

# A Computer-Based Technology for Processing and Visualization of Session's Data

I. Zheliazkova, P. Valkova, G. Georgiev

**Key Words:** Computer-based technology; experiment; test; task; exercise; session's data; processing; visualization.

**Abstract.** Student, teacher, and course evaluation are recognized as mutually connected decision making problems based on criteria involving multiple and possibly conflicting indicators. The paper justifies the need of such evaluation. A natural experiment is described. The experiment was divided into two sessions respectively for pretest and test-like exercise. A computer-based subject-independent technology for processing and visualization of session's data is described. Then, the descriptive statistics, correlation, regression, and cluster analysis are applied for some decision making problems.

## 1. Introduction

In the teaching process often a need raises to determinate the best student/teacher; select a subset of students/teachers, satisfying a given set of indicators; rank the students/teachers or categorize them in teams with similar attitudes. In these and other cases a precise student evaluation can provide an useful feedback for improvement of the learning/teaching practice and effectiveness. Teacher and course evaluation are recognized as decision making problems based on criteria involving several indicators and mutually connected with the student evaluation. The evaluation could also give the teacher a feedback on student group needs, an opportunity to learn new teaching technologies from other teachers and to counsel from the cognitive, pedagogical, and psychological principals of higher education [4,9].

The first statement from the limited literature review [2,3,4,5,6,7] was that the criteria for evaluation vary among courses and even within a given course. It is significant to find the indicators that have impact on the evaluation. The aim of the course, its knowledge level, teaching technology, and teacher's preferences has to be taken into consideration.

Another statement was that the technologies applied for this purpose are often selected arbitrarily. Sometimes the teach

ers are already familiar with the technology, other times the technology is developed in an ad hoc way, finally the technology is chosen simply because the appropriate tool is available


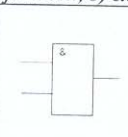
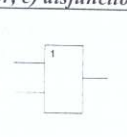
The aim of this paper is to present a computer-based technology for statistical processing and visualization of session data. The experimental study is similar to our previous ones ([1,10,13,14,15,16]) and the main deference is its application to test and exercise sessions covering a lecture topic. So the principal opportunity of implementation of a common and adaptive teacher tool for the needs of decisions making support will be confirmed, the sample size will become more representative, and a new evidence will be added to justify the statements.

## 2. Description of the Experiment

The experiment, described in this section, was aimed to prove the adaptation of the implementation of the tool for multi-criteria decision making by the teacher. The experimental group consisted of part-time Computer Systems and Technologies (CST) bachelor degree students (3-rd year, 2-nd semester) at Rousse University. The experiment was divided into two sessions respectively for pre-test and test-like exercise. Both sessions were carried out within the framework of the course in Discrete Structures and Modelling (DSM) on the topic „Basic elements of a computer“. The elapsed time between the written test on Synthesis and Analysis of Logical Circuits (SALC) and Computer Organization (CO) (other related courses, studied in the previous semester) and the experiment was after 5-6 months.

The test was planned by the lecturer of the DSM course as a pre-test in order to check if the students in the group have the necessary background knowledge to perform the exercise effectively and also to add missing knowledge or to correct wrong knowledge. The pre-test contained of 15 questions with total scores  $P_{max} = 158$  and planned duration  $T_{max} = 30$  min. The questions' types were four, namely: multiple choice, unordered keywords, ordered keywords, and unordered pairs [16]. An unanswered question was interpreted as missing knowledge, and incorrect answers as wrong knowledge that decreases the student's points (PO) to  $P_1$ . Depending on its type, each  $j^{th}$  question brings different number of points  $P_{maxj}$ . If the student's knowledge is evaluated higher than a threshold  $K_{min} = 0.40$  (minimum required background knowledge), then they can perform the exercise; if they fail the test, they are not allowed to do the exercise. They were also told that the time for the pre-test was unlimited and will be used as an assessment criteria ( $T_0/T_{max}$ ) only for research purposes. Each student was given a template Microsoft Word document, describing a number of questions (figure 1) they had to answer in an edit field. However

**Question 2: Unordered pairs**  
Point out the correspondence between numbers of figures and basic logical elements: a) conjunction, b) exclusive or, c) disjunction

		
Fig. 1	Fig. 2	Fig. 3

Answer: 2>a; 1>b; 3>c  
Parameters: L=2; Q=9; Cp=0.67  
Links: 1.1.2. AND, 1.1.6. eXclusive OR, 1.1.3. OR

Figure 1. The template for the student's questions



**Table 1.** The primary data of the test session

Student	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	P0	P1	T0	S1	
053527	St1	2	9	9	9	9	9	15	5	15	0	15	11	12	2	20	142	126	28	4.50
063526	St2	3	9	9	9	9	9	15	0	15	8	15	11	6	9	20	147	136	35	3.89
063533	St3	3	9	9	9	9	9	15	5	15	8	15	11	6	9	20	152	146	60	2.43
063535	St4	2	9	9	5	9	9	15	5	8	4	15	11	6	9	20	136	114	32	3.56
063536	St5	2	9	9	9	9	9	15	5	15	8	0	11	12	9	20	142	126	27	4.67
063537	St6	1	9	9	9	9	9	15	5	15	8	15	11	12	2	20	149	140	22	6.36
063538	St7	3	9	9	9	9	9	15	0	15	8	15	10	6	9	20	146	134	28	4.79
063539	St8	1	9	9	9	9	9	0	0	0	8	15	11	12	2	20	114	70	33	2.12
063544	St9	3	9	9	4	4	4	15	3	2	1	15	11	6	9	20	115	72	65	1.11
083666	St10	2	9	9	9	9	9	0	0	0	0	15	11	12	2	20	117	76	50	1.52
063545	St11	0	3	9	5	5	9	15	0	8	8	15	11	6	9	20	123	88	40	2.20

the usage of such technology, instead of an on-line environment, increases the teacher's labour cost for processing the student answers.

Table 1 represents the experimental data from the test session, brought into a Microsoft Excel table. It is a well known fact from the cognitive psychology, that each learner has his/her own tactic and speed of task performance. Speed of a learner's test session can be precisely computed by dividing the knowledge volume of a test/exercise to the time for performing it by the given learner.

In table 1 the following notations are used: Qij is the number of points the j<sup>th</sup> student was received for the i<sup>th</sup> question in the test; P0j is the total number of points for the j<sup>th</sup> student without punishment for wrong answers); P1j is the total number of points for the j<sup>th</sup> student with punishment for wrong answers; T0j is the time for completing the entire test for the j<sup>th</sup> student; S1j is the final mark of the j<sup>th</sup> student on a six-based scale.

The exercise session was carried out straight after the test session had been evaluated by means of a Windows-based standalone environment for modelling and simulation (described in more detail in [11,12]). The tasks related to 7 different logic

circuits, starting from elementary two-input gates (OR, AND, XOR, NOR) and progressing to somewhat more complex ones (RS trigger, full binary adder). Figure 2 represents the task for the two-input AND gate.

Each task has a fixed structure and consists of the following elements: type, formulation, instructions, parameters, and solution. The last could contain unlimited number of resource files, namely: the program in KDL, graphical, and truth table.

The environment that has been used saves the results in the form of a 'program' in KDL file, which stores student's knowledge for the system under investigation. For the needs of this experiment, two files are compared to compute the proximity of each learner's program in KDL to the teacher's one, and a mark R in the range [0,1] is formed. Second, the simulation and truth table are evaluated, assigning them marks (respectively S and T) in the range [0,1]. For task T1 to T5,  $S \in [0, 0.25, 0.5, 1.0]$ ; the same applies to test (there are only 4 combinations of input values). Finally, the total mark for each task is computed as a weighted

$$X = \frac{10R + 5S + 8T}{23}$$

The raw data from the exercise are presented in table 2. The E column represents the normalized value of wrong answers for the second subtask of each exercise task.

**Table 2.** The primary data of the exercise session

		X	E
053527	St1	0.51	1.00
063526	St2	0.71	0.39
063533	St3	0.46	0.62
063535	St4	0.53	0.37
063536	St5	0.74	0.32
063537	St6	0.16	0.75
063538	St7	0.50	1.00
063539	St8	0.16	0.75
063544	St9	0.21	0.81
083666	St10	0.20	1.00
063531	St11	0.41	1.00

**TASK 1:** Write down the truth table of a two-input "AND" logic gate, by simulating its behaviour in static mode and examining the output Y as a function of the inputs X1 and X2.

**Directions for using the environment:** Find out the proper BMP file. Generate a simulation program using the tool generator.exe; use 1.bmp as background image. Save the program as 1.trn; also, copy and paste its contents into Table 1.

- Table 1. Simulation program for two-input AND

1. Start the tool simulator.exe, open 1.trn and simulate different values for the inputs.

- After any change of input and/or output value, make screenshots and paste them as Fig. 1, 2, 3, 4.

<input type="text" value="&lt;paste here&gt;"/>	<input type="text" value="&lt;paste here&gt;"/>
Fig.1	Fig.2
<input type="text" value="&lt;paste here&gt;"/>	<input type="text" value="&lt;paste here&gt;"/>
Fig.3	Fig.4

- Fill in the truth table in Table 2.

Table 2

X <sub>1</sub>	X <sub>2</sub>	Y
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>

- Enter the time it took you to solve this task: ..... min.

**Figure 2.** The template for the student's exercise task

### 3. Application of the Descriptive Statistics

Application of the descriptive statistics gives informative, precise, and expressive results which facilitate studying the influence of different indicators, criteria, and candidates and the relationships among them. In general, this technique usually includes: 1) summaries using a mean or a mode; 2) expression variability in terms of range or inter quartile range; 3) representing of the observations in a dot plot, bar chart or histogram. The tables, constructed for calculating results and diagrams for their visualization, are suitable for different categories of decision making (sorting, selecting, and ranking). Add the average values, interpreted as group criteria, to the dynamic part of the group model, and to the dynamic part of the individual model of the teacher.

Table 3 has been composed based on the table with the test session results. The results of some students have been filtered out, based on their ID and some conflict data. Rows, for which the values in each cell are equal or very similar, can be treated as a reliable evidence for fraud by the corresponding student. They have to be dropped out from further consideration. Hereinafter St11 and St12 fail further consideration due to another reason - incorrect ID for the test and exercise session. Table 3 contains the calculated coefficients for each question and each student. Rows M0 and M1 contain the normalized values for the marks from table 1:

$$M0_{i,j} = \frac{P0_{i,j}}{P_{\max}}; M1_{i,j} = \frac{P1_{i,j}}{P_{\max}}$$

Rows M2 and M3 contain the marks from M0 and M1 with applied bonus/punishment according to the time it took each student to do the test:

$$M2_{i,j} = M0_{i,j} \frac{T_{\max}}{T_j}; M3_{i,j} = M1_{i,j} \frac{T_{\max}}{T_j}$$

The column „Avg“ shows the proximity of the actual student's scores to the maximum scores. For all students this average value is more than the frequently recommended in literature [2] threshold 0.40. This can be interpreted as well done self assessment before the organized testing.

The final results, which show dependence of the evaluation ranking on the chosen criteria, are illustrated by means of the series of bar diagrams on figure 3. The main conclusion regarding the 0.40 threshold is that all students pass the pre-test and can perform the exercise. It shows the best performance for this case is St6 with a score of 1.21 and St9 has scored only 0.21 for M3 which is the lowest result of the group. The picture changes slightly if M2 is considered - then St9 fails the pre-test. Even more significant changes are observed if M3 is taken into consideration - then one more student, St10, does not pass the pre-test. Table 4 and figure 4 are results of applying the same algorithm on the real data performance for the exercise session (table 2). The only difference is that instead of four series (M0, M1, M2, M3), here the series are two, denoted as E (normalized number of missing and redundant nodes) and X (normalized final student's evaluation).

Based on the formulated criteria for the exercise performance, we found experimentally that St5 shows the best exercise performance. On the other hand, the minimum performance points for this session 0.16 are shown by St6 and St8 from this group.

Together with St9 and St10 they have to perform the exercise again in order to improve their practical skills. But the academic group as a whole can continue, as it has an average of 0.42 points, which is above the accepted threshold. The value of Cp (degree of system prompt) in the test and exercise is approximately the same, but the exercise is more difficult (0.70) than the test (0.18). Cp is used by the authors as a measure of what part of the answer is made available to the student by the environment and is a real number between 0 and 1.

A similar processing of the questions/tasks data and their graphical interpretation has to confirm the expectation, that the test and exercise are valid, reliable and of high quality. These data are student-independent and present the criteria for questions/task and test/exercise quality usually with the pedagogical requirement for decreasing difficulty. The results from this processing is presented in table 5 and figure 5 for the test and table 6 and figure 6 for the exercise. The average results are added to the dynamic part of the teacher model.

The pre-test session also allows specifying the initial time for its performance planned only on the basis of the teacher's minutes and this means that the initially planned 30 minutes could be increased, for example, to 40 minutes whereas the time planned for the exercise can remain 90 minutes. Note, that in the standalone or Web-based testing environment, the mechanism "one task on one or more screens" is preferred to „one test on one screen“. This gives a possibility to precise the average time for performing each task in the task base.

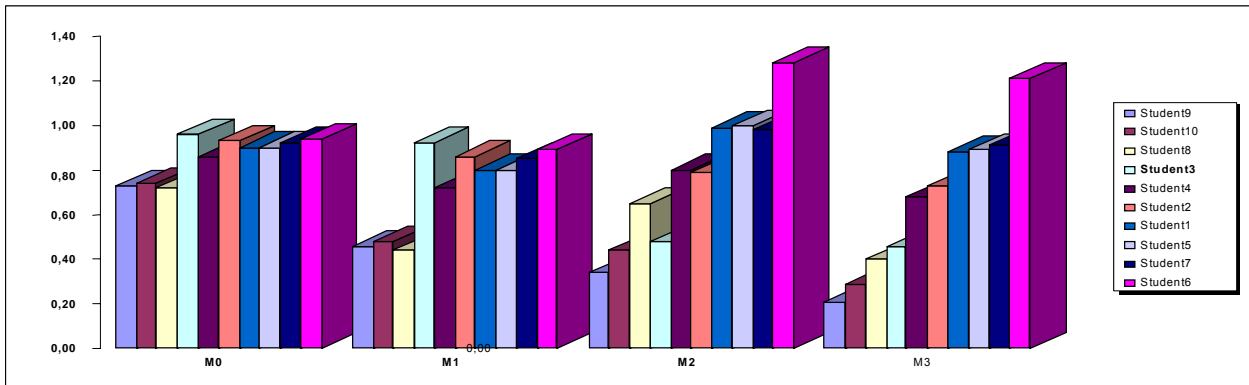
The classification of the questions/tasks in a test/exercise can be done by means of histograms under the given number of classes according to the average value of a chosen question/task indicator. For the needs of visualization 5 groups were chosen as in Likert's psychometric scale for the needs of questionnaires and in the traditional six-based scale in Bulgaria (table 7).

For example, the histogram on figure 7 together with the average approximation for the test corresponds the case when the test questions were divided into five groups respectively: very easy (VE) - from 0 to 0.29, easy (E) - 0.30-0.39, moderate (M) - 0.40-0.59, difficult (D) - 0.60-0.690, and very difficult (VD) - over 0.70 (table 8). The same grouping had been applied to the exercise tasks and similar table (table 9) and histogram (figure 8) were received.

Such kind of visualization could be useful in decisions about increasing the level of knowledge within different forms of teaching, distribution of the given indicator known from the pedagogy theory, or extracting questions/tasks from the database, which meet the given distribution. Most of the educators agree that the distribution of marks in an academic group has to correspond to the Normal Gauss distribution. The distribution of the test questions difficulty and the exercise tasks difficulty, recommended by the pedagogy science, has to be closer to linear.

**Table 3.** The exercise marks of application of descriptive statistics from the test for each student results

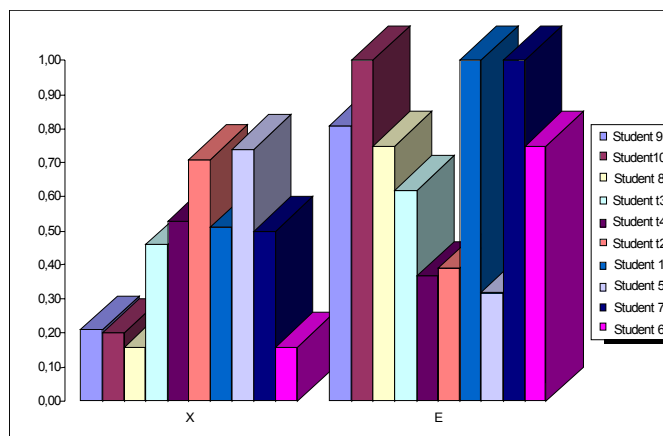
Student	St1	St2	St3	St4	St5	St6	St7	St8	St9	St10	Avg
M0	0.90	0.93	0.96	0.86	0.90	0.94	0.92	0.72	0.73	0.74	0.86
M1	0.80	0.86	0.92	0.72	0.80	0.89	0.85	0.44	0.46	0.48	0.72
M2	1.00	0.79	0.48	0.80	1.00	1.28	0.98	0.65	0.34	0.44	0.78
M3	0.89	0.73	0.46	0.68	0.89	1.21	0.91	0.40	0.21	0.29	0.67



**Figure 3.** The exercise marks of application of descriptive statistics from the test for each student results

**Table 4.** Marks from the exercise for each student

	St1	St2	St3	St4	St5	St6	St7	St8	St9	St10	Avg
E	1.00	0.39	0.62	0.37	0.32	0.75	1.00	0.75	0.81	1.00	0.70
X	0.51	0.71	0.46	0.53	0.74	0.16	0.50	0.16	0.21	0.20	0.42

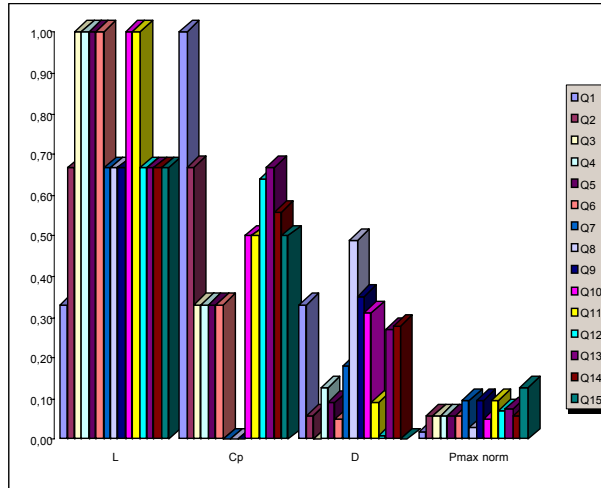


**Figure 4.** Marks from the exercise for each student



**Table 5.** Parameters of each test question

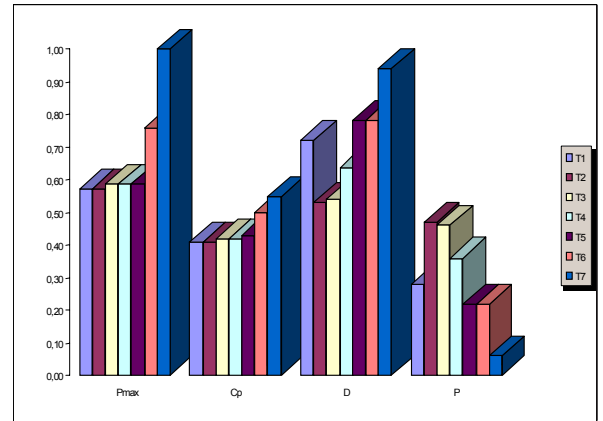
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Avg
Pmax	3.00	9.00	9.00	9.00	9.00	9.00	15.00	5.00	15.00	8.00	15.00	11.00	12.00	9.00	20.00	10.53
L	0.33	0.67	1.00	1.00	1.00	1.00	0.67	0.67	0.67	1.00	1.00	0.67	0.67	0.67	0.67	0.78
Cp	1.00	0.67	0.33	0.33	0.33	0.33	0.00	0.00	0.00	0.50	0.50	0.64	0.67	0.56	0.50	0.42
D	0.33	0.06	0.00	0.13	0.09	0.05	0.08	0.49	0.35	0.31	0.09	0.01	0.27	0.28	0.00	0.18
Pnorm	0.02	0.06	0.06	0.06	0.06	0.06	0.09	0.03	0.09	0.05	0.09	0.07	0.08	0.06	0.13	0.07



**Figure 5.** Parameters of each test question

**Table 6.** Parameters of each exercise task

	T1	T2	T3	T4	T5	T6	T7	Avg
Pmax	0.57	0.57	0.59	0.59	0.59	0.76	1.00	0.67
Cp	0.41	0.41	0.42	0.42	0.43	0.50	0.55	0.45
D	0.72	0.53	0.54	0.64	0.78	0.78	0.94	0.70
Pnorm	0.28	0.47	0.46	0.36	0.22	0.22	0.06	0.30



**Figure 6.** Parameters of each exercise task

**Table 7.** Normalized and traditional marks

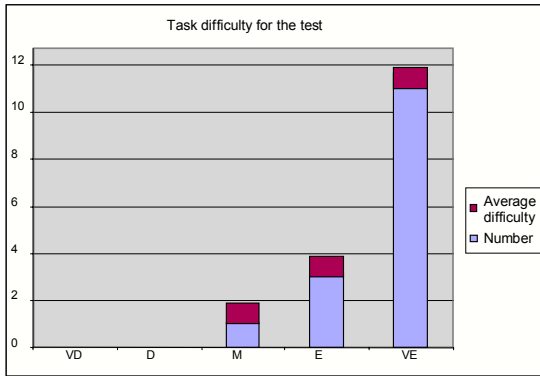
Normalized mark	Traditional mark
<0.40	Fail (2)
0.40 - 0.55	Fair (3)
0.56 - 0.70	Good (4)
0.71 - 0.85	Very good (5)
0.86 - 1.00	Excellent (6)

**Table 8.** The test results of cluster analysis

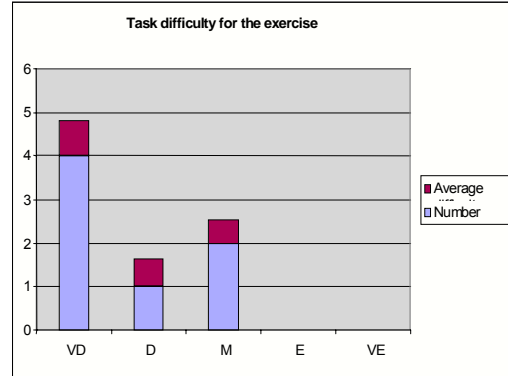
Category	Question ID	Number	Difficulty
VD	-	0	n.a.
D	-	0	n.a.
M	8	1	0,49
E	1,9,10	3	0,33
VE	2,3,4,5,6,7,11,12,13,14,15	11	0,11
Total:		0,31	

**Table 9.** The exercise results of cluster analysis

Category	Task ID	Number	Difficulty
VD	1, 5, 6, 7	4	0,80
D	4	1	0,64
M	2, 3	2	0,54
E	-	0	n.a.
VE	-	0	n.a.
Total		0,66	



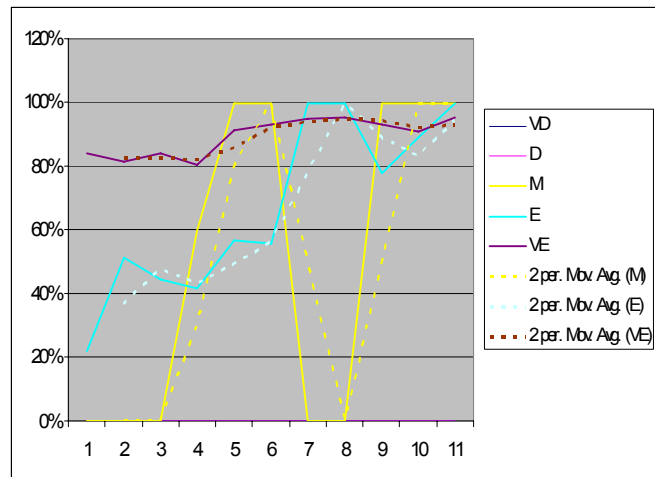
**Figure 7.** Histogram of test questions difficulty



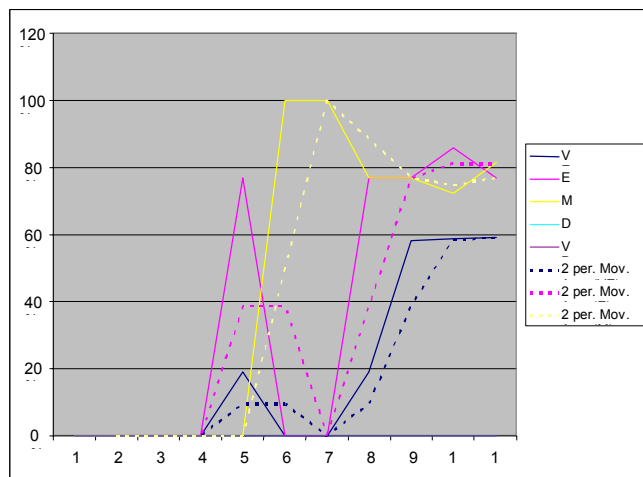
**Figure 8.** Histogram of exercise tasks difficulty

**Table 10.** The results of correlation analysis

Criteria	P1	T0	Cp	D	Pnom	L	TM	EM
P1	-	0.3997						
T0	0.3997	-						
Cp			-	0.1207	0.2084			
D			0.1207	-				
Pnom			0.2084		-	0.1070		
L					0.1070	-		
TM							-	0.4393
EM							0.4393	-



**Figure 9.** S-curves for the test questions



**Figure 10.** S-curves for the exercise tasks

The main graphical characteristic of the question difficulty is the typical S-shape curve described earlier by Kurata & Sato [8] and Bizhkov [2] and confirmed in our earlier experiment, mentioned above. The experimental (solid curves) and the their moving average approximation (dotted curves) for the test questions and exercise tasks are presented on *figure 9* and *figure 10* respectively. The interpretation of an S-curve will be reminded briefly here: 1) the width of the deviation along the X-axis of a curve corresponds to the difficulty of the corresponding question, because too few correct answers are given in the test; 2) the inclination describes the probability of guessing. The lower the grade is, the more even is the curve of the distribution of answers. The average curve should be close to the diagonal of the chart. The smaller angle of the approximation line with respect to the X-axis means lower sensitiveness with regards to the student's knowledge differences. From the visual comparison of both figures follows the expected conclusion that the test-like exercise is more sensitive than the intelligent test.

Comparison of two different forms of sessions (in our case test and exercise) on one and the same lecture material can be useful for improving the questions/task formulation and updating the value of test questions - difficulty.

#### 4. Application of Correlation Analysis

In general, correlation analysis is used to measure the strength of the relationship between two statistical factors. The linear correlation coefficient is usually used as a qualitative indicator of the relationship between final marks from different sessions. The value of this coefficient  $R$ , a real number in the range of  $[-1,1]$ , shows how strong is the relationship. For example, if  $R$  is in the range  $0.0 \div 0.3$  then the relationship is weak;  $0.3 \div 0.5$  - moderate;  $0.5 \div 0.7$  - significant;  $0.7 \div 0.9$  - high;  $0.9 \div 1.0$  - very high. The practically used algorithm is: 1) Construct a new table with dimension  $N \times N$ , where  $N$  is the number of indicators and/or criteria; 2) Compute  $R(i, j)$  - the

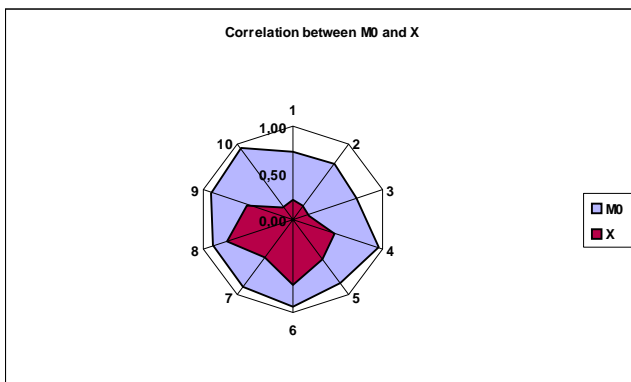


Figure 11. Web chart for  $R(M_0, X)$

correlation coefficient between the  $i$ th and  $j$ th factor; 3) Add blank for the diagonal cells, and use already computed values across the diagonal, as  $R(j, i) = R(i, j)$ ; 4) Draw a web chart, similar to these on *figure 11* and *figure 12*; 5) Draw conclusions based on the strength of the relationships.

The corresponding correlation coefficients received from the test data are presented in *table 10*. One of the most inter-

esting relations is between the marks from the test (TM) and the exercise (EM).  $R(TM,EM) = 0,4393 \gg 0,44$ , which means a moderate relationship. M1 was used as final mark from the test.

The degree of correlation between the test mark and the exercise mark on a six-based scale is quite good, and allows to conclude that both environments for knowledge testing and for training in dynamic systems are feasible and yield sustainable results.

Probably the main reason why the value of  $R(TM,EM)$  is not higher (*table 10*), is that the exercise tasks for modelling and simulation are more complex and the students use a new software environment to perform them. This is confirmed by some raw data, not presented here, namely the duration for solving each task. It took most students 2-3 times longer to complete the first task, than it took them to complete the following ones. So, unavoidably, the exercise partly measures students' skills to use a software environment, e.g. technological skills as well as their specific knowledge. Still, this side effect is reduced by the fact that in this BSc programme (Computer Systems and Technologies) they use many different software environments and quickly gain some experience in adapting to new ones.

The correlation analysis could also be applied to couples of question/task criteria. For example, of interest is the relationship between P1 and T0 (*table 10*). The computed value  $R(P1,T0) = -0,3997 \gg 0,40$  is negative and in practice means a moderate reverse relationship. Another interesting for the teaching practice relationship is lower and negative,  $R(Cp, D) = -0,1207$ . The conclusion is that as the degree of system prompt ( $Cp$ ) of the tests questions increases, the difficulty ( $D$ ) decreases but the relationship is very weak. This confirms the validity of the formulae for the computation of these parameters and the matches of the conclusions from our earlier study. The strength of the relationship between Pnorm and L has approximately the same value, but it is positive, i.e. a higher level question generally brings more scores ( $R(L,Pnorm) = 0.1070$ ). As expected, the relationship  $R(Cp, Pnorm) = -0.2084$  is negative and stronger.

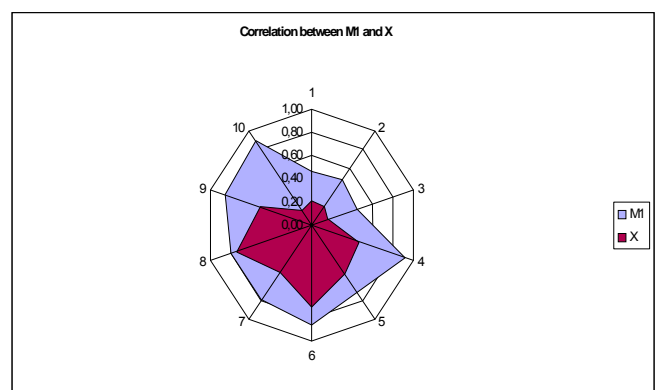


Figure 12. Web chart for  $R(M_1, X)$

#### 5. Application of Regression Analysis

In general, regression analysis is used to find the weights of each indicator presented by regression coefficients in a resulting complex criterion in the form of a regression equation. The practically used algorithm is: 1) Set up the input data table; 2) Construct a new table with the number of columns equal to

**Table 11.** Regression coefficients for the test methods

KT0	KP1	KP0	B	Param
0.00002	0.00677	-0.00724	1.07288	M0
-0.00005	0.00405	0.00455	-0.35694	M1
-0.01580	0.00946	-0.01277	2.03438	M2
-0.01379	0.00740	-0.00428	0.92811	M3

the number of indicators plus one; 3) Compute the regression coefficients for each factor.

In the case of *table 11*, each method M0, M1, M2, M3 presents a complex criterion, and the indicators are P0, P1, T0. The regression equation is  $M = K_{T0}T0 + K_{P1}P1 + K_{P0}P0 + B$ .

The regression equation for *table 12* is  $X = B_1M1 + B_2E + B_0$ , which represents the dependency of the

**Table 12.** Regression coefficients for the exercise method

B1	B2	B0	Param
0.55	-0.38	0.28	X

of the classes subjectively chosen by the teacher, was done by means of histograms. In contrast, cluster analysis is applied when the number of criteria is more than one and the number of classes is unknown in advance. The classical step by step algorithm is based on the Euclid distance and starts with the assumption that each candidate represents a separate cluster. Indicators are assumed to be independent. The practically used algorithm is: 1) Construct a new table with the number of columns equal to the combination of indicators; 2) Add rows as a result of applying the cluster analysis procedure; 3) Each column will contain the subset of candidates; 4) Make explicit conclusions and decisions.

**Table 13.** The test results of cluster analysis

indicators/clusters	P0	P0, P1	P1, T0	P0, P1, T0	P1, T0, Exercise
cluster 1	1,4,5	1,4,5	1,2,4,5,6,7	1,2,4,5,6,7	1,2,4,5,6,7
cluster 2	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10
cluster 3	2,3,6,7	2,3,6,7	3	3	3

final mark from the exercise (X) on the volume of errors in the exercise (E) and the test mark (M1 in this case).

Based on the regression coefficients given in *table 11*, the three evaluation indicators are ranked, thus, for M1 both P1 and P0 are equally important whereas for M3 the duration is the most important factor, next in importance is the number of correct answers and the wrong answers are the least important. The value of Kt0 for M0 and M1 is close to the method error and can be accepted as equal to zero, and its negative value for M2 and M3 can be interpreted as decreasing the corresponding final mark with increasing the time for completing the test.

Analogically this analysis could be applied in order to find the weights of different forms of teaching, e.g. lecture, test, exercise, course work, and so on and to compare these weights for different teachers instead of the weights chosen subjectively by the course team. For *table 12*, the free regression coefficient B0 could be interpreted as an assessment of the input knowledge, the coefficient B1 as the weight of the pre-test, and B2 the weight of the exercise. The negative value of B2 apparently corrects the somewhat too high marks from the pre-test.

As it is well known from the mathematical statistics, regression analysis can also be used by a teacher for forecasting, for example, of the exercise group results after the test results. If later the real results significantly differ from the forecast ones, that can be due to side effects and/or external factors not taken into consideration in the corresponding regression equation. Regarding students, they could be: fraud, unfamiliar type of tasks, insufficient attention or motivation. Other external factors, related to the teacher, could be: poor session planning, organization, and/or delivery.

## 6. Application of Cluster Analysis

In section 5 above, in case of one indicator/criteria the classification of the alternatives/ candidates under the number

*Table 13* presents the results of applying the above algorithm to students classification. As P0 and P1 are dependent indicators, the results are one and the same, whereas the classification based on the main independent indicators P1 and T0, as well as on P0, P1 and T0 change the classification. The number of clusters is equal to the number of categories received by using histograms for the test questions, but the contents is different. The main reason is that only one indicator, namely the difficulty, is taken under consideration.

The students' classification based on the test mark (TM) and exercise mark (EM) leads to 4 clusters. It can serve as a recommendation to form four student teams: three of them consist of 3 students, and one team of 2 students. Note, that indicators such as the communication skills and ability to work in a team are taken into account in neither criterion.

The application of cluster analysis to the test questions or exercise tasks could help the teacher in choosing the appropriate group of questions/tasks from the database for a given group of students. *Table 15* represents such an analysis for the test questions and *table 14* of the exercise tasks. Q1 and Q2 end up in separate clusters mainly due to unexpected combinations of Cp and D. For Q1 Cp = 1,00 and D = 0,33 and for Q2 Cp = 0,67 and D = 0,06. The reasons why a question with high Cp turned out to be difficult could be, for example, unclear formulation, insufficient attention paid to the lecture material, and so on. It has been found that using a single indicator (D), or even two indicators, such as (L, Pnorm) and (D, Pnorm) does not lead to

**Table 14.** The exercise results of cluster analysis

indicators/clusters	TM, EM
cluster 1	2,5
cluster 2	1,6,7
cluster 3	3,9,11
cluster 4	4,8,10



**Table 15.** The test results of cluster analysis

indicators/clusters	Cp, D	D, Pnorm, L	L, Cp, D, Pnorm
cluster 1	1	1	1
cluster 2	2	2	2
cluster 3	3,4,5,6,7,8,9,10,11,12,13,14,15	3,4,5,6,7,8,9,10,11,12,13,14,15	3,4,5,6,7,8,9,10,11,12,13,14,15

significant questions separation within the test. But the combination of Cp and D leads to 3 clusters, and even adding two more indicators (Pnorm and L) does not change the picture.

## 7. Conclusions and Future Work

Although the research is based on test and exercise sessions data covering a given subject topic, the proposed technology has a general character, i.e. it is platform- subject-, and task-independent. The technology is based on the assumption that several important learner's parameters such as coefficient of proximity, coefficient of knowledge retained, speed of learning, and so on based on the most important didactic parameters knowledge volume and time for its testing.

The application of statistical methods for description and analysis for the reasonable decisions concerning student, teacher and course evaluation is illustrated as useful. Specialized tables are constructed and charts are drawn, clearly showing different characteristics of the tests carried out, the group of students involved, and sessions provided by different teachers. The received results confirm some relationships from the authors' team earlier studies and imposes new ones, valid not only for tests with simple questions but for test-like exercises with complex tasks.

## References

- Atanasova, P. L., I. I. Zheliazkova, A. S. Levi. A Computer-Based Strategy for Practical Skills Teaching, Information Science & Computing, International Book Series, No. 6 (Methodologies and Tools of the Modern (e) Learning, 2008, 71-78.
- Bizhkov, G. Theory and Methodology of Didactic Tests. Sofia, Prosveta, 1996 (in Bulgarian).
- Carashtranova, E. L., D. I. Dureva & G. T. Tuparov. Assessment of the Students' Input Level Knowledge and Skills in Information Technologies - Reality and Perspectives. Proceedings of the Second Bulgarian Scientific-Practical Conference on New Technologies in Education, Plovdiv, 2004, 147-154 (in Bulgarian).
- European Association for Quality Assurance in Higher Education Standards and Guidelines for Quality Assurance in the European Higher Education. Pikseri Julkaisupalvelut, Helsinki, Finland, 2005.
- Jelev, G., & Y. Minkova. Determination of Representative Sample Size and Knowledge Assimilation Tests Results Processing. Proceedings of the International Conference „Computer Science“, Sofia, 2004, 274-279.
- Jelev, G., & D. Minkovska. Approaches for Definition the Validity of the Results of the Test for Knowledge Assessment. Proceedings of the International Conference „Computer Science“, Sofia, 2004, 268-273.
- Hristova M. Algorithmization of a Multifactor Model for Assessment of the Quality of Teaching in Higher Schools. - *Journal of Automatics and Informatics*, 2008, No. 2, .37-40 (in Bulgarian).
- Kurata, M. & T. Sato. An Educational and Psychological Test Item Data Base System. - *Journal of Information Processing*, 7, 1984, No. 4, 246-253.
- Quality Management System - Guidelines for Application of ISO 9001:2000 in Education. ISO, Geneva, 2004.
- Valkova, P. L., A. S. Levi, I. I. Zheliazkova. Preparing the Exercise Word Documents in a Task-Oriented Environment for Modeling and Simulation. Proceedings of the First International Conference „Actual Tasks and Multidisciplinary Sciences“, Lozenetz, 6-8 June, 4, 2007, 67-72.
- Zheliazkova, I. I., G. T. Georgiev, M. Andreeva. A Technology for Simulation-Based Training by Means an Intelligent Multimedia Environment. Proceedings of the 17th International Conference SAER, Varna, Bulgaria, 19th-21st September, 2003, 225-229.
- Zheliazkova, I. I., G. T. Georgiev. Representation and Processing of Domain Knowledge for Simulation-Based Training Systems. - *International Journal of Intelligent Systems*, 10, 2000, No. 3, 255-277.
- Zheliazkova, I. I., & R. T. Kolev. Multi-parametrical Analysis of Test Results. Proceedings of the International Scientific Conference „Computer Science“, Chalkidiki, Greece, 2005, part 1, 289-293.
- Zheliazkova, I. I., M. H. Andreeva & R. T. Kolev. Knowledge Testing in Algorithms - An Experimental Study. Proceedings of the International Conference Modern E-Learning, Varna, 2006, 55-62.
- Zheliazkova, I. I., R. T. Kolev, M. H. Andreeva. Knowledge Testing in Algorithms by Means of a Word Technology. Proceedings of the Second National Conference on E-Learning in the Higher Education, 14th - 16th July, Kiten, Bulgaria, 2006, 21-25.
- Zheliazkova, I. I., R. T. Kolev. Task Results Processing for the Needs of Task-Oriented Design Environments. - *International Journal Computers & Education*, 51, 2008, 86-96.

### Manuscript received on 19.07.2010

**Irina Zheliazkova** was born in Lozenetz, Burgas region in 1950. She graduated the Polytechnical University of Kiev, specialty Computer Systems in 1973. There she defended a PhD thesis „Design of Computer-Based Teaching Systems“. Since 1977 she has been working at the Department of Computer Systems and Technologies, University of Ruse where since 1985 as an associate professor. Her research interests are in application of artificial intelligence, multimedia, and internet technologies in environments for teaching, learning, and training. She was a supervisor of four PhD students defended. Irina Zheliazkova is author and co-author of over 200 publications (including patents and articles with impact factor) in English, Russian and Bulgarian.

Contacts: [irina@ecs.uni-ruse.bg](mailto:irina@ecs.uni-ruse.bg)

**Polina Atanasova** was born in 1981. She received both bachelor and master's degrees from the University of Ruse in 2003 and 2005 respectively. The results of her master's degree diploma work in the area of intelligent teaching environments have been already published in four papers. As a PhD student now she is interested in models and tools for structural knowledge teaching and learning.

Contacts: [valkova\\_99@yahoo.com](mailto:valkova_99@yahoo.com)

**Georgi Georgiev** was born in Ruse in 1963. He graduated from the University of Ruse in 1989 with a degree, equivalent to MSc in Computer Engineering. In 1996 he graduated from Warwick University, UK, with an MSc in Information technology for manufacture. He got a PhD in the area of System Programming in 2000. In 2010 he got a tenure as an Associate Professor at the Department of Computer Systems and Technologies in the University of Ruse, where he teaches at present. His research interests are in the areas of System Programming, e-Learning, Internet Technologies.

Contacts: [gtgeorgiev@ecs.uni-ruse.bg](mailto:gtgeorgiev@ecs.uni-ruse.bg)