

## Pupils' Summative Assessments in Mathematics as Dependent on Selected Factors

Vlastimil Chytrý<sup>1</sup>, Milan Kubiátko<sup>1\*</sup>

<sup>1</sup> J. E. Purkyne University, Faculty of Education, Department of Preschool & Primary Education, CZECH REPUBLIC

Received 3 May 2021 ▪ Accepted 18 July 2021

### Abstract

School assessment is a determining factor for a pupil in all functions. The aim of the research was to determine correlation between pupils' logical thinking and their school assessment and between pupils' mathematics skills and their school assessment. The sample size was created by 252 high school students (117 boys and 135 girls). As research tools were used reasoning mathematical, logical thinking and mathematical skills tests. Kruskal-Wallis test was used as statistical method. A significant effect was found at the 1% significance level between these factors and school assessment, where the greatest differences are caused mainly by evaluations in mathematics that were perceived either as the best or the worst. The research also monitored the use of the newly-developed GTOLT (Group Test of Logical Thinking), which is based on the GALT (Group Assessment of Logical Thinking) and TOLT (Test of Logical Thinking) tests, rooted in standardized tools.

**Keywords:** logical thinking, mathematics skills, motivation, summative assessment

### INTRODUCTION

It has been shown that school assessment is a determining factor for a pupil in all motivational, informative, regulating, educational, prognostic and differential functions. In the present article and the research described below, school evaluation in mathematics (at the end of a semester) is considered a demonstration of summative evaluation summarizing the pupils' learning results after completion of activity (Brookhart, 2013; Hoover & Abrams, 2013). The formative and summative assessment questions were possible to find out in studies of authors (Mohamadi, 2018; Puddy et al., 2014; Wei, 2014; Wholey, 1996). The summative assessment summarizes the pupils learning outcomes and the aim is to obtain holistic view about pupil's performance. This assessment is providing to pupil at the moment, where is impossible to influence the evaluation. (Sewall & Santaga, 1986). The summative assessment in Czech Republic has got the form (1 - 5), which is similar to (A - E). These authors draw attention to the fact that summative evaluation is often given to the pupils at the moment when they can no longer change their results. It is believed that formative assessment is a much more effective tool for adapting

teaching to help students master the material (Garrison & Ehringhaus, 2007). On this place is appropriate to mention in brief form, what formative assessment is. This concept is realized during the time of learners' activity due to understating how learners are in the process of representation of the concept (Kuh et al., 2014). The formative and also summative assessment serve the similar purposes, but their aims are different. The summative assessment is used for the long time (Greenstein, 2010). Currently, there is no indication that this assessment should change in any way, although it is increasingly criticized. Based on the formative assessment, students can understand the issue and justify why this or that subject is so important (Brophy, 1999). Influencing students' learning based on formative assessment is mentioned by Weurlander et al. (2012) who state that formative assessment: i) promotes motivation, especially in the long run, ii) leads to more precise study by students. The presented article points out the fact that although the summative evaluation is criticized, it is encouraging that the evaluation by teachers marks the pupil's performance, both in terms of the level of his logical thinking and in terms of success in the mathematical diagnostic test. Also Marinho, Leite and Fernandes (2017) summarized the positive effect of

### Contribution to the literature

- The article focuses on the determination of correlation between pupils' logical thinking and their school assessment and between pupils' mathematics skills and their school assessment.
- Pupils with better school results in mathematics achieve higher levels of logical thinking and pupils with better school results in mathematics have better mathematics skills.
- The study provides the spectrum of statistical methods, which could be used on the analysis of presented variables.

summative assessment, when it has a classification and measurement purpose, it simultaneously assumes a formative function.

Also, the fact is, that in the discipline like mathematic is the assessment has got great importance not only for learners and their parent, however also for teachers, policy-makers, curriculum-makers and for future employers (e.g., Boud & Soler, 2016). The similar statement is possible to find in the studies of Brown and Lally (2018), King et al. (2017), Potvin et al. (2020), and Rakoczy et al. (2019). Summative assessment is also questionable in terms of reliability, as teachers must be sure that retesting a student using the same assessment will produce consistent results (Popham, 2014). Summative assessment is considered by pupils themselves to be a satisfactory form of assessment and is preferred over other assessment methods (Rokos et al., 2019). There are many disadvantages to quantitative assessment; pupils may tend to learn "for grades" only, even though this represents an extremely simplified, abstract way of evaluating pupils' performance. Abandoning the practice of quantitative assessment also has also created a certain discourse in the context of inclusive education (Smetáčková, 2018).

The main goal of the paper is to find out if there are differences to be found in school assessment outcomes in relation to pupils' logical thinking and mathematics skills. Within the skills for the 21st century it is necessary to use different kinds of thinking (inductive reasoning, deductive, etc.). Liu, Ludu and Holton (2015) support this view and consider valid logical reasoning to be the key element of healthy critical thinking. Although the reasoning in mathematics differs significantly from day-to-day reasoning (Bronkhorst et al., 2020), the rationale in a mathematical proof is not just a formal procedure, but also involves discussion, exploration and examination (Bronkhorst et al., 2020) and shows the need for a more informal method of solving formal reasoning tasks. Widana et al. (2018) analyzed the effectiveness of thinking skills assessment towards critical thinking skills of high school students in mathematics lesson. Authors used quasi-experimental design and the results showed, that thinking skills assessment could improve student's critical thinking skills in mathematics lesson effectively. The similar results were possible to find in the study of Ramdani, Syamsuddin and Sirajuddin (2019).

It has been shown that the key problem in research on logical thinking is certain inconsistencies in its definition, where different definitions and interpretations were provided by Albrecht (1984), Chuechote et al. (2020), and Labouvie (1992). Neither definition covers logical thinking in the way it has been understood for the purposes of this paper. We have thus created a new definition, combining these two: Logical thinking is a process in which an individual looks away from the contents of individual statements and thoroughly exploits his or her own judgements to ensure correct conclusions. The intermediate, individual steps of this process form a relation between preconditions and conclusions by demonstrating their connection to the judgements. With regard to these findings, other authors show that the development of formal argumentation should be the key priority in science education (DeCarcer et al., 1978; Lawson, 1982). This is an important issue because, for instance, logical reasoning creates a certain style of thinking that affects pupils' ability to solve tasks in physics. This claim is based in particular on the fact that tasks in physics include problems, which, need various types of logical reasoning, mathematical operation and experiment, respectively, and these then promote the formation of learners' thinking (Korsun, 2019). Logical thinking also helps the pupil to work with interactive multimedia which can improve skills for more advanced thinking (Hartini et al., 2017). In order to use the potential of logical thinking, efforts have been made to quantify it, such as described, for example, in Fadiana et al. (2019).

Cresswell and Speelmann (2019) focused on the effect of logical thinking to students' achievement in mathematics. Authors declared positive relationship between these two variables. The similar results were possible to find in the research of Jeon and Park (2014) and also Sartika and Fatmanissa (2020). Insorio and Librada (2021), and Yang and Chang (2013) demonstrated significant improvements in critical thinking skills, and academic achievement. Ali (2010) investigated the difference between academic achievement of students who have high critical thinking dispositions and of students who have low critical thinking disposition and whether this difference change with students' gender. Result of this study, there is no statistically significant difference between students' academic achievement and critical thinking. Jawad,

Maiwall and Hussein (2019) focused on the effect of logical thinking on the achievement of biology students with respect to gender. Authors used experimental design and summarized, that there is no effect in achievement with respect to gender. The research studies regarding to relationship mathematical skills and pupils school assessment are lower frequent. Some findings is possible to read in the study of Finn et al. (2014) with the positive relationship between these two variables. Similar results are possible to find in research works of authors Gonda and Tirpakova (2018, 2021), Watt et al. (2014). Guo and Yan (2019) found out negative affective attitudes towards summative assessment. Girls had more positive instrumental attitudes towards this kind of assessment than boys. It was found that students' affective and instrumental attitudes to formative assessment positively predicted students' affective and instrumental attitudes to summative assessment. As it is possible to observe number of research studies related to presented topic is little bit narrow. This statement is supported by other studies like Martinovic and Manizade (2018), Nortvedt and Buchholtz (2018), and Ubuz and Aydin (2018).

## PROBLEM OF RESEARCH

The research problem was to determine correlation between school assessment and pupils' logical thinking and mathematics skills and research problem is defined in the research questions. The research questions are rooted in our belief that assessment should provide opportunities for all pupils to prove their mathematical skills and should respond to a variety of students (Klieme et al., 2004; Klinger et al., 2015). Within these research questions, the issues of logical thinking and school evaluation in mathematics are also emphasized. The aim of the research was to find the answers to the following research questions:

RQ1: What is the correlation between pupils' logical thinking and their school assessment?

RQ2: What is the correlation between pupils' mathematics skills and their school assessment?

The research problems are linked to two hypotheses:

H1: Pupils with better school results in mathematics achieve higher levels of logical thinking.

H2: Pupils with better school results in mathematics have better mathematics skills.

## METHOD

### Sample of Research

The research was carried out with pupils aged 13 – 16 ( $M = 15$ ,  $SD = 0.49$ ). The main method used was a questionnaire and further partial methods (e.g., based on

qualitative research). The main target was to process data collected from a total of 252 respondents (117 boys and 135 girls). In terms of gender division, the sample is balanced. The actual test tools were distributed to more than 300 pupils from randomly selected classes in the Czech Republic, and without this selection, it would not be possible to verify the statistical level of significance. The data was collected by university students who were informed in detail about the individual research steps. The research involved an evaluation from the end of the last semester. The obtained values were first entered into Excel 2015 (Microsoft, 2016) and then transferred to Statistica 12 (Statsoft, 2016). The basic unit of the research sample was not individual pupils but whole school classes. Choosing the multi-stage random selection in which a pupil would be the basic unit would have been very time-consuming and very difficult to organize. Due to the fact that the main data collection was carried out by means of a questionnaire and takes about two hours for each pupil. As a part of the data analysis itself, there was a reduction in case the student did not answer the question. It means, if student did not fill whole research tool, he/she was eliminated from further analysis. In the case of the school evaluation, it was not possible to take into consideration evaluation 5 as it was achieved by almost none of the students.

### Statistical Methods Used

A detailed descriptive and frequency analysis was preceded by cleaning the data. When outliers were identified, for instance, using a quartile graph, it is preferable (rather than using very rigorous statistical methods) to examine why outliers have occurred at all. If outliers were identified, it was checked whether there had been a measurement error. Given the size of the sample outlier values can always be expected. It is not possible to proceed mechanically when removing outliers or extreme values.

A large variety of data was acquired. It was necessary to select dependent and independent samples, nominal, ordinal and metric random variables, to assess normality in the metric variables and thus make a choice between parametric or non-parametric statistical methods. The following statistical methods and techniques were used throughout analysis of the survey:

- Normality test
- Non-parametric hypothesis testing
- Non-parametric dispersion analysis followed by post hoc analysis (multiple observations).

The data analysis was performed with the Robinson and Levin (1997) two-step model. As a guideline for assessing the significance of the results, statistically unimagined by the range of the analyzed data set, effect size coefficients were used (Cohen, 1988; Sheskin, 2003; Thomas & Nelson, 2001). These coefficients eliminate the influence of positive dependence on statistical

**Table 1.** Characteristics of some GTOLT items

Items	Dimensions	Success	Difficulty
Item 1	Searching for numerical laws	0.57	0.43
Item 2	Disjunction	0.46	0.54
Item 3	Implication	0.82	0.18
item 4	Reverse sentence	0.78	0.22
Item 5	Conjunction	0.81	0.19
Item 6	General quantifier	0.72	0.28
Item 7	General quantifier negation	0.21	0.79
Item 8	Zebra type task	0.78	0.22
Item 9	Conjunction negation	0.19	0.81
Item 10	General quantifier negation	0.35	0.65
Item 11	Existential quantifier negation	0.46	0.54
Item 12	Volume preservation	0.60	0.40
Item 13	Combinatory	0.17	0.83
Item 14	Volume preservation	0.42	0.58
Item 15	Proportions	0.17	0.83
Item 16	Control variable	0.05	0.95
Item 17	Control variable	0.24	0.76
Item 18	Probability	0.08	0.92
Item 19	Probability	0.08	0.92
Item 20	Combinatory	0.49	0.51

significance on the sample size (Rosenthal, Rosnow, & Rubin, 2000). Eta squared ( $\eta^2$ ) and Cohen's  $d$  were used to measure effect size.

Other statistical method was regression analyses with summative assessment as independent variable and achievement from test of didactic test from mathematics and test of logical thinking as dependent variables. The regression analyses was performed with model ENTER.

## Instruments

### Logical thinking test

The authors choose research tools, which were possible in the form of pencil-paper. (Lawson, 1978; Tobin & Capie, 1980). The preference tools are GALT a TOLT. The research tool TOLT is focused on the formative operative thinking (Piaget & Inhelder, 1955), where every item is paired and answer have to be justified, because justification is needed for understanding (Weber et al., 2014). The research tool GTOLT is focused primarily on the academic achievement in the connection with formal justification. This cognitive dimension was in last two decades in marginal awareness of researchers. In our previous research (Chytry, 2015) we described pilot testing of the tool GALT (Group Assessment of Logical Thinking) for the Czech Republic, it is important to notice, the study was in Czech language. Our current research includes features of the TOLT (Test of Logical Thinking) as it was presented in important journals abroad. We are obliged to mention that not only GALT and TOLT were used, as they do not cover the full scope of logical thinking tests in the way we understand the problem. The newly-created test is labelled GTOLT (Group Test of Logical

Thinking). The presented test includes 20 items which focused on various attributes of logical thinking. The answers to the questions were assessed dichotomously, 0 – if the pupil's answer was incorrect, and 1 – if the pupil's answer was correct. If the pupil did not answer a question, an empty set was used in coding. This type of coding allowed the results to be interpreted in the following way: the arithmetic average of the measured values is an adequate point estimation of parameter  $p$  of alternative distribution, which is the probability that a randomly selected pupil would answer the question correctly. Some of the test entries required pupil's reasoning. In such cases, the pupil scored only on the condition that their answer, as well as the reasoning, was correct. The entries given were then evaluated as a unique set, including the reasoning. The maximum score was 20 points. The reliability was determined by two ways: a) Split-Half with Spearman-Brown Adjustment (0.68), b) KR20 = 0.71. The research tool is reliable. Some basic characteristics of research tool are presented in Table 1.

The problematic items were 16, 18 a 19, but these items were left in analyses after consultation with experts in the field of didactics of mathematics due to two reasons: a) the work with control variable and probability is lack during teaching of mathematics; b) the research tool was created due to as unit was matched with the logical thinking definition. The research tool was in initial phase piloted by 30 respondents due to content validity. On the basis of their results and comments the tasks were revised mainly in stylistic way.

### Testing mathematics skills

To analyze the mathematics skills of a pupil, we used a freed CERMAT (Centre for High School Graduation



**Table 2.** Item analysis of mathematical skills test

Items	Success	Difficulty	Items	Success	Difficulty
Item 1	0.58	0.42	Item 15	0.60	0.40
Item 2	0.88	0.12	Item 16	0.66	0.34
Item 3	0.51	0.49	Item 17	0.54	0.46
Item 4	0.40	0.60	Item 18	0.70	0.30
Item 5	0.70	0.30	Item 19	0.50	0.50
Item 6	0.27	0.73	Item 20	0.83	0.17
Item 7	0.54	0.46	Item 21	0.34	0.66
Item 8	0.36	0.64	Item 22	0.26	0.74
Item 9	0.44	0.56	Item 23	0.51	0.49
Item 10	0.33	0.67	Item 24	0.68	0.32
Item 11	0.46	0.54	Item 25	0.29	0.71
Item 12	0.47	0.53	Item 26	0.42	0.58
Item 13	0.86	0.14	Item 27	0.15	0.85
Item 14	0.40	0.60	Item 28	0.16	0.84

**Table 3.** Pupils' success rate in the GTOLT in relation to school performance evaluation

Used tests	Points in GTOLT				Mathematics test			
School performance assessment	A (1)	B (2)	C (3)	D (4)	A (1)	B (2)	C (3)	D (4)
Arithmetic mean	9.84	9.51	7.15	6.65	15.18	12.38	9.26	7.50
Median	8.75	9.90	8.00	6.50	15.00	12.00	9.00	6.00
Mode	7.00	6.00	8.00	5.00	12.00	12.00	7.00	6.00
Standard deviation	4.10	3.21	2.99	2.80	6.95	5.66	4.31	4.29
Maximum	19.00	16.00	15.00	14.00	26.00	26.00	24.00	23.00
Minimum	2.00	1.00	1.00	0.00	1.00	2.00	3.00	0.00
<i>p</i> -level (Shapiro-Wilk test normality)	0.02	.01	<.01	.12	.03	<.01	<.01	.02

Exam Reform - test M9PID15C0T01)1 test. The test involves 17 tasks, and the equipment allowed included pens and geometry instruments. The test covers three general topics: i) number and variable, ii) processing data, dependences and relations, iii) geometry (computational and constructional geometry on the planar and spatial levels). The original evaluation of the CERMAT test allowed the respondent to score as many as 50 points; we adapted the assessment so that each answer had a value of one point if the answer was correct and no points if the answer was incorrect. The respondents in our testing were able to score up to 28 points. This is not the school evaluation mentioned above (on the scale A - E), but success in the didactic test in mathematics. The reliability was tested by two ways a) Split-Half with Spearman-Brown Adjustment (0.92), b) KR20 = 0.91. The research tool is reliable. The item analysis was realized due to basic psychometric characteristics of the test. Table 2 includes basic characteristics, as it is possible to see some items were out of standard level of difficulty.

The tasks were used from the standardized test and the items were tested on the respondents of similar age as were used in this research. The coding was dichotomy (0 - incorrect answer; 1 - correct answer). The authors are aware by limitations of this method, but for the conception of this study it is the best way.

## RESULTS

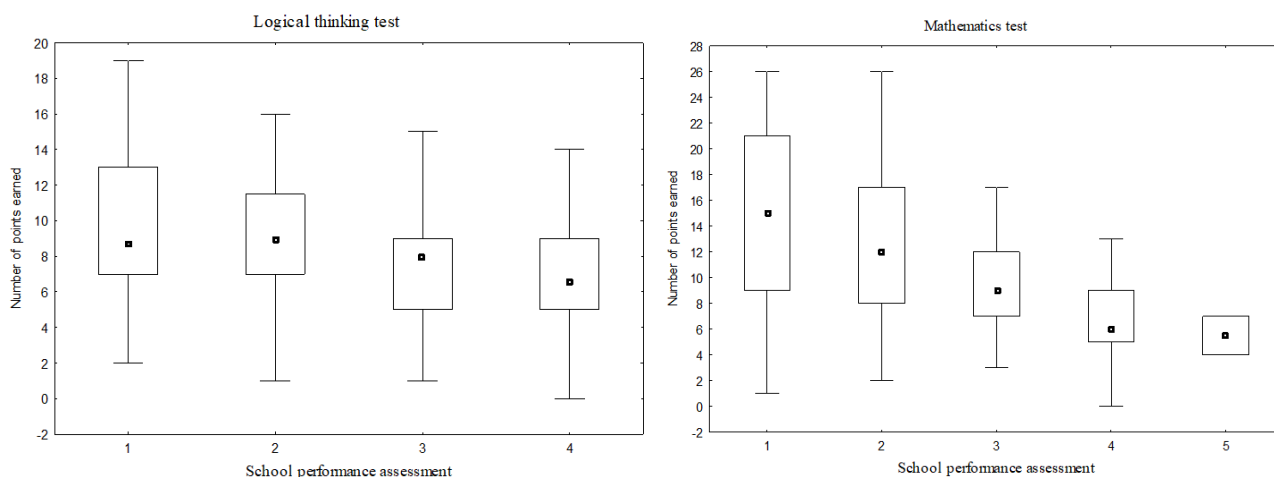
As already mentioned above, the respective tables and analyses shown below do not cover the fifth grade of assessment (E), as only a low number of respondents scored that particular grade. On the basis of descriptive analysis (Table 1) we may notice a trend in the following outcomes: the less satisfactory the school performance (evaluated at level 1 - 4 or A - D), the lower a pupil's success rate within the GTOLT and the poorer the mathematics skills. The correlation matrices are presented below. The relevant *p*-level values for the Shapiro-Wilk normality test are also added to Table 3.

We tested the hypothesis H1 a H2 through the use of the Kruskal-Wallis test. The general formulation of the null hypothesis in this case is: the medians are identical, separate school performance assessments for both tools. The *p*-level found for logical thinking (Kruskal-Wallis test:  $H(3, N = 233) = 31.92, p < 0.001$ ) shows that dependence is significant, at the 1% level. In the case of effect size, the results are:  $\eta^2 = 0.14$  and  $d_{Cohen} = 0.80$ , therefore we conclude that we notice a large effect here (Cohen, 1988). We found similar conclusions with mathematics skills, as the values measured were (Kruskal-Wallis test:  $H(3, N = 237) = 54.75, p < 0.001$ ), meaning the values are  $\eta^2 = 0.28$  and  $d_{Cohen} = 1.24$ . Also here we may state that it is of large effect. It is interesting to compare the results of the post-hoc analysis of both tests.

**Table 4.** Post hoc analysis to determine dependence of efficiency in the GTOLT together with mathematics skills and school performance assessment

		School performance assessment				School performance assessment			
		A (1)	B (2)	C (3)	D (4)	A (1)	B (2)	C (3)	D (4)
A (1)			$p = 0.82$	$p > 0.05$	$p > 0.001$		$p = 0.90$	$p > 0.001$	$p > 0.001$
B (2)	$p = 0.89$			$p > 0.01$	$p > 0.001$	$p = 0.90$		$p > 0.05$	$p > 0.001$
C (3)	$p > 0.05$	$p > 0.01$			$p = 0.69$	$p > 0.001$	$p > 0.05$		$p = 0.19$
D (4)	$p > 0.001$	$p > 0.001$	$p = 0.69$			$p > 0.001$	$p > 0.001$	$p = 0.19$	

Note: Values in bold are statistically significant.



**Figure 1.** The dependence of logical thinking and mathematical skills on school evaluation in mathematics

**Table 5.** Values of regression analysis

$R = 0.53$			
$F = 46.47 ; p < 0.001$			
	Coef.	$p$ value	$t$
K	4.11	$p < 0.001$	25.00
$X_1$	-0.07	$p < 0.001$	6.56
$X_2$	-0.08	$p < 0.001$	4.41

Table 2 shows that in terms of statistical significance, the findings are identical for both studied domains (logical thinking and mathematical skills). In both cases, when the two best ratings (1 - 2 / A - B) and the two worst ratings (3 - 4 / C - D) are compared, there is no statistically significant difference between them. In all other cases the differences are statistically significant.

The figure shows that the worse a pupil’s summative school assessment in mathematics is, the less successful the pupil appears to be in the logical thinking test as well as in the mathematics skills test.

Other way is to use correlation and regression analyses, where the variable summative school assessment was as dependent variable and other variables had character of independent variables following: a) achievement in didactic test from mathematics; b) achievement in logical thinking test. The values of regression analysis are presented in Table 5.

From Table 3 was possible to read, that all variables had got the significant effect. The next step was to calculation of Spearman correlation coefficient ( $\rho$ ). The  $p$  value between summative assessment and achievement

in didactic test from mathematic was  $\rho = -0.48$  ( $p < 0.001$ ) and between summative assessment and achievement of logical thinking test was  $\rho = 0.37$  ( $p < 0.001$ ).

## DISCUSSION

The first research question is rooted in the assumption that assessment in Maths is an adequate tool for monitoring pupils’ skills in solving mathematics problems (Rosli et al., 2013). The second research question deals with a similar issue: the perspective of mathematics skills also in relation to school performance assessment, which should consider both the content of and the procedures for solving problems, as suggested in NCTM (2014), and Pellegrino et al. (2001). We claim that both our hypotheses have been confirmed (H1: Pupils with better school results in mathematics achieve higher levels of logical thinking H2: Pupils with better school results in mathematics have better mathematics skills) and pupils’ assessments in mathematics reflect their logical thinking skills together with their mathematics skills, even under the condition that this type of evaluation is implemented for monitoring

reasons only (e.g., in Parke et al., 2003). Also, the similar findings are possible to find in the study of Firdaus et al. (2015), positive effects of use mathematical learning module on problem-based learning to enhance learning skills. The pupils have better skills to argument and to evaluate (Pitchford et al., 2016). The mean values / medians matching the individual grades in education in both testing procedures (Table 1) are as follows: grade A (1) (9.84/8.75), grade B (2) (9.51/9.90), grade C (3) (7.15/8.00), grade D (4) (6.65/6.50); and for mathematics skills - grade A (1) (15.18/15.00), grade B (2) (12.38/12.00), grade C (3) (9.26/9.00), grade D (4) (7.50/6.00), suggesting that with a higher grade (lower evaluation) the pupil's ability to think logically decreases along with their mathematics skills. This is not a surprising conclusion. What is interesting is the diversification itself and the possibilities of generalizing the conclusions with respect to the statistically significant differences. Based on the research, it can be said that pupils are (in terms of school evaluation in mathematics) divided by teachers into two groups, where the first group consists of those who have grades A and B and the second group, those with grades C or D. Taking Gauss's normal distribution of frequency into consideration, it could be expected that the same differences as between pupils with A and B grades will be found between all pairs. However, the research shows that differences are more likely to be between pupils with B and C, and between any other groups (meaning the groups where the difference in school evaluation is one grade). This fact could be supported by the theory of average pupils in summative assessment have got better logical thinking skills as the best and worst pupils. This statement is supported by study of Eckstein and Shemesh (1989) and it is valid nowadays. Also, Simonton (1992) stated, that the success in mathematics is not so important in the future career of pupils, but the level of logical thinking skills is predicting factor of their career and success in life.

The research shows that pupils belong to two diverse groups, where the first group are assessed with A and B and the latter group are assessed with C or D. From the point of view of logical thinking, the conclusions are in line with earlier research that we tested on elementary school respondents (in total 162 respondents), 108 grammar school students, and 23 students of different secondary schools (Chytrý, 2015). We saw that a pupil's ability to find numeral rules and the laws of geometry, as well as correct judgement in terms of working with logical connectives, depends on their assessment (grade). It means, cognitive level to understand these concepts was on higher level, when the summative assessment was better. Such an outcome was not applicable to students outside elementary schools and eight-year grammar schools. Zaman (2011) also reached the same conclusion in the field of mathematics skills. He carried out research which found a correlation between

the mathematical thinking of pupils and their success rate in their school performance assessment in mathematics. The testing involved a specially designed test focusing on mathematical thinking in pupils. The sample involved 500 randomly selected respondents. The statistics analysis was realised through regressive correlation techniques. A strong dependence was revealed between mathematical thinking and school assessment in the subject of mathematics.

## CONCLUSION

The most significant limitation of our research is the scope of the sample ( $n = 252$ ). The size of the base set does not guarantee strong representativeness. Maximum effort was put into balancing proportions within the sample participating, therefore effect size eliminating the size of the sample was used. The testing involved a minimum of pupils qualifying as "E" - failing. For this reason, these students were excluded from the data matrix. Other limitation is the answers of respondents, it is very hard to distinguish between skills and if respondent remembered the correct answer or incorrect answer was caused only by single mistake. These difficulties in the evaluation of data are common for research typical in social studies.

As it is seen in results part and also in Discussion there is positive correlation between summative assessment and level of logical thinking. So, for the all representatives of educational environment should be focus on the process how to improve the level of logical thinking among learners. The logical thinking is in the connection with the creativity level of every person, so the teachers in every school grade should focus in the teaching process on the developing of creativity and logical thinking. More concretely, it is possible to use different educational games, which support divergent thinking, creativity and as it was mentioned above also logical thinking. Also, to focus on the reflection of the connection theory with practice. In the many schools in Czech Republic the learning process is focused on the theoretical presentation of topic without any practical connection.

Also, the mathematics skills are in narrow relationship with summative assessment in mathematics. The ways, how to improve mathematics skills among learners are similar like are presented in previous paragraph. For many learners is difficult to find and understand the connection of mathematics with real life, so the higher amount of tasks, which are connected with everyday problems could lead to developing of mathematical skills and then to better summative assessment. There are others forms, how to improve mathematics skills, for example to focus on the misunderstandings in mathematics to their complete understanding, to offer more than one possible manners of solutions etc.

**Author contributions:** All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

**Funding:** This article was supported by the J. E. Purkyne University in Usti nad Labem, internal grant no. UJEP-SGS-2021-43-003-2.

**Declaration of interest:** No conflict of interest is declared by authors.

## REFERENCES

- Albrecht, K. (1984). *Brain building: Easy games to develop your problem-solving skills*. Prentice-Hall Press.
- Ali, A. Z. A. R. (2010). The effect of critical thinking dispositions on students' achievement in selection and placement exam for university in Turkey. *Journal of Turkish Science Education*, 7(1), 61-73.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413. <https://doi.org/10.1080/02602938.2015.1018133>
- Bronkhorst, H., Roorda, G., Suhre, C., & Goedhart, M. (2020). Logical reasoning in formal and everyday reasoning tasks. *International Journal of Science and Mathematics Education*, 18, 1673-1694. <https://doi.org/10.1007/s10763-019-10039-8>
- Brookhart, S. M. (2013). The Use of Teacher Judgement for Summative Assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69-90. <https://doi.org/10.1080/0969594X.2012.703170>
- Brophy, J. (1999). Toward a model of the value aspects of motivation in education: developing appreciation for. *Educational Psychologist*, 34(2), 75-85. [https://doi.org/10.1207/s15326985ep3402\\_1](https://doi.org/10.1207/s15326985ep3402_1)
- Brown, K., & Lally, V. (2018). Rhetorical relationships with students: A higher education case study of perceptions of online assessment in mathematics. *Research in Comparative and International Education*, 13(1), 7-26. <https://doi.org/10.1177/1745499918761938>
- Chuechote, S., Nokkaew, A., Phongsasithorn, A., & Laosinchai, P. (2020). A neo-piagetian analysis of algorithmic thinking development through the "sorted" digital game. *Contemporary Educational Technology*, 12(1), ep261. <https://doi.org/10.30935/cet.685959>
- Chytry, V. (2015). *Logika, hry a myšlení* [Logic, games and thinking]. Usti nad Labem: Univerzita J.E. Purkyne v Usti nad Labem.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Cresswell, C., & Speelman, C. P. (2020). Does mathematics training lead to better logical thinking and reasoning? A cross-sectional assessment from students to professors. *PloS One*, 15(7), e0236153. <https://doi.org/10.1371/journal.pone.0236153>
- DeCarcer, I. A., Gabel, D. L., & Staver, J. R. (1978). Implications of Piagetian research for high school science teaching: A review of literature. *Science Education*, 62, 571-583. <https://doi.org/10.1002/sce.3730620417>
- Eckstein, S. G., & Shemesh, M. (1989). Development of children's ideas on motion: Intuition vs. logical thinking. *International Journal of Science Education*, 11(3), 327-336. <https://doi.org/10.1080/0950069890110309>
- Fadiana, M., Amin, S. M., Lukito, A., Wardhono, A., & Aishah, S. (2019). Assessment of seventh grade students' capacity of logical thinking. *Indonesian Journal of Science Education*, 8(1), 75-80. <https://doi.org/10.15294/jpii.v0i0.11644>
- Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., Sheridan, M. A., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science*, 25(3), 736-744. <https://doi.org/10.1177/0956797613516008>
- Firdaus, F., Kailani, I., Bakar, M. N. B., & Bakry, B. (2015). Developing critical thinking skills of students in mathematics learning. *Journal of Education and Learning*, 9(3), 226-236. <https://doi.org/10.11591/edulearn.v9i3.1830>
- Garrison, C., & Ehringhaus, M. (2007). *Formative and summative assessments in the classroom*. Association for Middle Level Education.
- Gonda, D., & Tirpakova, A. (2021). An alternative approach to pupils' acquiring a concept parameter in solving inequalities in school mathematics. *International Journal of Mathematical Education in Science and Technology*, 52(2), 178-193. <https://doi.org/10.1080/0020739X.2019.1670365>
- Gonda, D., & Tirpakova, A. (2018). A new teaching method aimed at eliminating the causes of students' unsuccessful algorithmic problem solving with parameter. *Problems of Education in the 21st Century*, 76(4), 499-519. <https://doi.org/10.33225/pec/18.76.499>
- Greenstein, L. (2010). *What teachers really need to know about formative assessment*. ASCD.
- Guo, W. Y., & Yan, Z. (2019). Formative and summative assessment in HongKong primary schools: students' attitudes matter. *Assessment in Education: Principles, Policy & Practice*, 26(6), 675-699. <https://doi.org/10.1080/0969594X.2019.1571993>
- Hartini, S., Misbah, M., Dewantara, D., Oktovian, R. A., & Aisyah, N. (2017). Developing learning media using online prezi into materials about optical equipment. *Indonesian Journal of Science Education*, 6(2), 313-317. <https://doi.org/10.15294/jpii.v6i2.10102>



- Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education*, 26(3), 219-231. <https://doi.org/10.1080/08957347.2013.793187>
- Insorio, A. O., & Librada, A. R. P. (2021). Enhancing mathematical critical thinking and problem-solving skills through emergentics® as a grouping mechanism. *Contemporary Mathematics and Science Education*, 2(1), ep21002. <https://doi.org/10.30935/conmaths/9289>
- Jawad, H. Q., Majwall, M., & Hussien, A. (2019). The effect of teaching on the skills of logical thinking in the achievement of student's achievement in biology. *Indian Journal of Public Health Research & Development*, 10(6), 1125-1129. <https://doi.org/10.5958/0976-5506.2019.01440.2>
- Jeon, S., & Park, J. H. (2014). Analysis of relationships of scientific communication skills, science process skills, logical thinking skills, and academic achievement level of elementary school students. *Journal of the Korean Association for Science Education*, 34(7), 647-655. <https://doi.org/10.14697/jkase.2014.34.7.0647>
- King, D., Varsavsky, C., Belward, S., & Matthews, K. (2017). Investigating students' perceptions of graduate learning outcomes in mathematics. *International Journal of Mathematical Education in Science and Technology*, 48(1), 67-80. <https://doi.org/10.1080/0020739X.2017.1352044>
- Klieme, E., Avenarius, H., Blum, W., & Döbrich, P. (2004). *The development of national educational standards: An expertise*. Federal Ministry of Education and Research.
- Klinger, D. A., McDivitt, P. R., Howard, B. B., Munoz, M. A., Rogers, W. T., & Wylie, E. C. (2015). *The classroom assessment standards for PreK-12 teachers*. Kindle Direct Press.
- Korsun, I. (2019). The forming of learners' interest towards science using inventive tasks in physics teaching. *Physics Education*, 54(2), 025014. <https://doi.org/10.1088/1361-6552/aaff91>
- Kuh, G., Jankowski, N., Ikenberry, S., & Kinzie, J. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in us colleges and universities*. University of Illinois; Indiana University: National Institute for Learning Outcomes Assessment (NILOA).
- Labouvie, G. (1992). A neo-piagetian perspective on adult cognitive development. In R. J. Sternberg & C.A Berg (Eds.), *Intellectual development*. Cambridge University Press.
- Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1), 11-24.
- Lawson, A. E. (1982). Formal reasoning, achievement, and intelligence: An issue of importance. *Science Education*, 66(1), 77-83.
- Liu, H., Ludu, M., & Holton, D. (2015). Can K-12 math teachers train students to make valid logical reasoning? *Emerging Technologies for STEAM Education*, 331-353. [https://doi.org/10.1007/978-3-319-02573-5\\_18](https://doi.org/10.1007/978-3-319-02573-5_18)
- Marinho, P., Leite, C., & Fernandes, P. (2017). Mathematics summative assessment practices in schools at opposite ends of performance rankings in Portugal. *Research in Mathematics Education*, 19(2), 184-198. <https://doi.org/10.1080/14794802.2017.1318085>
- Martinovic, D., & Manizade, A. G. (2018). The challenges in the assessment for knowledge for teaching geometry. *ZDM Mathematics Education*, 50(4), 613-629. <https://doi.org/10.1007/s11858-018-0934-4>
- Mohamadi, Z. (2018). Comparative effect of online summative and formative assessment on EFL student writing ability. *Studies in Educational Evaluation*, 59, 29-40. <https://doi.org/10.1016/j.stueduc.2018.02.003>
- National Council of Teachers of Mathematics (NCTM). (2014). *Principles to action: Ensuring mathematical success for all*. National Council of Teachers of Mathematics.
- Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM Mathematics Education*, 50(4), 555-570. <https://doi.org/10.1007/s11858-018-0963-z>
- Parke, C., Lane, S., Silver, E. A., & Magone, M. (2003). *Using assessment to improve mathematics teaching and learning: Suggested activities using quasar tasks, scoring criteria, and student work*. National Council of Teachers of Mathematics.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science of design and educational assessment*. National Academy Press.
- Piaget, J., & Inhelder, B. (1955). *De la logique de l'enfant à la logique de l'adolescent* [From the logic of the child to the logic of the adolescent]. Presses Universitaires de France.
- Pitchford, N. J., Papini, C., Outhwaite, L. A., & Gulliford, A. (2016). Fine motor skills predict maths ability better than they predict reading ability in the early primary school years. *Frontiers in Psychology*, 7, 783. <https://doi.org/10.3389/fpsyg.2016.00783>
- Popham, W. J. (2014). *Classroom Assessment: What Teachers Need to Know* (7th ed.). Pearson Education.
- Potvin, P., Hasni, A., Ayotte-Beaudet, J. P., & Sy, O. (2020). Does individual interest still predict achievement in science and technology when

- controlling for self-concept? A longitudinal study conducted in Canadian schools. *Eurasia Journal of Mathematics Science and Technology Education*, 16(12), em1904. <https://doi.org/10.29333/ejmste/8938>
- Puddy, R. W., Boles, R. E., Dreyer, M. L., Maikranz, J., Roberts, M. C., & Vernberg, E. M. (2007). Demonstrating support for the formative and summative assessment paradigm in a school-based intensive mental health program. *Journal of Child and Family Studies*, 17(2), 253–263. <https://doi.org/10.1007/s10826-007-9164-z>
- Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, 60, 154–165. <https://doi.org/10.1016/j.learninstruc.2018.01.004>
- Ramdani, R., Syamsuddin, A., & Sirajuddin, S. (2019). Development of mathematical module-problem solving approach to train student's reflective thinking. *Pedagogical Research*, 4(4), em0040. <https://doi.org/10.29333/pr/5861>
- Robinson, D., H., & Levin, J. R. (1997). Reflections on statistical and substantive significance with a slice of replication. *Educational Researcher*, 26(5), 21–27. <https://doi.org/10.3102/0013189X026005021>
- Rokos, L., Lišková, J., Váchová, L., Cihlářová, M., Chadová, M., & Strapková, J. (2019). Pohled žáků vybraných základních škol a gymnázií na hodnocení při hodinách přírodopisu a biologie s akcentem na formativní hodnocení [The view of pupils of selected elementary schools and grammar schools on the evaluation during the lessons of natural history and biology with an accent on the format evaluation]. *Scientia in Educatione*, 10(1), 90–112. <https://doi.org/10.14712/18047106.1234>
- Rosenthal, R., Rosnow, R.L., & Rubin, D.B. (2000). *Contrasts and effect sizes in behavioral research: a correlational approach*. Cambridge University Press.
- Rosli, R., Goldsby, D., & Capraro, M. M. (2013). Assessing students' mathematical problem-solving and problem-posing skills. *Asian Social Science*, 9(16), 54–60. <https://doi.org/10.5539/ass.v9n16p54>
- Sartika, I., & Fatmanissa, N. (2020). Analysis of students' error in solving trigonometric function problems which assess higher order thinking skills. *Contemporary Mathematics and Science Education*, 1(1), ep20002. <https://doi.org/10.30935/conmaths/8437>
- Sewall, T., & Santaga, M. (1986). *A reference guide to program evaluation in adult education*. Wisconsin Assessment Centre, University of Wisconsin-Green Bay.
- Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. Chapman and Hall/CRC.
- Simonton, D. K. (1992). The social context of career success and course for 2,026 scientists and inventors. *Personality and Social Psychology Bulletin*, 18(4), 452–463. <https://doi.org/10.1177/0146167292184009>
- Smetáčková, I. (2018). Obliba školní matematiky a její souvislost s externím hodnocením a sebehodnocením [The popularity of school mathematics and its connection with external evaluation and self-evaluation]. *Scientia in Educatione*, 9(2), 44–56. <https://doi.org/10.14712/18047106.1049>
- Thomas, J. R., & Nelson, J. K. (2001). *Research methods in physical activity* (4th ed.). Human Kinetics.
- Tobin, K. G., & Capie, W. (1980). *The test of logical thinking, development and application* [Paper presentation]. National Association for Research in Science Teaching, Boston, MA.
- Ubuz, B., & Aydın, U. (2018). Geometry knowledge test about triangles: Evidence on validity and reliability. *ZDM Mathematics Education*, 50(4), 659–673. <https://doi.org/10.1007/s11858-018-0964-y>
- Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, 43(7), 352–360. <https://doi.org/10.3102/0013189X14553660>
- Weber, K., Inglis, M., & Mejia-Ramos, J. P. (2014). How mathematicians obtain conviction: implications for mathematics instruction and research on epistemic cognition. *Educational Psychologist*, 49(1), 36–58. <https://doi.org/10.1080/00461520.2013.865527>
- Wei, W. (2014). Using summative and formative assessments to evaluate EFL teachers' teaching performance. *Assessment & Evaluation in Higher Education*, 40(4), 611–623. <https://doi.org/10.1080/02602938.2014.939609>
- Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: Students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, 37(6), 747–760. <https://doi.org/10.1080/02602938.2011.572153>
- Wholey, J. S. (1996). Formative and summative evaluation: Related issues in performance measurement. *Evaluation Practice*, 17(2), 145–149. [https://doi.org/10.1016/s0886-1633\(96\)90019-7](https://doi.org/10.1016/s0886-1633(96)90019-7)
- Widana, I. W., Parwata, I., Parmithi, N. N., Jayantika, I. G. A. T., Sukendra, I. K., & Sumandya, I. W. (2018). Higher order thinking skills assessment towards

- critical thinking on mathematics lesson. *International Journal of Social Sciences and Humanities*, 2(1), 24-32. <https://doi.org/10.29332/ijssh.v2n1.74>
- Yang, Y. T. C., & Chang, C. H. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. *Computers & Education*, 68, 334-344. <https://doi.org/10.1016/j.compedu.2013.05.023>
- Zaman, A., Jumani, N. B., Ali, A., & Hussain, M. A. (2011). Predictive validity of scores in mathematics for reasoning ability in mathematics for grade 9 students in Khyber Pakhtunkhwa based on curriculum of mathematics. *Procedia-Social and Behavioral Sciences*, 12, 588-594. <https://doi.org/10.1016/j.sbspro.2011.02.072>

**<http://www.ejmste.com>**