6.** Ranking of students.

Position	Student	Position	Student
1	S2	10	S14
2	S17	11	S16
3	S9, S15	12	S4
4	S1	13	S11
5	S10	14	S3
6	S18	15	S12
7	S6	16	S7
8	S8	17	S13
9	S5	18	–



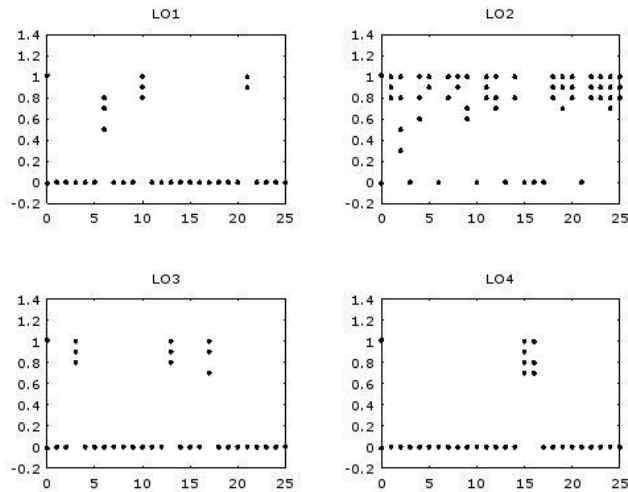
At first we can notice that the students take different positions in the ranking, only two students S9 and S15 have got the same position in the ranking. Now the best student is S2 and the poorest student is S13.

This ranking shows that student S2 is the best one when learning outcome LO1 is the most important one. Of course, if we choose the different order of importance of learning outcomes, the ranking will be different. The possibility of preparing different rankings according to specific criteria is really important for recruitment officers because they need candidates with specific abilities and skills.

Comparing the fuzzification and the IRT method we can see that the IRT takes into consideration only the difficulties of items and the examinee's abilities. Our method enables to calculate the levels of learning outcomes' acquirement taking into consideration one, a few or all learning outcomes.

## 5 Second Type Fuzzification

Now, we fuzzify further the relation between learning outcomes and items by letting the experts who defined levels of verifying learning outcomes define their own value belonging to the interval  $[0,1]$ . Hence the sample membership functions of the relation between the given learning outcome and items are presented in Fig. 3.



**Fig. 3.** The membership functions of the relation between learning outcomes LO1-LO4 and items.

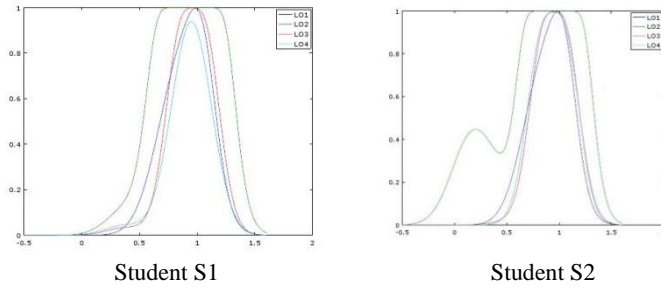
In order to prepare another ranking of students, we will use type 2 fuzzy relations [11,13].

Let  $A = \{-0.5, -0.49, \dots, 0, \dots, 1.6\}$  be the basic membership for all secondary membership functions for  $j = 1, 2, 3, 4$ ,  $i = 1, 2, \dots, 149$  and  $k = 1, 2, \dots, 25$ . Let  $m_{j,k}$  and  $s_{j,k}$  denote the average mean and standard deviation of values set by experts for learning outcome  $j$  and item  $k$ . Let each secondary membership function of the relation between learning outcome and item be defined the Gauss function  $\mu_1(x, j, k) = \exp(-(x - m_{j,k})^2 / s_{j,k})$  for each  $j, k$  and  $x \in A$ .

Since students can earn 0 or 1 points, so the secondary membership functions of the type 2 fuzzy relation between items and students can be the Gauss functions defined as follows:  $\mu_2(x, i, k) = \exp(-(x - m_{i,k})^2 / s_{i,k})$  for each  $i, k$  and  $x \in A$ , where  $m_{i,k} = 0, 1$  and  $s_{i,k} = 0.1$ .

Now let the S-T composition between relations  $R_1$  and  $R_2$  be defined as follows:

$R_3(x, j, i) = 1 - (1 - R_1(x, j, 1) \cdot R_2(x, 1, i)) \cdot \dots \cdot (1 - R_1(x, j, 25) \cdot R_2(x, 25, i))$  for each  $j, i$  and  $x$ . After the S-T composition, the sample secondary membership functions of the type 2 fuzzy relation between learning outcomes and students are presented in Fig. 4.



**Fig. 4.** The membership functions of the relation type 2 fuzzy relation between learning outcomes LO1-LO4 and students.

The next step is to find the method of comparing the results of students on the basis of the calculated secondary membership functions of the relation between learning outcomes and students. Let the level of acquirement of learning outcome  $j$  by the student  $i$  be equal to first coordinate  $x_{j,i}$  of the maximal point of the secondary membership function. Moreover, as we can see (Figure 3) some of the functions are “slim” and some are “wide”. Hence we can assume that if the function is “slim”, so the likelihood of the result is higher and when the function is “wide”, so the likelihood is smaller.

Thus the likelihood of the level of acquiring learning outcome  $j$  by student  $i$ , called the range, is defined as follows:

$$range(j,i) = A(a_2) - A(a_1)$$

where

$$a_1 = \min_{x \in A} \mu_3(a_1, j, i) > 0.5 \text{ and } a_2 = \max_{x \in A} \mu_3(a_1, j, i) > 0.5.$$

Thus for student  $i$  and learning outcome  $j$  we get the pair  $(x_{j,i}, range(j,i))$ . Hence we got the set of values of acquiring learning outcomes with their ranges. The pairs of these values for learning outcomes LO1 and LO2 and all students are put to Table 8.

**Table 8.** Levels of acquirement of learning outcomes LO1 and LO2 with their ranges.

Students	Learning outcomes			
	LO1-value	LO1-range	LO2-value	LO2-range
S1	0.98	0.49	0.96	0.8
S2	0.98	0.49	0.97	0.78
S3	0.98	1.32	0.95	0.82
S4	0.98	1.26	0.97	0.79
S5	0.98	1.26	0.96	0.79
S6	0.98	0.49	0.97	0.81
S7	0.07	0.38	0.96	0.81
S8	0.86	0.32	0.96	0.79
S9	0.98	0.49	0.96	0.8
S10	0.98	0.49	0.95	0.81
S11	0.98	1.26	0.95	0.82
S12	0.07	0.38	0.95	0.79
S13	0.07	0.38	0.94	0.82
S14	0.98	1.26	0.96	0.79
S15	0.98	0.49	0.95	0.81
S16	0.86	0.32	0.97	0.81
S17	0.98	0.49	0.95	0.8
S18	0.98	0.49	0.97	0.81

To prepare the ranking we have to defuzzify the achieved results [11,4]. We assume that the first value should be greater and for students who achieved the same first values, the second one should be smaller. Thus on this basis we can prepare the ranking of students according to each learning outcome but as in the previous sections, we will present the ranking based on the lexicographic order assuming that learning outcome LO1 is most important, then LO2, LO3 and LO4. Hence we achieve the following ranking of students presented in Table 9.

**Table 9.** Ranking of students.

Position of students			
Position	Student	Position	Student
1	S2	10	S5
2	S6, S18	11	S11
3	S9	12	S3
4	S1	13	S16
5	S17	14	S8
6	S15	15	S7
7	S10	16	S12
8	S4	17	S13
9	S14	–	–

Comparing the rankings presented in Tables 6 and 9 we can notice that the first and last positions are the same and position of other students are similar. However, using the type 2 fuzzy relations we have more information because we can also describe the likelihood of acquirement of the learning outcomes by given students.

Even if this ranking did not differentiate students S6 and S18, it is not worse than the previous rankings. Moreover, we have got more information about likelihood of acquirement of learning outcomes.

## 6 Results

Nowadays rankings of students are prepared very often. For example, in Poland all universities to admit students for the first year course prepare the ranking of candidates based on the results of the final secondary school examination and the grades of specific subjects. Universities could choose students based on the levels of acquirement of learning outcomes, not only on the average mean. All students discussed in the paper are on the same position after the first part of the examinations.

The paper presents different methods of preparing rankings based on the IRT algorithm and three different level of fuzzification of the relation between learning outcomes and items: crisp (two different position in the ranking), type 1 (we can prepare the ranking using additionally the lexicographical order of levels of acquirement of learning outcomes by students) and type 2 (the ranking based on the lexicographical order of levels of acquirement of learning outcomes by students gives additionally the information about likelihood of this acquirement).

The IRT method takes into account only difficulties of items and examinee's abilities described as learning outcomes. Moreover, when we prepared the ranking based on examinee's abilities of all discussed learning outcomes we have got only 3 positions.

The ranking prepared on the basis of the crisp relation between learning outcomes and items had only 2 positions. After the first and second fuzzifications, we have got the rankings with 17 positions and in the case of type 2 fuzzy relations we have got more information (the likelihood of levels of learning outcomes acquirement).

The next step is to find more precise measure than range to calculate the likelihood of acquirement learning outcomes by students, for example the area between x-axis and Gauss functions.

## References

1. European Higher Educational Area, Homepage, [www.ehea.info](http://www.ehea.info), last accessed 2016/07/12.
2. Rozporządzenie Ministra Edukacji Narodowej z dnia 23 grudnia 2008 r. w sprawie podstawy programowej wychowania przedszkolnego oraz kształcenia ogólnego w poszczególnych typach szkół.
3. Zasady oceniania rozwiązań zadań, Egzamin maturalny w roku 2014/2015, Centralna Komisja Egzaminacyjna, [www.cke.edu.pl](http://www.cke.edu.pl), last accessed 2016/07/12.
4. F.B. Baker, F.B.: The basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation USA (2001).
5. Dobrosielski W.T., Szczepański J., Zarzycki H.: A Proposal for a Method of Defuzzification based on the Golden Ratio – GR, CONFERENCE IWIFSGN, Cracow 2015, Novel Developments in Uncertainty Representation and Processing, pp. 75-84, Springer International Publishing (2016).
6. Duch W., Wieczorek T., Biesiada J., Blachnik M.: Comparison of feature ranking methods based on information entropy, Proc. of Int. Joint Conf. on Neural Networks (IJCNN), Budapest 2004, IEEE Press, pp. 1415-1420 (2004).
7. Duch W., Winiarski T., Biesiada J., Kachel, A.: Feature Ranking, Selection and Discretization, Proc. Joint Int. Conf. on Artificial Neural Networks (ICANN) and Int. Conf. on Neural Information Processing (ICONIP), Istanbul, pp. 251-254 (2003).
8. Hambleton R.K., Swaminathan H.: Item Response Theory, Principles and applications, Springer Science+Business Media, LLC (1991).

9. Mreła A., Sokołov O., Katafiasz T.: Types of fuzzy relations' composition applied to validation of learning outcomes at mathematics during final high school examination, in: Mreła A., Wilkoszewski P. (ed.): *Nauka i technika u progu III tysiąclecia*, Wydawnictwo Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy, Bydgoszcz, pp.119-132 (2015).
10. Нейман Ю.М, Хлебников В.А.: Введение в теорию моделирования и параметризации педагогических тестов, М., Прометей (2000).
11. Rutkowski R.: *Metody i techniki sztucznej inteligencji*, PWN, Warsaw (2009).
12. Zadeh L.A.: Fuzzy sets, *Information and Control* 8, pp.338-353 (1965).
13. Zadeh L.A.: The Concept of a Linguistic Variable and Its Application to Approximate Reasoning–1, *Information Sciences*, vol. 8, pp. 199–249 (1975).