2015, **4**(2): 37–49
DOI: 10.1515/ijicte-2015-0008

ADAPTIVE TESTING IN PRACTICE

Tatiana Prextová
Department of Information and Communication Technologies, Pedagogical Faculty, University of Ostrava, Fráni Šrámka 3, Ostrava-Mariánské Hory, Czech Republic
tatiana.prextova@osu.cz

Abstract

Testing as such is important for diagnostics and evaluation. It is used as feedback by both the pedagogues and the ones being tested. The more a teacher learns from the test results, the better is their chance to correct, clarify, or modify the test itself; i.e. to carry out the changes in their instruction or the education process. The more a student learns from the test, the better is their chance to thoroughly learn and master information, and to clarify problematic issues of a particular curriculum. Moreover, their motivation to further study is growing as they deal with more demanding tasks. Adaptive testing carried out in a suitable LMS offers such possibilities.

This paper is aimed at the introduction of basic principles and rules of computerized adaptive testing. Moreover, it provides information about the process and results of computerized adaptive testing, which was experimentally carried out on the sample of 53 ninth-grade pupils at the Porubská elementary school in Ostrava.

Keywords

computerized adaptive testing, Barborka 4 LMS, mathematics, experiment, repetition with tutorial mode.

Computerized Adaptive Testing

Basic Principle

The theory of computerized adaptive testing (TCAT) is derived from the theory of adaptive education (TAE) where the process of the Virtual Teacher is proposed – automatically controlled program of personalized education based on pupils' learning characteristics. According to the theory of adaptive education the instruction is carried out by the adaptive Barborka 4 LMS, which was developed by a group of informatics within the scope of the cooperation of the Faculty of Electrical Engineering and Computer Science of the VŠB-Technical University of Ostrava and the Department of Information and Communication Technologies of the Pedagogical Faculty of the University of Ostrava. The rules and author's principles of computerized adaptive testing have been incorporated into this system.

The main principle of the TCAT is the division of tasks into several levels of difficulty from the least to the most demanding. By solving the tasks gradually a student realizes their knowledge and skills; tests the boundaries between knowing and not knowing; can be notified of their mistakes or offered additional instruction concerning a particular part of the curriculum (they are offered a tutorial in the form of a reference to a study material); and in the case of complete not knowing they are provided with an entire solving process including a correct result (Help).

A detailed algorithm for computerized adaptive testing (including both the reference and Help) was described in detail in the article "Evaluating Student's Knowledge Through The Use of Adaptive Testing" (Prextová, 2014). The author's principles and adaptive rules emerged from the created algorithm, which can be seen – in the form of flowchart – in the following picture:

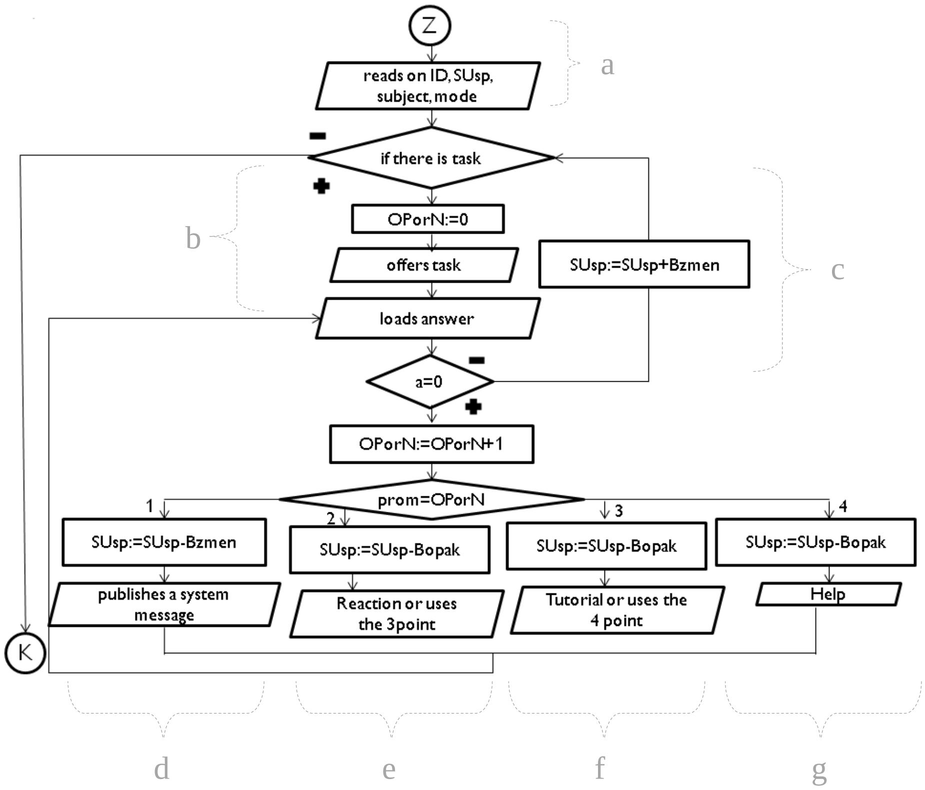


Fig. 1: Algorithm of Computerized Adaptive Repetition with Tutorial

a. Reads an ID of a current student.
    - Loads the current SUsp value.
b. On the basis of a student's choice of subject, a lesson or the repetition mode, the algorithm controls the education process in this mode according to the following procedure of repetition with tutorial: If there are tasks in the sequence of tasks of a selected subject (lesson), then
    - the algorithm calculates a student's current Obod value from their current SUsp value,

- the Virtual Teacher uses the ChooseTask procedure and offers a student a task with the Obod difficulty corresponding to the current SUsp,
- loads a student's answer and evaluates it.

c. If the answer is correct, it raises the SUsp value by the Bzmen value.

d. If the answer is not correct at the first try, it publishes a system message on the incorrectness of the answer and offers a student the second try option; it reduces the SUsp by the Bzmen.

e. If ihe answer is incorrect at the second try, then:

- it reduces the SUsp by the Bopak,
- if there is a Reaction to an answer (expected incorrect answers or generally incorrect answers), it offers the Reaction to a student or uses the 3 point.

f. If the answer is incorrect at the third try, then:

- it reduces the SUsp by the Bopak,
- if there is a reference to a particular tutorial reference as a contextual instruction in any of the previous layers, it offers such a layer to a student or uses the 4 point.

g. If the answer is not correct at the fourth try, then:

- it reduces the SUsp by the Bopak,
- if "help" exists, it provides a student with a detailed solving process including the correct result and invites them to write the correct result down, or simply tells a student the correct result.

Author's Principles

The author's principles for the creation of testing tasks can be divided into two groups.

As far as the *instruction* is concerned, the author can:

- Set the level of difficulty by assigning points for each testing task from the $<l, n>$ interval; where n is chosen by the author according to how many difficulty levels they intend the test to have (for our testing we used 5 levels of difficulty);
- For each task, set a group of equivalent tasks, which differ in their numeric values, opposite symbols, etc.;
- For each task, set not only correct answers but also Reactions (based on the author's experience – at which point they expect a half-correct or incorrect answers to occur);
- For each task, set Reference (the author refers to one of the Barborka 4 LMS layers with the so called tutorial – a study material for a particular task);

- For each task, set Help (it is a special layer of the Barborka 4 LMS that contains an entire solving process including a correct result).

As far as the rules are concerned, the author can:

- Define the relation between the difficulty of a level and the speed of a tested pupil's success rate change (how correct or repeated incorrect answers will change a pupil's success rate value);
    - If the author prefers not to set the relation, the implicitly set relation will be used.

Note: At the start of testing each pupil's quality – success rate value – is registered. This value is then dynamically modified (increased/decreased) in relation to how they advance through the test and the answers to particular tasks.

Rules

The requirements for computerized adaptive testing were incorporated into the mentioned Barborka 4 LMS (Takács, 2014). These are elementary rules of the if-then type, which determine:

I. The relation between a pupil's success rate and the level of difficulty;

II. The change of a pupil's success rate for correct and incorrect answers;

III. The change of a pupil's success rate for repeated incorrect answers;

IV. The reaction of the Virtual Teacher to a pupil's correct answer:

If the answer is correct, publish a system message on the correctness of the answer

V. The reaction of the Virtual Teacher to repeated incorrect answers:

If the answer is incorrect one time, publish a system message on the incorrectness of the answer.

If the answer is incorrect two times, then:

a) Publish Reaction; if it is not available, omit it and continue to b).

If the answer is incorrect three times, then:

b) Display the layer with the Reference to tutorial; if it is not available, omit it and continue to c).

If the answer is incorrect four times, then:

c) Display Help; if it is not available, omit it and display only the correct result.

If need be, a new rule can be added or an existing rule can be modified.

Elementary School Experiment

Creation of Testing Tasks and Questions

The subject of the experiment for verifying computerized adaptive repetition with tutorial was Mathematics for elementary schools – final revision as a preparation for the high school entrance examination. The set of test tasks – questions and tasks from the subject Mathematics for the 9th grade – was divided into the following thematic areas: *Number and variable; Terms and formulas; Data, graphs, and tables; Functions; Plane geometry; Space geometry*.

For each area 25 theoretical questions and practical tasks were created. The theoretical questions test definitions and theorems while the practical tasks test particular life situations. For each task two equivalent tasks were created – they test the same issue but differ in the way they are formulated, numeric values, or in the offered variants of answers (450 test tasks altogether). All tasks refer not only to the individual study materials (Reference), but also to a particular solving process of a particular test task (Help).

In order for the experiment to be carried out properly, it was necessary that the set of tasks corresponded with computerized testing. At the same time, the individual tasks had to be divided into different categories of difficulty – this was necessary in order to fulfill the requirement of adaptivity. On the basis of *formal classification*, the tasks were divided into: automatically inevaluable and automatically evaluable (it is the second group that is suitable for adaptive testing). Based on the taxonomy of learning tasks by Dana Tollingerová (1970) and following a detailed analysis, a *content classification* with five categories, which represent the five levels of difficulty (with level 1 representing the most demanding group), was created. The final set was made of tasks fulfilling both classifications.

Adaptive Testing System

Because computerized adaptive testing cannot be carried out without a suitable system, the already mentioned Barborka 4 LMS was used. The subject Mathematics was incorporated into the system. The subject consists of individual lessons – thematic areas. Each lesson has five frames; each frame has five test tasks of various levels of difficulty. Each task also has:

- Reference with tutorial thanks to which a pupil gains access to the current study material;
- The complete solving process (the so called Help) with the correct answer.

The option of assigning Reaction to half-correct and incorrect answers was not used.

Realization of Experiment

Before the experiment started, several hypotheses were formulated the content of which could be summarized as follows: *How the use of a computerized adaptive system in practice helps increase pupils' level of knowledge*. The individual hypotheses were aimed at finding out the impact of the adaptive testing system on individual thematic areas and thus evaluating the overall influence of the adaptive system on the entire subject (Mathematics).

The Porubská 832 elementary school in Ostrava agreed to take part in the experiment. The school has three 9th grade classes (70 pupils altogether). Eventually, 53 pupils took part in the experiment, 19 of which were boys and 34 were girls. Before the experiment started, each student was assigned a unique identification number under which they went through the entire experiment. The experiment took place near the end of the academic year when all the pupils had already taken the high school entrance exams and their final Mathematics grades could no longer be changed. Therefore, the experiment results could not influence their school grades and they were motivated only by taking part in the experiment, testing the adaptive system and having the opportunity of being part of the process that would enable the next generation of pupils preparing for the high school entrance exams to work with the system.

The experiment consisted of three parts. In the first phase a pre-test was carried out. To verify the efficiency of all six thematic areas, six samples of pre-test were created. Each sample consisted of nine tasks. The first five tasks were of the medium difficulty (group 3) while the last four tasks were of the 1, 2, 4, 5 difficulties. The process of assigning a particular sample of pre-test to a pupil was random.

In the second phase the electronic adaptive system using the Barborka 4 LMS was launched. All pupils who took part in the pre-test were registered in the system. They were instructed on how to work with the system. A thematic area (lesson) on which a pupil focuses in the system is selected according to which of the six samples of pre-test they took. Throughout the course of one week, the pupils had the opportunity to work with the system, test themselves, and improve in the tasks, which they found problematic.

In the third phase a post-test was carried out. Again, six samples we created, which contained nine tasks equivalent to the pre-test tasks (the first five tasks were of the medium difficulty (group 3) while the last four tasks were of the 1, 2, 4, 5 difficulties). The post-test was filled out by the same 53 ninth-grade pupils. The process of assigning a particular sample of post-test to a pupil depended on the pre-test sample.

Experiment Results

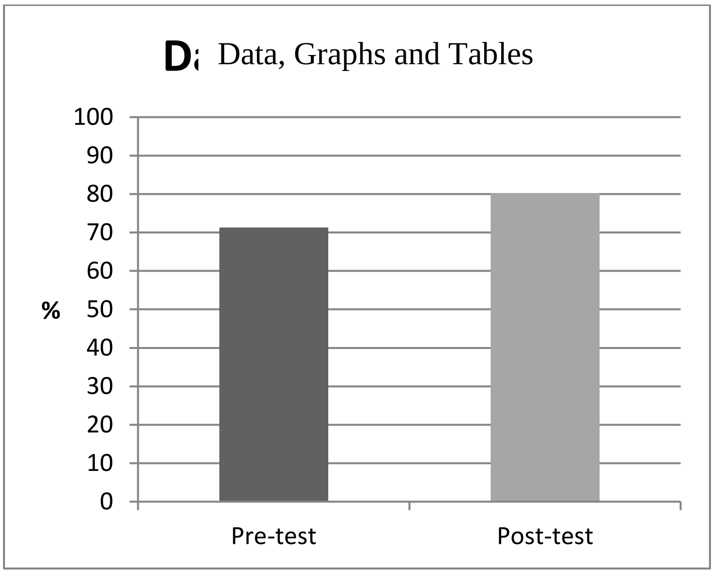
The experiment brought results which were analyzed and processed into tables. Moreover, graphs were added to the tables. A t-test was used to compare the pre-test and post-test results and to find out their statistical importance. The following are the most interesting results which emerged from the experiment. Each hypothesis is supported by a table and the overall pre-test value, which either denies or confirms the respective zero hypothesis.

Data, graphs and tables

As far as the thematic area Data, graphs, and tables is concerned, what is the difference between the pre-test and post-test in the level of knowledge in the experimental group?

H_0: As far as the thematic area Data, graphs, and tables is concerned, there is no statistically important difference between the pre-test and post-test in the level of knowledge in the experimental group.

H_A: As far as the thematic area Data, graphs, and tables is concerned, there is a statistically important difference between the pre-test and post-test in the level of knowledge in the experimental group.



Graph 1: Comparison of Pre-test and Post-test Results

	Post-test	*Pre-test*
Middle value	722,222	641,667
Dispersion	24444,4	37031,3
Observation	9	9
Hypothetical difference of middle values	0	
Difference	8	
t Stat	**2,82675**	
P(T<=t) (1)	0,01113	
t krit (1)	1,85955	
P(T<=t) (2)	0,02226	
t krit (2)	2,306	

Tab. 1: Analysis of Statistical Importance

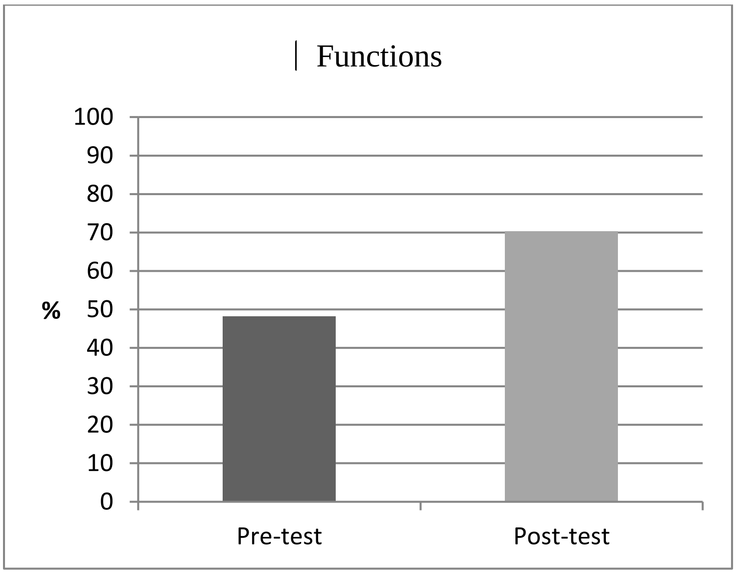
The table shows that out of the maximum success rate of 900, pupils' success rate in the pre-test is 641.667 and in the post-test 722.222. On the basis of the calculated t-test value T = **2.82675**, we can deny the H_0 hypothesis and confirm the alternative H_A hypothesis. Within the scope of the Data, graphs, and tables thematic area, there was a statistically important improvement in pupils' level of knowledge.

Functions

As far as the thematic area Functions is concerned, what is the difference between the pre-test and post-test in the level of knowledge in the experimental group?

H_0: As far as the thematic area Functions is concerned, there is no statistically important difference between the pre-test and post-test in the level of knowledge in the experimental group.

H_A: As far as the thematic area Functions is concerned, there is a statistically important difference between the pre-test and post-test in the level of knowledge in the experimental group.



Graph 2: Comparison of Pre-test and Post-test Results

	Post-test	*Pre-test*

Middle value	633,3	434,329
Dispersion	8816,39	18843,1
Observation	9	9
Hypothetical difference of middle values	0	
Difference	8	
t Stat	**9,75008**	
P(T<=t) (1)	5,1E-06	
t krit (1)	1,85955	
P(T<=t) (2)	1E-05	
t krit (2)	2,306	

Tab. 2: Analysis of Statistical Importance

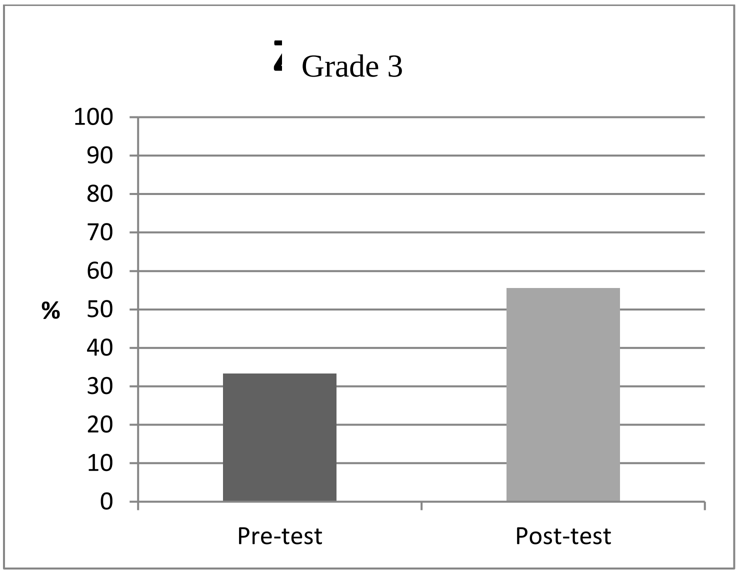
The table shows that out of the maximum success rate of 900, pupils' success rate in the pre-test is 434.329 and in the post-test 633.3. On the basis of the calculated t-test value T = **9.75008**, we can deny the H_0 hypothesis and confirm the alternative H_A hypothesis. Within the scope of the Functions thematic area, there was a statistically important improvement in pupils' level of knowledge.

Grade 3

As far as Mathematics is concerned, what is the difference between the pre-test and post-test in the level of knowledge of the pupils with the mid-year grade 3 in the experimental group?

H_0: As far as Mathematics is concerned, there is no statistically important difference between the pre-test and post-test in the level of knowledge of the pupils with the mid-year grade 3 in the experimental group.

H_A: As far as Mathematics is concerned, there is a statistically important difference between the pre-test and post-test in the level of knowledge of the pupils with the mid-year grade 3 in the experimental group.



Graph 3: Comparison of Pre-test and Post-test Results

	Post-test	*Pre-test*
Middle value	450	323,936
Dispersion	7875	23413,1
Observation	11	11
Hypothetical difference of middle values	0	
Difference	10	
t Stat	**3,47156**	
P(T<=t) (1)	0,003	
t krit (1)	1,81246	
P(T<=t) (2)	0,00601	
t krit (2)	2,22814	

Tab. 3: Analysis of Statistical Importance

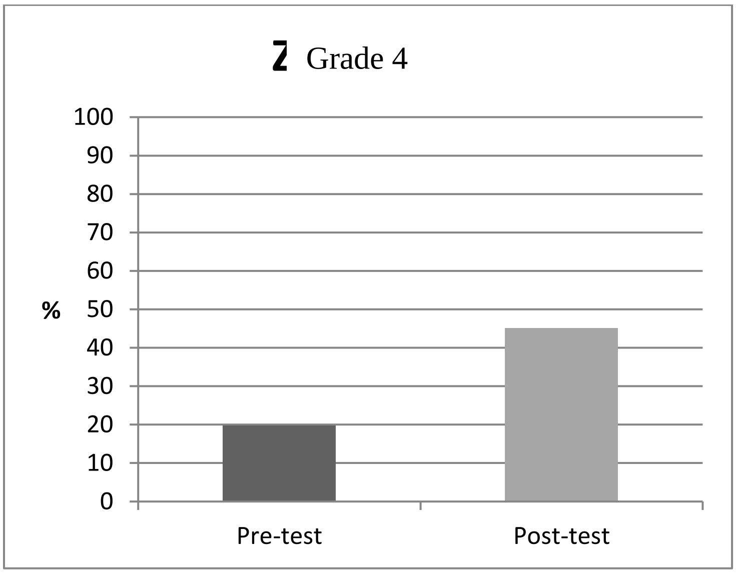
The table shows that out of the maximum success rate of 900, the success rate of the pupils with the grade 3 in the pre-test is 323.936 and in the post-test 450. On the basis of the calculated t-test value T = **3.47156**, we can deny the H_0 hypothesis and confirm the alternative H_A hypothesis. Within the scope of Mathematics, there was a statistically important improvement in the level of knowledge of the pupils with the mid-year grade 3.

Grade 4

As far as Mathematics is concerned, what is the difference between the pre-test and post-test in the level of knowledge of the pupils with the mid-year grade 4 in the experimental group?

H_0: As far as Mathematics is concerned, there is no statistically important difference between the pre-test and post-test in the level of knowledge of the pupils with the mid-year grade 4 in the experimental group.

H_A: As far as Mathematics is concerned, there is a statistically important difference between the pre-test and post-test in the level of knowledge of the pupils with the mid-year grade 4 in the experimental group.



Graph 4: Comparison of Pre-test and Post-test Results

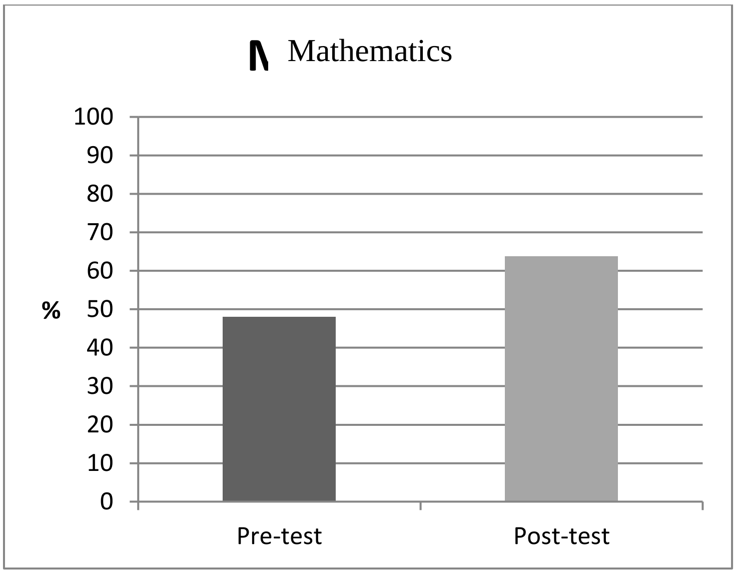
	Post-test	*Pre-test*
Middle value	406,25	177,5
Dispersion	24322,9	21691,7
Observation	4	4
Hypothetical difference of middle values	0	
Difference	3	
t Stat	**16,1123**	
P(T<=t) (1)	0,00026	
t krit (1)	2,35336	
P(T<=t) (2)	0,00052	
t krit (2)	3,18245	

Tab. 4: Analysis of Statistical Importance

The table shows that out of the maximum success rate of 900, the success rate of the pupils with the grade 4 in the pre-test is 177.5 and in the post-test 406.25. On the basis of the calculated t-test value T = **16.1123**, we can deny the H_0 hypothesis and confirm the alternative H_A hypothesis. Within the scope of Mathematics, there was a statistically important improvement in the level of knowledge of the pupils with the mid-year grade 4.

Conclusion

The experiment results show that in each of the six thematic areas there was a statistically important improvement in the level of knowledge of the experimental group pupils. Although the differences between the pre-test and post-test in individual topics differed, in each and every case they denied the zero hypothesis and confirmed the alternative one. As can be seen in the following graph, there is an improvement of the level of knowledge of Mathematics as a whole.



Graph 5: Entire Subject Comparison of Pre-test and Post-test Results

The experiment showed that the proposed algorithm for the Repetition with tutorial mode and the proposed adaptive rules contribute to the improvement of pupils' knowledge. The improvement is most noticeable in the initially less successful pupils.

Positive factors:

- Motivation – Motivate using adaptive testing to better results and reinforcing knowledge (by taking small steps).
- Individuality – Respect for the knowledge level of each student. The student gets the question whose difficulty depends on the responses to the previous task.
- Elimination of stress – The student is not frustrated that he cannot answer any questions.
- Immediate feedback – The provision of immediate feedback (study material – Tutorial or Help).

References

DRÁPELA, Radek. *Systém řízení výuky adaptované podle učebního stylu studenta*. Ostrava, 2013. Diploma thesis. VŠB – Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky, Katedra informatiky.

KOSTOLÁNYOVÁ, Kateřina. *Teorie adaptivního e-learningu.* Ostrava: Ostravská univerzita, 2012. 118 p. ISBN 978-80-7464-014-8.

TAKÁCS, Ondřej. *Automatizované řízení adaptivní výuky v e-learningu podle stylů učení studenta.* Dissertation. VŠB – Technická univerzita. Ostrava, 2014.

TOLLINGEROVÁ, Dana. *Teorie programového učení,* Středisko pro výzkum učebních metod a prostředků, 1970.

PREXTOVÁ, Tatiana. Adaptive Testing of Student's Knowledge Level. In: *Proceedings of the 6th International GUIDE Conference*. Athens, 2013. ISBN 9788897772026.

PREXTOVÁ, Tatiana. Evaluating student's knowledge through the use adaptive testing. In: *Information and Communication Technology in Education*. Ostrava: Repronis Ostrava, 2014, 7 p. ISBN 978-80-7464-561-7. Available from: https://konference.osu.cz/icte/dokumenty/2014/proceedingsICTE2014.pdf