Full research paper

PUPILS' SCHOOL PERFORMANCE AND THEIR COGNITIVE ABILITIES TO SOLVE PROBLEMS

Eva Hejnová✉
Petr Eisenmann
Lucie Loukotová
Jiří Přibyl

Faculty of Science, University of Jan Evangelista Purkyně in Ústí nad Labem, Czech Republic

✉ eva.hejnova@ujep.cz

ABSTRACT

The paper describes the results of a study whose main aim was to find the interrelationship between pupils' school grades in Czech language (native), mathematics, and physics and pupils' cognitive predispositions to problem-solving in science and mathematics diagnosed by the Lawson Classroom Test of Scientific Reasoning and the Culture of Problem Solving test. A total of 180 pupils from the Czech Republic aged 14-15 took part in this study.

The results show that pupils with better grades in the monitored subjects achieve better results in both tests. It also turns out that there are generally statistically insignificant differences between the results of pupils assessed by grades 1 or 2 and between those assessed by grades 3 or 4. Pupils' performance in the two tests might help to strengthen the objectivization of grading at school. They might also help identify the indicators important for developing problem-solving skills. The research specifically points to the need to develop algebraic thinking, the conception of infinity, spatial imagination, geometric imagination in the plane, proportional reasoning, and the ability to control variables.

KEYWORDS

Culture of problem solving, lower secondary school, mathematics education, school performance, scientific reasoning

HOW TO CITE

Hejnová E., Eisenmann P., Loukotová L., Přibyl J. (2024) 'Pupils' School Performance and Their Cognitive Abilities to Solve Problems', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 54-65. http://dx.doi.org/10.7160/eriesj.2024.170105

Article history
Received
September 13, 2022
Received in revised form
June 25, 2023
Accepted
September 25, 2023
Available on-line
March 31, 2024

Highlights

- *Pupils' classroom assessment shows a connection to their performance in cognitive skills tests.*
- *Pupils with better grades (classification) achieve better results in the test of predispositions to solving mathematical problems and in Lawson's Test of Scientific Reasoning.*
- *There are almost no statistically significant differences in the performance of pupils assessed by grades 1 or 2 and in the performance of pupils assessed by grades 3 or 4.*

INTRODUCTION

The paper is one of the outcomes of the research project concerning mutual relations between two constructs that are related to problem-solving (the Culture of problem solving and Scientific reasoning) and school performance in the Czech language, mathematics, and physics. The construct *Culture of problem solving* (CPS) was introduced by some authors of this paper as a tool for describing a pupil's ability to solve mathematical problems (Eisenmann et al., 2015). The other construct is *Scientific reasoning* (SR), which includes the thinking and reasoning skills involved in systematically exploring a problem, formulating and testing hypotheses, evaluating experimental outcomes, etc. (Bao et al., 2009; Gormally et al., 2012). In (Cihlář et al., 2017), we present the results of a small research conducted among 23 pupils aged 14-15 in the Czech Republic in 2016. The results of this research proved the legitimacy of the idea of exploring mutual relations between individual components of CPS and SR dimensions. This research was then followed up with extensive research in 2017 to describe the dependency between all components of the CPS and the SR dimensions (Cihlář et al., 2020). The objective of the article is to provide a new perspective on the investigation issue based on the data obtained and to answer the question about the relationship between school performance (classification) and the constructs of CPS and SR.

The following sections describe two constructs (CPS and SR). This is discussed in detail in (Hejnová et al., 2018). The following text is a slightly shortened version essential for the purpose of this paper.

The Culture of Problem Solving

The CPS construct was developed to better describe the pupils' dispositions to solve mathematical problems. In creating the CPS, we tried to make this description independent of the problem itself and equally independent of the pupils' knowledge. The development of the construct is well described in (Eisenmann et al., 2015). At this point, we should mention that in its development, we drew mainly from the work of (Schoenfeld, 1982; Sriraman, 2005; Wu and Adams, 2006).

The CPS consists of four components: mathematical intelligence, reading comprehension, creativity, and the ability to use existing knowledge.

In the development of *mathematical intelligence*, we focused on the perception and understanding of six selected mathematical phenomena. We chose Gardner (1993) and Juter and Sriraman (2011) as our inspiration, but our primary goal was not to seek out mathematically gifted pupils. Our focus was to determine the level of disposition of all pupils in the areas that we consider important for successful mathematical problem-solving. We emphasize that we are not concerned with mathematical content per se but only with content to determine the level of the specific phenomenon. These phenomena are logical reasoning, the conception of infinity, spatial imagination, geometric imagination in the plane, algebraic thinking, and arithmetic patterns.

In accordance with the PISA framework, we understand *reading comprehension* as a functional literacy, that is, as a set of knowledge, skills, strategies, and attitudes needed to understand, use, and evaluate all kinds of texts in different contexts. Reading comprehension is one of the key competencies needed to successfully solve problems, especially in the case of word problems (Akbaşl et al., 2016; Pape, 2004; Vilenius-Tuohimaa et al., 2008).

As far as *creativity* is concerned, we focused only on the part of it that relates to divergent thinking (Guilford, 1967). We recognize that convergent thinking also plays a vital role in the task solving process, but it is the production of different ideas that we see as very important in this process, as confirmed by (Kwon et. al., 2006). By divergent thinking, we mean the production of diverse but appropriate answers to an open-ended question or problem. It can be said that the higher the level of creativity, the more difficult the problems can be presented to pupils (Chamberlin and Moon, 2005).

Although some parts of the construct CPS are either investigated from a domain perspective (mathematical intelligence) or are only found in specific school contexts (reading comprehension), creativity per se has been the focus of attention in other areas of human endeavour and is seen as a prerequisite for problem-solving across different domains (Zhou, 2012).

The *ability to use the existing knowledge* is the last component of the CPS. This ability is a necessary condition for the successful solution of non-routine problems. We have developed it to determine the level of formalism in the pupils. Eisenmann et al. (2015) have shown one connection with general intelligence: Those pupils who have a higher indicator of intelligence also show a higher indicator of the ability to use the existing knowledge.

Scientific Reasoning

SR can be regarded as a complex process that is widely defined as "the skills involved in inquiry, experimentation, evidence evaluation, and inference that are done in the service of conceptual change or scientific understanding" (Zimmerman, 2007, p. 172). These general skills, referred to as science process skills (Padilla, 1990), are considered crucial components of STEM education (Science, Technology, Engineering, and Math). As Coletta and Phillips (2005) and Han (2013) have shown, their development is closely related to the cognitive abilities of the pupils and their prior knowledge of the content. In our research, we focus on measuring scientific reasoning skills. That is why we deal with the operational definition of SR. Similarly, to Lawson (1978, 1982, 2004, 2005), we suppose that its structure is determined by the hypothetical-deductive nature of science and includes dimensions such as proportional thinking, identification, and control of variables, probabilistic thinking, correlational thinking and inductive and deductive reasoning (Han, 2013). However, other skills could be involved because of the multidimensional structure of SR (Opitz et al., 2017).

Classroom Assessment in Czech Schools

We understand classroom assessment to be the process of measuring performance and gaining evaluation information in lessons. This information is primarily intended for pupils. This assessment relates to the specific learning objectives and is an integral part of the teaching (different types of assessment of pupils' knowledge and skills, oral and written testing, peer assessment, training of pupils in self-assessment) or directly supports and complements it (on-the-fly diagnostic information for parents about pupils' school performance).

Assessment is regarded as the key tool for the improvement of school performance (see, e.g., Black et al., 2003, 2004; Naylor et al., 2005). Together with the process of enriching and updating the objectives of education, also the educational content that has to be assessed is modified. The quality and efficiency of the educational process are significantly influenced by the way they are evaluated. That is why teachers from developed countries ask for assessments that will be well thought out with respect to the effects they will have on school practice. Concerns that assessment of only some (well-measurable) educational objectives will result in deformation of teaching brought about the need to verify educational goals that had not been previously assessed (critical thinking, problem-solving, social skills, etc.). This puts new demands on assessment methods and used tools (Chvál et al., 2015).

The current conception of mathematics and science education that should be acquired by all students at school emphasizes, in particular, the development of a general understanding of important concepts, understanding of methods by which science gets evidence to support its claims, understanding of strengths of science and its limitations in the real world; the ability to draw correct and well-founded conclusions from presented facts and information, to critically assess people's statements on the basis of the evidence presented and to distinguish opinions from evidence-based claims.

Tests used by teachers in the Czech Republic often target lower-level skills and knowledge of facts. However, international research and surveys (e.g., PISA[1]) show a clear tendency to set to students the so-called context tasks that refer to some situation and show the applicability of mathematics and science in real-life situations or tasks that test such knowledge and skills of students that they will need in their future lives. Tasks also often focus on overall student abilities that permeate all subjects (not only mathematics and science), on the use of their own way of thinking and understanding in specific life situations.

In our research, we focus primarily on summative assessment, the purpose of which is to obtain an overall overview of students' performance. The aim of this assessment is to diagnose a student and evaluate their performance with respect to the assessed group of students.

Classification (assessment by grades) is still predominantly used in Czech schools as the standard for summative assessment of students. Its advantages are simplicity and systematicity when used in practice, as well as its long tradition and comprehensibility to parents and the public. In Czech schools, students are evaluated by grades 1 to 5. Grade 1 corresponds to the best performance, and 5 describes insufficient performance and failure.

Research on classification shows a relative stability of achievement during school attendance, which is attributed to a relatively stable and evenly developing dispositional basis, which determines a student's school performance (Hrabal, 1989). The existence of a grade has a constant influence on the character, intensity, and focus of a student's learning activities, affects the classroom atmosphere, and also represents simple informative feedback for the student and their parents (Slavík, 1999).

One of the tools for assessing students in science and mathematics is problem-solving. Suurtamm et al. (2016) highlight the relationship between students' problem-solving success and their grades. Tasks on whose basis students should be assessed, according to (Swan and Burkhardt, 2012), should present a balanced view of the curriculum in terms of all aspects of performance that the curriculum wants to encourage. Each student has specific cognitive predispositions to problem-solving. Students with better cognitive predispositions can be expected to have better grades in mathematics and physics. As expected, it was confirmed (Česká školní inspekce (ČŠI), 2019), for example, that students who had a better grade in mathematics on the school final report from the previous school year achieved higher average success in the mathematical literacy test, with a slightly more pronounced effect in case of ninth-graders than in case of 2nd-year students of upper secondary schools.

Objectives

In this paper, we focus on the relationship between classroom assessment (classification) and cognitive predispositions of lower secondary 14 to 15-year-old students to solve problems in the area of mathematics and science. Their predispositions were diagnosed using the Lawson Classroom Test of Scientific Reasoning (LT) and the CPS test. With respect to the focus of these two tests, we were working with grades in three major subjects - Czech language, mathematics, and physics as we believe these play a significant role in the development of the ability to solve problems.

We asked the following research question: To what extent are the variables determined by the CPS and LT tests related to grades in Czech, mathematics, and physics?

MATERIALS AND METHODS

Culture of Problem Solving

The test to determine the values of all four CPS components was designed to be administered in one teaching lesson and took 45 minutes. The individual components were timed as follows: mathematical intelligence - 13 minutes, reading comprehension - 13 minutes, creativity - 9 minutes, and the ability to use the existing knowledge - 9 minutes. Each student worked independently during the assessment and could only use a simple calculator in addition to writing tools. All parts of the test were evaluated by the authors of this paper.

The test of *mathematical intelligence* consisted of 8 problems determining sensitivity to the above-mentioned six specific phenomena: logical reasoning, a conception of infinity, spatial imagination, geometric imagination in a plane, algebraic thinking, and arithmetic patterns. All test problems, with one exception, were closed multiple-choice tasks with one correct answer. The tested student could get 2 points for each problem. The sum of all points formed the total score.

The *reading comprehension* test was created on the same principle as the tests in the PISA research. The test proceeded as follows: the students were presented with a text of 15 lines about kangaroos. The students then had to answer 4 closed and 2 open questions. The text was available to them all the time.

Since we chose to measure *creativity* by divergent thinking, we chose Guilford's Alternative Uses Test as our instrument. This test is based on the fact that the student is presented with a word that expresses a certain object, and the student is asked to suggest as many different and unusual uses of this object as possible, counting only feasible uses. In the test, four words were presented to the students to rule out a certain inclination towards a subject and to observe phenomena such as fluency, originality, and elaborateness. The qualitative assessment of each part of the test was then converted into scores, and the total score indicated the creativity index.

The test of *the ability to use the existing knowledge* consisted of four simple, non-routine problems. At the beginning of each problem, the students revised the knowledge that was necessary to solve the task (How to find the part from the whole with percentage, the circumference of a circle, the surface of a cuboid, and the lowest common multiple). For each task, the student could get one point, and the total sum informs the overall score in this test.

Scientific Reasoning

SR was tested by the Czech version of LT including 24 items (Dvořáková, 2011) with small corrections in items 8a and 8b according to (Han, 2013), allowing one to examine six

1 Programme for International Student Assessment

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

dimensions: conservation of matter and volume (items 1 to 4), proportional reasoning (items 5 to 8), control of variables (items 9 to 14), probability reasoning (items 15 to 18), correlation reasoning (items 19, 20), and hypothetical-deductive reasoning (items 21 to 24) (Lawson, 1978).

LT is a two-tier multiple-choice test with items of increasing difficulty, each of the two-tier items including a question offering answers and possible reasons for the response to the question. A student could get two points for questions 1 through 22 if he chose the correct answer and concurrently its correct justification. Only the answers to questions 23 and 24 were evaluated separately; that is, the student received one point for each correctly answered question or its correct justification. The students had a maximum of 45 minutes to solve the test.

The results of the test allow us to determine what level of scientific reasoning a particular student has achieved (Dvořáková, 2011). The first developmental level is concrete-level reasoning. This stage consists of students who get 0-8 points in the LT test. The second level is transitional and is made up of students who get 9-16 points. The highest level is formal-level reasoning. This stage is achieved by students who get 17-24 points.

Sample

A total of 180 students (76 girls and 104 boys) aged 14-15 from eight classes from six lower secondary schools and one class of an eight-year secondary grammar school took part in our study in May and June 2017. All schools were located in three towns in the Ústí nad Labem Region in the Czech Republic. None of the classes were specialized and integrated for pupils with physical or mental disabilities or low socioeconomic status. All the pupils were native Czech speakers.

Statistical Evaluation

As part of the statistical processing, the pupils' grades and the results of the CPS and LT tests were first analyzed with respect to the pupils' scores.

Also, the relationships between grades in the Czech language (CZECH), mathematics (MATH), and physics (PHYS) and the overall scores in the CPS and LT tests were examined. The Kruskal-Wallis test was used to evaluate the dependency between grades and overall scores in these tests, and partial differences among different classes of grades were studied using Wilcoxon tests. The effect size coefficient ξ^2 was used to determine the strength of the association. The guidelines of Field (2013) were followed to interpret the size of the effect. For the purpose of evaluation, the category of pupils with grade 5 was merged with the category of pupils with grade 4. This modification was necessary because of the low number of pupils with grade 5 in the sample. In the case of the CPS test, the results in individual subtests were normed in such a way that all subtests had the same weight.

The calculation was realized by STATISTICA 12.0 (StatSoft, Inc.). The level of significance $\alpha = 0.05$ was used in all tests.

RESULTS

The pupils' school performance was assessed according to their grades in Czech language, mathematics, and physics on the final school report in the school year 2016/2017. The structure of grades in the monitored subjects is given in Figure 1. The structure of grades in Czech language and mathematics is very similar; the proportion of grades 1 to 4 is roughly the same in both subjects. About 60% of pupils have a grade 1 or a grade 2. In contrast, a greater proportion of grades 1 can be observed in physics (about 30% of pupils in contrast to 20% in Czech language and mathematics) and a lower proportion of grades 4. The proportion of grades 3 is about the same in all three subjects (slightly more than 20%). Grades 5 only appeared in mathematics. Average grades in individual subjects correspond to the above. The average grade in Czech language and mathematics was 2.30, and in physics, 2.06.

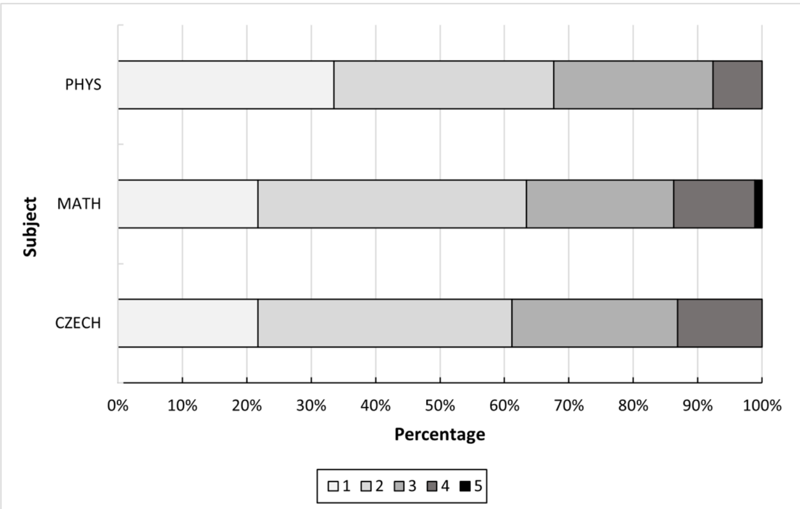

Figure 1: Structure of grades, 2017 (source: own calculation)

As far as the LT test is concerned, the average score was 7.9 points (out of 24 points) with a standard deviation of 5.3 points. Figure 2 illustrates that the distribution of the achieved score does not show a maximum in the proximity of the average gained score. On the contrary, the pupils' scores are scattered on the whole scale of possible scores.

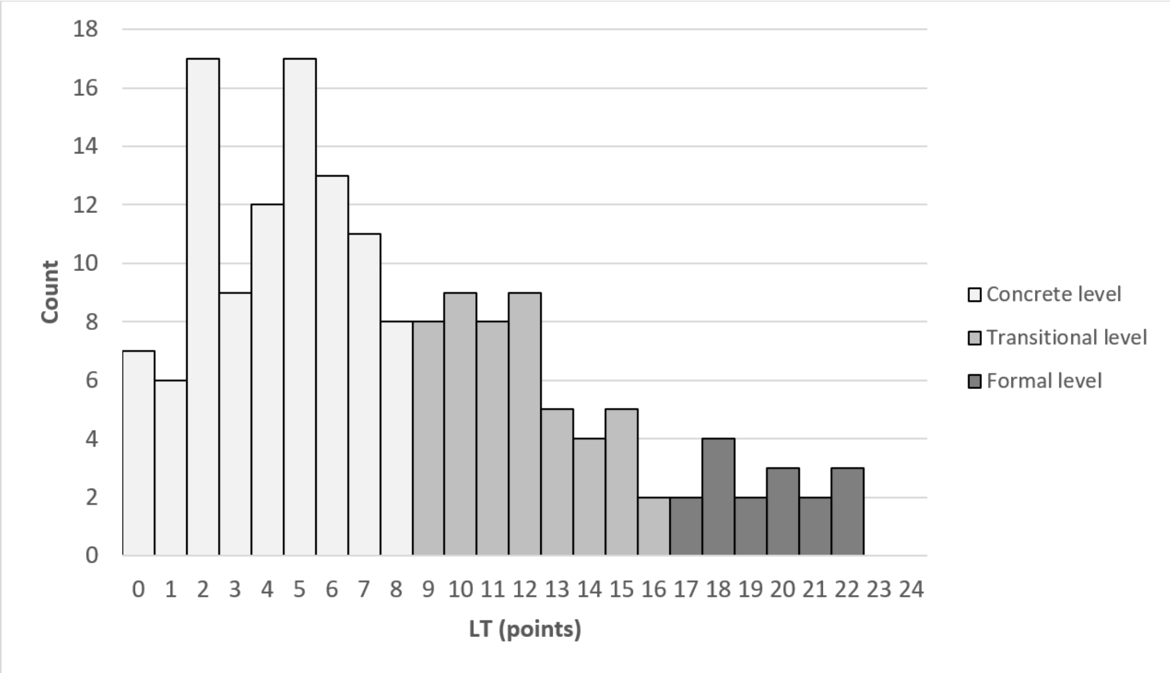

Figure 2: Points from Lawson test (LT), 2017, (source: own calculation)

Pupils most often had two to seven points in the LT test, namely 48% of them. 45% of the pupils had an above-average score, that is, 8 points and more, and 19% of the pupils gained more than half of the possible points (i.e., more than 12). Roughly 8% of the pupils had only one or no point. None of the pupils achieved the maximum number of points. 60% of the pupils are on the lowest, which is concrete-level reasoning. 30% of the pupils are on the transitional level of scientific reasoning. Only 10% of the pupils have reached the highest, i.e., formal-level reasoning.

The maximum number of points was 24, also in the CPS test. The average score achieved by pupils was 12.1, with a standard deviation of 4.4 points. The score on the CPS test is distributed normally (see Figure 3).

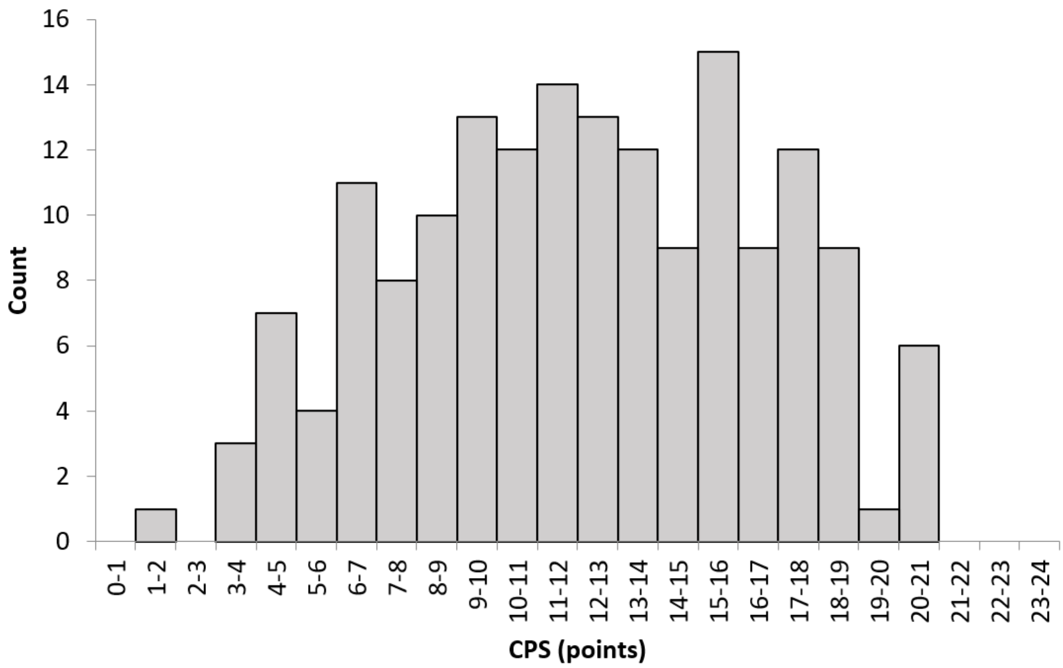

Figure 3: Points from CPS test, 2017, (source: own calculation)

The coefficient of variation in LT is 67.9%, which means that the scores of individual pupils have a moderate range. In contrast, the coefficient of variation in the CPS test is only 36.6%, which means that the scores of the CPS test are much more consistent. Kruskal-Wallis tests show a significant association between the overall score on the CPS test and grades in Czech language, mathematics, and physics; the effect of the grade on CPS and LT tests is large. The results of the tests are presented in Table 1.

Pairs of variables	Kruskal-Wallis test		Effect size	
	H	*p*	ε^2	size of effect
CPS & CZECH	68.87	< .001	.409	large
CPS & MATH	44.99	< .001	.261	large
CPS & PHYS	50.42	< .001	.304	large

Table 1: Relationship between the overall score in the CPS tests and grades in Czech, mathematics and physics, 2017 (source: own calculation)

The results of Kruskal-Wallis tests in the case of the search for relations between the score in LT and grades in Czech, mathematics, and physics are very similar. The results of the tests are presented in Table 2.

Pairs of variables	Kruskal-Wallis test		Effect size	
	H	*p*	ε^2	size of effect
LT & CZECH	68.87	< .001	.208	large
LT & MATH	44.99	< .001	.189	large
LT & PHYS	50.42	< .001	.171	large

Table 2: Relationship between overall score in the LT test and grades in Czech, mathematics and physics, 2017 (source: own calculation)

Pupils with better grades generally achieve better scores both in the CPS and LT tests. The grade has a larger effect on the CPS score than on the LT score.

A more detailed look at the relationship between the grades in the selected subjects and the result in the tests is very interesting. Wilcoxon subtests show that in the case of grades in Czech and physics, the scores in CPS and LT tests are not significantly different for pupils with grades 1 or 2 (Czech: p = .10 for CPS, p = .06 for LT, physics: p = 1.00 for both tests), and for pupils with grades 3 or 4 (p = 1.00 for both of subjects and tests). This allows us to define, for the case of grades in Czech and physics, a group of "more successful pupils" (with grades 1 or 2) and of "less successful pupils" (with grades 3 or 4).

Basic descriptive statistics for the results of CPS and LT tests in relation to grades in the Czech language are shown in Figure 4 and Figure 5.

Similarly, Figures 6 and 7 present the basic descriptive statistics for the results of CPS and LT tests in relation to the grades in physics.

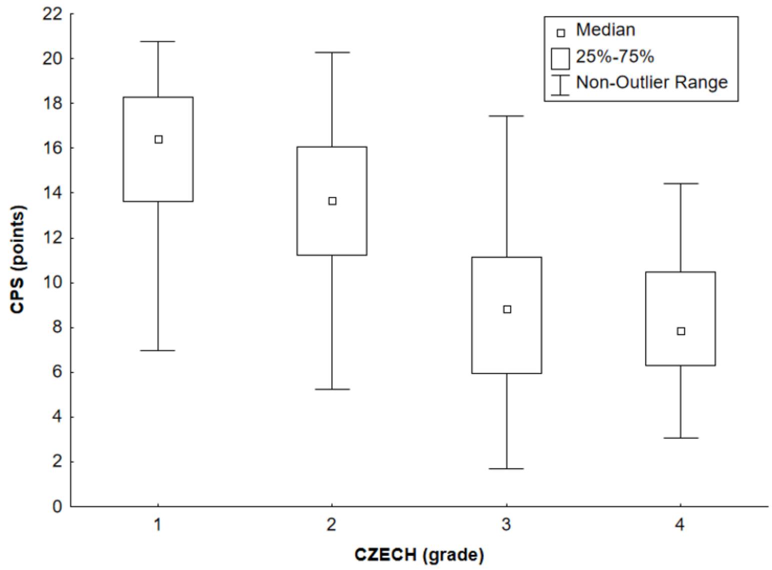

Figure 4: Results of CPS test with respect to grades in Czech, 2017 (source: own calculation)

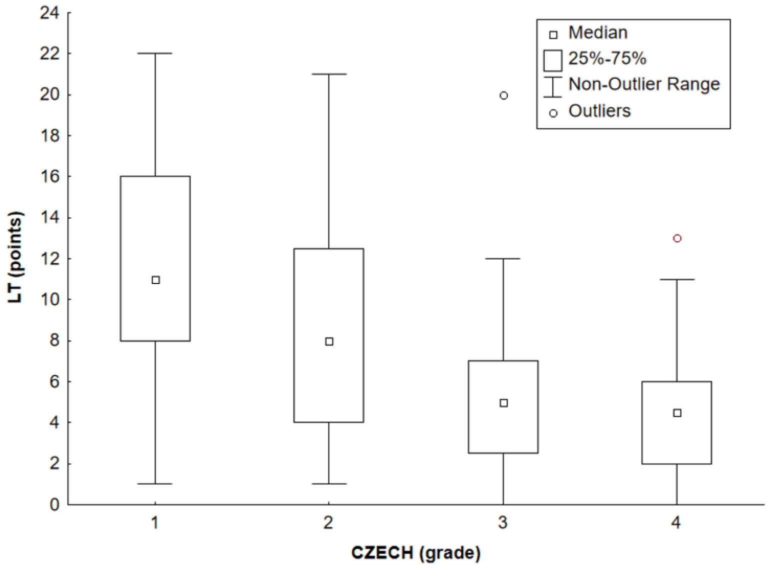

Figure 5: Results of LT with respect to grades in Czech, 2017 (source: own calculation)

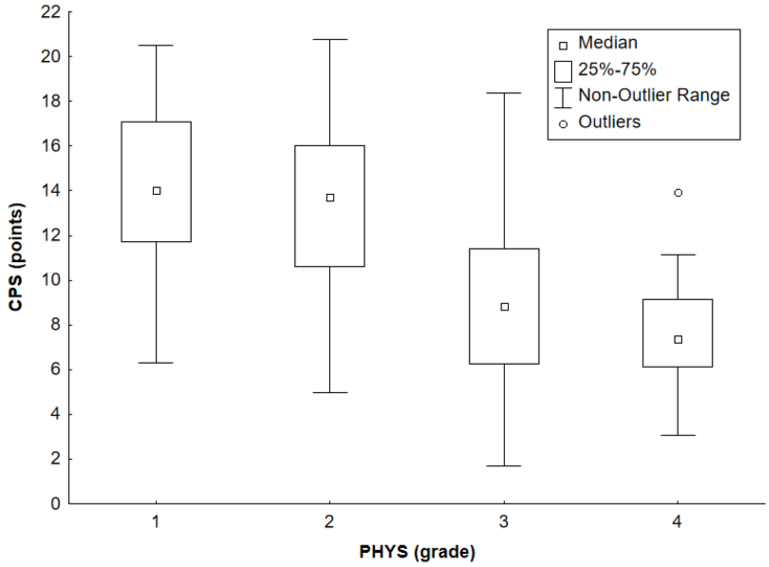

Figure 6: Results of CPS test with respect to grades in physics, 2017 (source: own calculation)

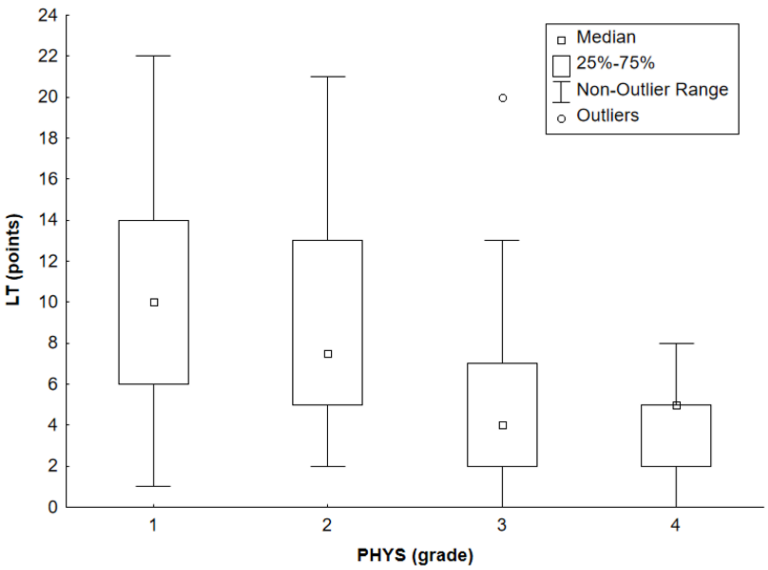

Figure 7: Results of LT with respect to grades in physics, 2017 (source: own calculation)

In the case of grades in mathematics, there are no statistically significant differences between pupils with grades 3 or 4 in CPS and LT tests ($p = 1.00$ for both tests; see Figures 8 and 9 for basic descriptive statistics). However, unlike in the case of Czech and physics, pupils with grades 1 in mathematics are significantly better than all other groups both in CPS and LT tests ($p < .02$ in all cases). Moreover, in the case of LT, pupils with grades 2, 3, or 4 show similar results (the result of pupils with grades 2 or 3 is borderline).

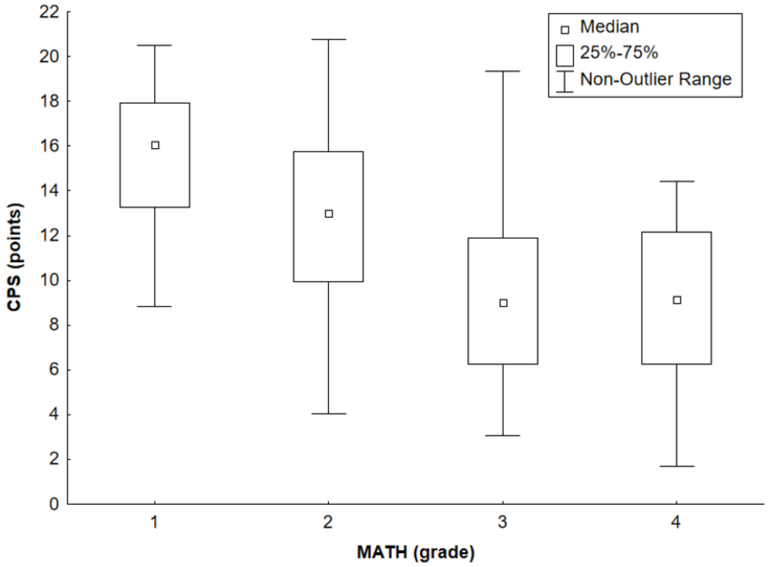

Figure 8: Results of CPS test with respect to grades in mathematics, 2017 (source: own calculation)

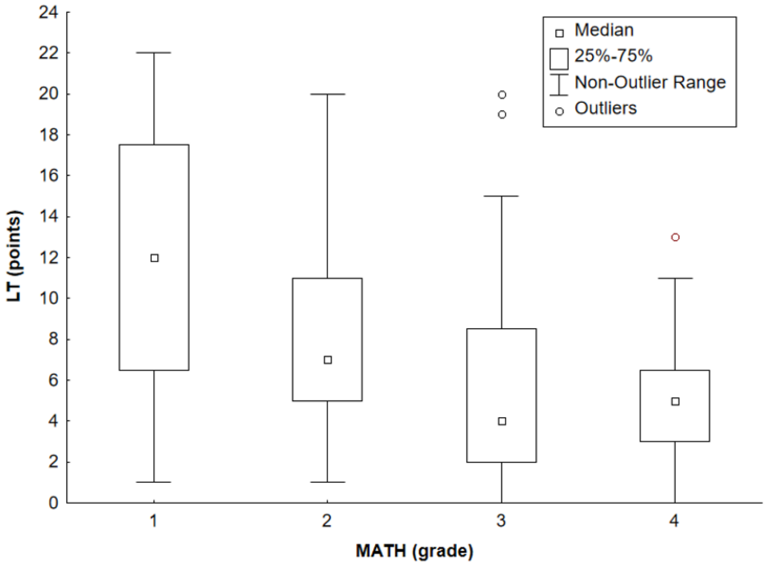

Figure 9: Results of LT tests with respect to grades of mathematics, 2017 (source: own calculation)

DISCUSSION

The study was designed to generate data relevant to the question of the relationship between the variables determined by the CPS and the LT tests and the grades in Czech, mathematics, and physics among a cohort of Czech pupils aged 14-15. All statistical tests showed a significant association between the overall score of the CPS and LT tests and the grades where the association of the score of the CPS test with the grades was found to be stronger than in the case of LT.

The dependency between variables measured by the CPS and LT tests and school performance given by pupils' grades is not much described in the literature. Valanides (1997) examined the relationship between performance on the Test of Logical Thinking (TOLT) and gender, a section of the study, and measures of school achievement of 12th-grade students. He concluded that gender, section of study, achievement in mathematics, and grade point average, but not achievement in science and Greek language, contributed significantly to predicting performance on TOLT. Generally, a pupil with a high IQ is expected to have good results at school (see, e.g., Jalili et al., 2018), and a learner with higher levels of SR could be expected to be a superior problem solver (Tajudin and Chinnappan, 2015). Research also has shown (e.g., Hand et al., 2001; Samková et al., 2021) that reasoning skills represent a set of broadly transferable skills across science and mathematics, and teaching scientific reasoning has a lasting impact on general learning ability not only in science but also in mathematics. Hilbert et al. (2019) exposed in their research with Austrian students that the grade in mathematics was best predicted by reasoning at the age of 11-12. The reasoning was assessed using the standard progressive matrices (Raven et al., 2000). The test requires participants to complete matrices based on visual patterns. Such tasks were used by us in the test of mathematical intelligence (one component of the CPS structure) as well.

As stated in the chapter Results, there are almost no statistically significant differences in the performance in the CPS and LT test of pupils assessed by grades 1 or 2 in Czech and physics and in the performance of pupils assessed by grades 3 or 4 in these subjects. We can talk about a group of "more successful pupils" (with grades 1 and 2) and of "less successful pupils" (with grades 3 and 4). A similar grouping of pupils can be found in Tajudin and Chinnappan's (2015) study, where groups are referred to as high-achievement and low-achievement groups. Also, in PISA research (Potužníková et al., 2019), all pupils are divided into two basic groups: pupils with critical reading skills (they do not reach basic level 2 according to PISA) and others.

Teachers who diagnose their pupils using the CPS and LT tests can use these results to divide pupils into two groups and thus better account for pupil heterogeneity in instruction. Dividing pupils and then treating the two groups individually during a lesson can be a powerful tool that allows the teacher to better take into account the different cognitive backgrounds of the pupils in the class and to work with each group in a slightly different way. Pupils who are likely to have difficulties in understanding the material can be given more attention and given, for example, appropriate preparatory tasks or exercises to practice the material. Pupils who have been more successful on the CPS and LT tests can work more independently in the lessons or receive more challenging tasks. For both groups, this approach can be effective in increasing their motivation to learn.

However, this division into two groups may not always be 100% effective and may have limitations, for example, due to differences in pupils' individual abilities or their social interactions in groups. It is also important to note that the CPS and LT tests alone may not fully reveal all aspects of the abilities and learning needs of the pupils. In addition, some pupils may not perceive the division into two groups positively and may feel, for example, a certain sense of injustice or stigma. If a teacher decides to create non-heterogeneous groups for the above reasons, for example, to implement group learning, this can effectively strengthen pupils' social bonds and social skills. Non-heterogeneous groups can also help create an environment in which weaker pupils feel accepted and included regardless of their individual abilities or skills, which can increase their self-esteem and motivation to learn. However, each teacher must always individually weigh the advantages and disadvantages of both approaches before deciding on a particular teaching method.

If teachers want to eliminate a pupil's reading comprehension deficit, they can be inspired by the so-called Singapore mathematics based on the triad: concrete manipulation - image representation - abstract model (Kaur and Yeap, 2009; Wong and Lee, 2009). The image representation of the initial situation is a key tool that Singaporean students consciously learn to deal with. Drawing an image converts verbal information into a visual one. This process helps the student to realize the relationships between the individual pieces of information in the task assignment. The structure of the task parameters is analyzed and converted into a visual representation. This principle is also in accord with our research (Eisenmann et al., 2015) or, for example, with (Nunes and Bryant, 2015). Similarly, if we want to compensate for a pupil's deficiency in mathematical intelligence, we can successfully present him with mathematical tasks of different types. We must now note that each of the six named phenomena is differently sensitive to stimuli. For example, the perception of infinity develops very slowly and matures only with the individual. In contrast, arithmetic patterns and geometric imagination in a plane are phenomena that develop quite well (Alsina and Nelsen, 2006; Rezaie and Gooya, 2011). In the case of logical reasoning, it is sometimes difficult to find out what is causing the problems. In fact, the cause may be a low level of reading literacy. In a situation where we have ruled out this cause, we present the problems to the students and go through the problem statement with them to see if they are able to identify the relationships between objects. For these tasks, it is recommended that they also plot these relationships. Pólya and Conway (2004) and Boaler and Dweck (2016) clearly recommend this approach.

In the present study, we found out that a large part of the pupils (60%) had acquired only concrete-level reasoning, a third of the pupils were in the transitional (30%), and a smaller part of the pupils were in formal (10%) reasoning levels. However, in the 9th grade, most pupils should optimally be on the transitional level of reasoning (Han, 2013). Concrete-level reasoning refers to thinking patterns that allow pupils to grasp concepts and statements that directly refer to well-known actions. At this level, pupils can follow instructions step by step, provided each step is fully specified. On the transitional level of reasoning, pupils also remain limited to being only capable of partial formal reasoning. Given the positive association between the grades of the three subjects and the scores on the LT test, the "more successful pupils" are expected to achieve good results on the LT test. However, the scores on the LT test were satisfactory for only 40% of the students who achieved transitional or formal reasoning levels, which does not correlate with the number of pupils with grades 1 or 2, who represented approximately 60% of all study participants. A possible explanation for this is that the mastery of the individual dimensions of the SR is not reflected in the grades to a sufficient extent.

One of the strategic lines of the Strategy for the Education Policy of the Czech Republic up to 2030+ is to move away from a broad body of expected knowledge and to foster the ability to understand issues in a deeper context (Fryč et al., 2020). Assessment practice should, therefore, mirror the curriculum we want to develop: its goals, objectives, content, and instructional approaches. The Framework Education Programme for Basic Education (Ministerstvo školství, mládeže a tělovýchovy, 2017), which is a basic curricular document in the Czech Republic, emphasizes the development of key competencies, which also includes problem-solving competencies. Evaluation of such general competencies requires a much broader and more holistic view of student performance, e.g., also in mathematics and science. In contrast to knowledge assessment, however, assessment of progress toward competencies is more difficult, and teachers need to receive useful support in this regard.

At the end of this chapter, we would like to mention that in our study, we tested only pupils from the Ústí nad Labem Region due to the practical feasibility of the research. However, we believe that this fact does not have a significant impact on the generality of our conclusions, as schools were selected in such a way as to minimize possible bias caused by regional limitations, and the sample of pupils was sufficiently representative of the entire population in the age category 14-15 years. Thus, although our research was limited to one region, we believe that our findings may be relevant to the education of pupils in other regions of the Czech Republic and bring new insights into the overall context of educational research.

If we could generalize our research question, we could also discuss whether the three subjects we have chosen (Czech language, mathematics, and physics) really play a significant role in the development of students' ability to solve problems.

We can say that a certain unifying element of both CPS and SR is critical thinking. Indeed, this mode of reasoning is largely present in the background of both constructs (Dowd et al., 2018; Syafril et al., 2020) but the given triad of CPS, SR and critical thinking has not yet been examined together. Research that addresses the constructs of CPS, SR, and critical thinking can be perceived as a possible challenge for the future.

CONCLUSION

Our study proved that pupils with better classroom assessment achieve better results both in CPS and LT tests. This conclusion refers to all three monitored subjects. At the same time, it has been shown that there were no statistically significant differences between the results of pupils with grades 1 or 2, and also between the results of pupils with grades 3 or 4. This finding seems to point to the fact that when assessed by grades, pupils are evaluated not only for their cognitive skills but also for other competences and for implementation of other learning objectives that are not targeted by the CPS and LT tests. However, we believe that the results of pupils in tests such as CPS and LT can help to reinforce the objectification of grades within the summative assessment of pupils. Evaluation by grades still has a significant impact on the degree of effort in learning. Thus, it makes sense to develop those skills in pupils who are involved in problem-solving.

These skills correspond to the variables that were subject to testing using the CPS test. In particular, we mean reading comprehension and some variable components of mathematical intelligence that can be developed: algebraic thinking, the conception of infinity, spatial imagination, and geometric imagination in the plane.

In the case of the LT test, it turned out that nearly two-thirds of the pupils only reached concrete-level reasoning. Pupils should be able to solve problems on the conservation of matter and volume already at the end of the primary level. This should be followed by practicing problems on proportional reasoning at both primary and lower secondary school levels. With respect to the development of scientific reasoning, we believe that the development of the ability to control variables is essential at the lower secondary school level. A thorough acquisition of the above-mentioned skills is a prerequisite to the development of other skills, such as probability reasoning, correlation, and hypothetical-deductive reasoning.

Pupils who perform better on the CPS and Lawson tests are more likely to develop problem-solving competences: they are better equipped to perceive a variety of problem situations both in and out of school, to recognize and understand the problem, and to think through and plan how to solve it. Additionally, they may be more effective in searching for information that is suitable for solving a problem and identifying its commonalities and differences. These pupils are then also more likely to succeed in entrance examinations to schools where science and engineering are taught.

REFERENCES

Akbaşl, S., Şahin, M. and Yaykiran, Z. (2016) 'The Effect of Reading Comprehension on the Performance in Science and Mathematics', *Journal of Education and Practice*, Vol. 7, No. 16, pp. 108-121.

Alsina, C. and Nelsen, R. B. (2006) *Math made visual: creating images for understanding mathematics*, Washington DC: Mathematical Association of America.

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y, Wang, Y., Li, L. and Wu, N. (2009) 'Learning and Scientific Reasoning', *Science*, Vol. 323, No. 5914, pp. 586-587. https://doi.org/10.1126/science.1167740

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003) *Assessment for learning: putting it into practice*, Maidenhead: Open Univ. Press.

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2004) 'Working inside the Black Box: Assessment for Learning in the Classroom', *Phi Delta Kappan*, Vol. 86, No. 1, pp. 8-21. https://doi.org/10.1177/003172170408600105

Boaler, J. and Dweck, C. (2016) *Mathematical mindsets: unleashing students' potential through creative math, inspiring messages, and innovative teaching*, San Francisco: Wiley

Česká školní inspekce (ČŠI) (2019) *Rozvoj matematické gramotnosti na základních a středních školách ve školním roce 2017/2018 [Development of mathematical literacy in primary and secondary schools in the school year 2017/2018]*, Praha: Česká školní inspekce.

Chamberlin, S. A. and Moon, S. M. (2005) 'Model-Eliciting Activities as a Tool to Develop and Identify Creatively Gifted Mathematicians', *Journal of Secondary Gifted Education*, Vol. 17, No. 1, pp. 37-47. https://doi.org/10.4219/jsge-2005-393

Chvál, M., Procházková, I. and Straková, J. (2015) *Hodnocení výsledků vzdělávání didaktickými testy [Evaluation of learning outcomes by didactic tests]*, Praha: Česká školní inspekce.

Cihlář, J., Eisenmann, P., Hejnová, E., and Přibyl, J. (2017). 'Science reasoning and culture of problem solving', *Proceedings of the 14th International Conference (ERIE 2017)*, Prague, pp. 33-40.

Cihlář, J., Eisenmann, P., Hejnová, E., and Přibyl, J. (2020). 'Problem Solving in Mathematics and Scientific Reasoning', *The New Educational Review*, Vol. 41, No. 3, pp. 97-108. https://doi.org/10.15804/tner.20.61.3.08

Hejnová, E., Eisenmann, P., Cihlář, J. and Přibyl, J. (2018) 'Relations between Scientific Reasoning, Culture of Problem Solving and Pupil's School Performance', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 11, No. 2, pp. 38-44. https://doi.org/10.7160/eriesj.2018.110203

Coletta, V. P. and Phillips, J. A. (2005) 'Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability', *American Journal of Physics*, Vol. 73, No. 12, pp. 1172-1182. https://doi.org/10.1119/1.2117109

Dowd, J. E., Thompson, R. J., Schiff, L. A. and Reynolds, J. A. (2018) 'Understanding the Complex Relationship between Critical Thinking and Science Reasoning among Undergraduate Thesis Writers', *CBE Life Sciences Education*, Vol. 17, No. 1, pp. ar4. https://doi.org/10.1187/cbe.17-03-0052

Dvořáková, I. (2011) 'Scientific reasoning of pupils - How to develop and test it' [Online], Available: https://kdf.mff.cuni.cz/lide/dvorakova/Plzen_prispevek_Dvorakova.pdf [17 Jun 2023]

Eisenmann, P., Novotná, J., Přibyl, J. and Břehovský, J. (2015) 'The development of a culture of problem solving with secondary students through heuristic strategies', *Mathematics Education Research Journal*, Vol. 27, No. 4, pp. 535-562. http://doi.org/10.1007/s13394-015-0150-2

Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics*, Sage: London.

Fryč, J., Matušková, Z., Katzová, P., Kovář, K., Beran, J., Valachová, I., Seifert, L., Běťáková, M. and Hrdlička, F. (2020) '*Strategy for the Education Policy of the Czech Republic up to 2030*', Praha: Ministerstvo školství, mládeže a tělovýchovy. [Online], Available: https://www.msmt.cz/uploads/Brozura_S2030_online_CZ.pdf [20 Jun 2023]

Gardner, H. (1993) *Frames of mind: the theory of multiple intelligences*, New York: BasicBooks.

Gormally, C., Brickman, P. and Lutz, M. (2012) 'Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates' Evaluation of Scientific Information and Arguments', *CBE—Life Sciences Education*, Vol. 11, No. 4, pp. 364-377. https://doi.org/10.1187/cbe.12-03-0026

Guilford, J. P. (1967) *The nature of human intelligence*, New York: McGraw-Hill.

Han, J. (2013) *Scientific Reasoning: Research, Development, and Assessment,* [Ph.D. thesis], Ohie: The Ohio State University.

Hand, B. M., Prain, V. and Yore, L. (2001) 'Sequential Writing Tasks' Influence on Science Learning', in Tynjälä, P., Mason, L. and Lonka, K. (ed.) *Writing as a Learning Tool*, Dordrecht: Springer. https://doi.org/10.1007/978-94-010-0740-5_7

Hilbert, S., Bruckmaier, G., Binder, K., Krauss, S. and Bühner, M. (2019) 'Prediction of elementary mathematics grades by cognitive abilities', *European Journal of Psychology of Education*, Vol. 34, No. 3, pp. 665-683. https://doi.org/10.1007/s10212-018-0394-9

Hrabal, V. (1989) *Pedagogicko-psychologická diagnostika žáka [Pedagogical and psychological diagnosis of the pupil]*, Praha: Státní pedagogické nakladatelství.

Jalili, A., Hejazi, M., Fomani, G. E. and Morovvati, Z. (2018) 'The Correlation between IQ with Educational Performance and Problem Solving Mediation', *Journal of Health Promotion Managment*, Vol. 7, No. 1, pp. 1-8. https://doi.org/10.21859/jhpm-08011

Juter, K. and Sriraman, B. (2011) 'Does High Achieving in Mathematics = Gifted and/or Creative in Mathematics?', in Sriraman, B. and Lee, K. H. (ed.) *The Elements of Creativity and Giftedness in Mathematics*, Rotterdam: SensePublishers. https://doi.org/10.1007/978-94-6091-439-3_4

Kaur, B. and Yeap, B. H. (2009) *Pathways to reasoning and communication in the primary school mathematics classroom: A resource for teachers by teachers*, Singapore: Centre for Research in Pedagogy and Practice, National Institute of Education.

Kwon, O. N., Park, J. H. and Park, J. S. (2006) 'Cultivating divergent thinking in mathematics through an open-ended approach', *Asia Pacific Education Review*, Vol. 7, No. 1, pp. 51-61. https://doi.org/10.1007/BF03036784

Lawson, A. E. (1978) 'The development and validation of a classroom test of formal reasoning', *Journal of Research in Science Teaching*, Vol. 15, No. 1, pp. 11-24. https://doi.org/10.1002/tea.3660150103

Lawson, A. E. (1982) 'The nature of advanced reasoning and science instruction', *Journal of Research in Science Teaching*, Vol. 19, No. 9, pp. 743-760. https://doi.org/10.1002/tea.3660190904

Lawson, A. E. (2004) 'The Nature and Development of Scientific Reasoning: A Synthetic View', *International Journal of Science and Mathematics Education*, Vol. 2, No. 3, pp. 307-338. https://doi.org/10.1007/s10763-004-3224-2

Lawson, A. E. (2005) 'What is the role of induction and deduction in reasoning and scientific inquiry?', *Journal of Research in Science Teaching*, Vol. 42, No. 6, pp. 716-740. https://doi.org/10.1002/tea.20067

Ministerstvo školství, mládeže a tělovýchovy (2017) 'Rámcový vzdělávací program pro základní vzdělávání [Framework Education Programme for Basic Education]', [Online], Available: https://www.msmt.cz/file/43792_1_1/ [10 June 2023]

Naylor, S., Keogh, B. and Goldsworthy, A. (2005) *Active Assessment for Science: Thinking, Learning and Assessment in Science*, London: Routledge.

Nunes, T. and Bryant, P. (2015) *Handbook of Child Psychology and Developmental Science: Volume II - Cognitive Processes*, Hoboken: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy217

Opitz, A., Heene, M. and Fischer, F. (2017) 'Measuring scientific reasoning - a review of test instruments', *Educational Research and Evaluation*, Vol. 23, No. 3-4, pp. 78-101. https://doi.org/10.1080/13803611.2017.1338586

Padilla, M. J. (1990) *The Science Process Skills*, [Online], Available: https://narst.org/research-matters/science-process-skills [20 June 2022]

Pape, S. J. (2004) 'Middle School Children's Problem-Solving Behavior: A Cognitive Analysis from a Reading Comprehension Perspective', *Journal for Research in Mathematics Education*, Vol. 35, No. 3, pp. 187. https://doi.org/10.2307/30034912

Pólya, G. and Conway, J. H. (2004) *How to solve it: a new aspect of mathematical method,* Princeton: Princeton University Press.

Potužníková, E., Janotová, Z. and Blažek, R. (2019) *Mezinárodní šetření PISA 2018: Koncepční rámec hodnocení čtenářské gramotnosti [PISA 2018: conceptual framework for reading literacy]*, Praha: Česká školní inspekce.

Raven, J., Raven, J. C. and Court, J. H. (2000) *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 3: The standard progressive matrices (including the parallel and plus versions)*, Oxford: Oxford Psychologists Press.

Rezaie, M. and Gooya, Z. (2011) 'What do I mean by combinatorial thinking?', *Procedia - Social and Behavioral Sciences*, Vol. 11, pp. 122-126. https://doi.org/10.1016/j.sbspro.2011.01.046

Samková, L., Rokos, L. and Vízek, L. (2021) 'A Joint Assessment of Reasoning about General Statements in Mathematics and Biology', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 14, No. 4, pp. 270-287. https://doi.org/10.7160/eriesj.2021.140406

Schoenfeld, A. H. (1982) 'Measures of Problem-Solving Performance and of Problem-Solving Instruction', *Journal for Research in Mathematics Education*, Vol. 13, No. 1, pp. 31. https://doi.org/10.2307/748435

Slavík, J. (1999) *Hodnocení v současné škole: Východiska a nové metody pro praxi [Evaluation in the present school: Starting points and new methods for practice]*, Praha: Portál.

Sriraman, B. (2005) 'Are Giftedness and Creativity Synonyms in Mathematics?', *Journal of Secondary Gifted Education*, Vol. 17, No. 1, pp. 20-36. https://doi.org/10.4219/jsge-2005-389

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S. and Vos, P. (2016) *Assessment in Mathematics Education*, Cham: Springer. https://doi.org/10.1007/978-3-319-32394-7

Swan, M. and Burkhardt, H. (2012) 'A Designer Speaks: Designing Assessment of Performance in Mathematics', *Educational Designer*, Vol. 2, No. 5.

Syafril, S., Aini, N. R., Netriwati, Pahrudin, A., Yaumas, N. E., and Engkizar (2020) 'Spirit of Mathematics Critical Thinking Skills (CTS)', *Journal of Physics: Conference Series*, Lampung, Vol. 1467, No. 1, pp. 012069. https://doi.org/10.1088/1742-6596/1467/1/012069

Tajudin, N. M. and Chinnappan, M. (2015) 'Exploring relationship between scientific reasoning skills and mathematics problem solving', in Marshman, M., Geiger, V. and Bennison, A. (ed.) *Mathematics education in the margins: Proceedings of the 38th annual conference of the Mathematics Education Research Group of Australasia*, Australia: Mathematics Education Research Group of Australasia.

Valanides, N. (1997) 'Cognitive Abilities among Twelfth-grade Students: Implications for Science Teaching', *Educational Research and Evaluation*, Vol. 3, No. 2, pp. 160-186. https://doi.org/10.1080/1380361970030204

Vilenius-Tuohimaa, P. M., Aunola, K. and Nurmi, J. (2008) 'The association between mathematical word problems and reading comprehension', *Educational Psychology*, Vol. 28, No. 4, pp. 409-426. https://doi.org/10.1080/01443410701708228

Wong, K. Y. and Lee, N. H. (2009) 'Singapore Education and Mathematics Curriculum', in Wong, K. Y., Lee, P. Y., Kaur, B., Foong, P. Y. and Ng, S. F. (ed.) *Series on Mathematics Education*, Singapore: World Scientific. https://doi.org/10.1142/9789812833761_0002

Wu, M. and Adams, R. (2006) 'Modelling mathematics problem solving item responses using a multidimensional IRT model', *Mathematics Education Research Journal*, Vol. 18, No. 2, pp. 93-113. https://doi.org/10.1007/BF03217438

Zhou, C. (2012) 'Teaching engineering students creativity: a review of applied strategies', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 5, No. 2, pp. 99-114. https://doi.org/10.7160/eriesj.2012.050205

Zimmerman, C. (2007) 'The development of scientific thinking skills in elementary and middle school', *Developmental Review*, Vol. 27, No. 2, pp. 172-223. https://doi.org/10.1016/j.dr.2006.12.001