Athens Journal of Education 2024, 11: 1-16
https://doi.org/10.30958/aje.X-Y-Z

Development of Critical thinking through the Creation of Mathematical Problems

By Valéria Švecová[*], Marta Balgová[±] & Veronika Uhríková[°]

Through creative thinking, it is possible to acquire new attitudes and a view of the world, not only of mathematics but from several sides (Kamp, 2016). Creativity in mathematics can be defined as a process based on sensitivity to problems, deficiencies, gaps in knowledge and sensitivity to identifying missing elements, revealing difficulties in finding solutions, making estimates, formulating hypotheses, and verifying them and then forming a conclusion (Mann, 2006). The concept of critical thinking is closely related to creativity. Its essence, not only in mathematics, lies in the way of formulating and asking questions. Broader assigned tasks or complex problems lead students to new data and knowledge, which naturally creates the ability to think critically. The paper focuses on problem posing and the development of critical thinking through the creation of tasks based on Bloom's taxonomy and criterial testing requirements. Part of the paper will also be a description of the process of creating math problems, their evaluation, compiling a test based on various criteria, diagnosing students based on the results and proposing further work with students.

Introduction

Nowadays, when, among other things, emphasis is placed on testing students in different grades of elementary school, teachers are also required to have a creative approach to creating assignments. Creating tasks is a creative and demanding process, especially if it is necessary to create tasks according to precisely defined criteria, e.g. Bloom's taxonomy. In practice, we encounter that teachers have a problem with this activity. Therefore, we consider it necessary to emphasize the creation of tasks already in the training of mathematics teachers, whether in the 1st or 2nd grade of elementary schools.

Through creative thinking, the solver/student acquires new attitudes and insight into the world, not only mathematics, but from several sides (Kamp, 2016). Creativity in mathematics can be defined as a process based on sensitivity to problems, deficiencies, gaps in knowledge, and sensitivity to identifying missing elements, revealing difficulties in finding solutions, making estimates, formulating hypotheses, and verifying them, and then forming a conclusion (Mann, 2006). The concept of critical thinking is closely related to creativity. Its essence, not only in mathematics, lies in the way of formulating and asking questions. Broader assigned tasks or complex problems lead pupils/students to new data and knowledge, which naturally creates the ability to think critically.

[*]Assistant Professor, Constantine the Philosopher University in Nitra, Slovakia.
[±]Department of Evaluation and monitoring of Education, NIVAM Bratislava, Slovakia.
[°]Elementary School with Kindergarten, Slovakia.

In the PISA monitoring of mathematical literacy, Slovak students achieve long-term results below the OECD average (PISA, 2018). Slovakia as well as some other countries, e.g. Denmark responded to the PISSA results by introducing a set of national tests (Andersen & Nielsen, 2016). So far, nationwide measurements of pupils' educational results in mathematics have been implemented in the Slovak Republic with the aim of maximum objectivity and differentiated assessment of pupils' performances - normative tests. They are comparative tests, the aim of which is to create a ranking of the tested students according to their success in the given test. Part of the ongoing education reform in Slovakia will also be innovation in the evaluation of educational results, primarily using criterion-referenced tests. In Sweden, research was conducted to examine the relationship between TIMMS testing and Swedish national measures. The result is a very strong correlation between the two tests, meaning that if a student was successful in the national testing, they were also successful in the TIMMS measure (Wiberg, 2019). This is also the reason for changing the national testing in Slovakia, we want to get closer to the PISA and TIMMS measurements. The goal of the contribution was to bring closer the process of creation and evaluation of test tasks suitable for national testing for pupils of the 3rd grade of primary schools.

Curricular reform is currently underway in Slovakia. Its aim is to create educational content organized in three multi-year cycles in primary schools. The intention is that teaching, instead of handing over ready-made information, creates situations in which students can interpret information in confrontation with real experience (Ministry of Education, Slovak Republic, 2023).

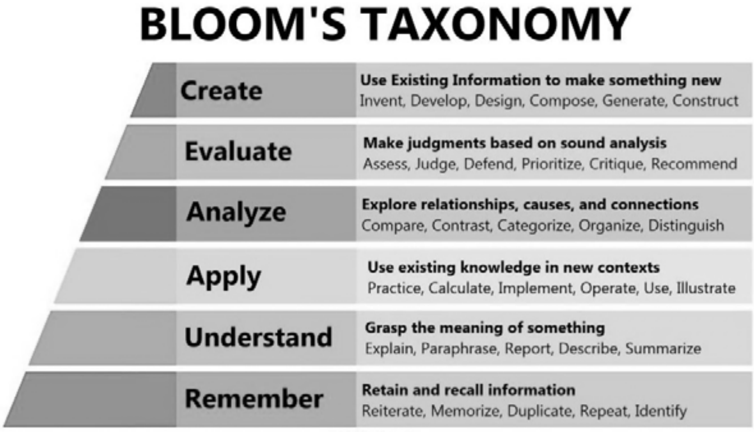
The aim of the article is to bring closer the process of creation and evaluation of tasks suitable for the prepared criterion tests in mathematics for the 3rd grade of elementary school (9-10 years) within the framework of the planned curricular reform. Curricular reform also includes innovation in the evaluation of educational outcomes through new testing methods. The aim of our research is to create a tool for criterion testing in mathematics. The final nationwide criterion testing of third graders will be in 2026. Until then, it is necessary to create tests, pilot, evaluate and possibly change individual tasks.

As stated by (Ficová, Pichaničová, 2023) criterion tests are created with the aim of reflecting on the absolute performance of the student. The task of such tests is to check the level of the student's knowledge and abilities in a precisely defined content area. However, the performance of the tested pupil is not compared with the performance of other pupils. The goal is to decide at what level the student controls the educational content and competencies - from the lowest level to the highest. The tests prepared in this way reflect the extent to which the student mastered the specified subject area, what is the minimum, medium, or optimal standard in the given content; to what extent the student knows what he is supposed to know; which specific parts of the curriculum content need to be focused on during tutoring. When constructing tests that are aimed at absolute performance, it is necessary to identify the subject matter that the student must master without problems. This curriculum is subsequently transformed into test tasks. To verify the mastery of a certain subject matter, it is required that each tested phenomenon be covered by a larger number of test tasks. They then need to be piloted, verified, and evaluated. Based on the evaluation, the task can then be included in the final test.

Literature Review

Problem posing and problem solving are one of the important topics in mathematics education. Problem posing covers a whole group of new problems that relate to creating a problem or reshaping an already formulated problem. Problem posing can also be used as a diagnostic or evaluation tool, which helps teachers to reveal deficiencies and obstacles in students' knowledge. Problem posing has the potential to be used as a diagnostic/evaluation tool (Papadopoulos, 2021). Both methods have the potential to improve the educational process at the primary level by strengthening students' problem-solving, creativity and independent thinking skills. Their effective use depends on teacher preparation and appropriate integration into the curriculum. According to Tichá and Hošpesová (2013) based on the raised problems, it is possible to examine both the level of understanding and the difficulty of a specific mathematical concept. Not all tasks are equal if we assess them from the point of view of their demands on the cognitive activities that need to be performed when solving them. Correct problem posing is very important for effective mathematics teaching. Teachers must also be able to formulate and set meaningful mathematical problems for their students. The problems that a teacher poses can affect the learning of mathematics in the classroom and "their other mathematical goals for the classroom" (Ferrini-Mundi, 2000, p. 53). Teachers can use problem-based tasks to find out how their students understand and apply mathematics in real life (Cai & Hwang, 2002; Cai & Hwang, 2020; Kotsopoulos & Cordy, 2009). Research by Hamidi, Soleymani, Dazy, & Meshkat (2024) also points to the connection between problem posing and problem solving on TIMMS measurement. Many experts from practice, as well as ministries of education of several countries, try to incorporate problem posing into teaching to increase the level of mathematical literacy as part of curricular changes (e.g. Brown & Walter, 1983; Healy, 1993; Ministry of Education, Slovak Republic, 2023). We can use different taxonomies by problem solving. Bloom's taxonomy is used in our conditions when assessing cognitive demands. The framework elaborated by Bloom and his collaborators consisted of six major categories: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation(Armstrong, 2010). In 2001 Bloom's taxonomy was revised (Figure 1).

Figure 1. Revised Bloom's Taxonomy



Source: Drew (2023)

Methodology

Three pilot tests were created and evaluated in 2023. Each one consisted of 30 tasks. We present two of them passed pilot testing and subsequent evaluation.

Since tasks from criterion tests can be repeated, they will not be published. After testing, the tasks that will contain them will be selected. The same principle is also used in the international PISSA or TIMMS tests.

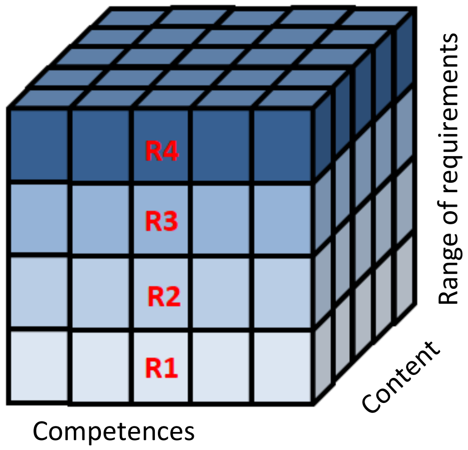
Each created task, which is included in the database, contains a table. It contains the characteristics of the task according to the theoretical framework, as well as the description of the task itself.

Mathematical problems for criterion tests are created based on a theoretical framework.

The theoretical framework is created due to the need to identify each task that is included in the task database. This database is then used to compile individual criterion tests. The creation of the theoretical framework was based on a holistic model that is uniform for all areas of mathematics teaching and levels of education. Existing frameworks used in mathematics education worldwide (PISA, TIMMS, VERA model, existing models in the USA and other countries) served as inspiration (Švecová, Balgová, 2023).

The theoretical framework contains four levels of the range of requirements (characteristics of tasks and their cognitive level) and five levels of competence, each of which has 3 more sub-levels. All domains are mutually independent (or existing dependencies are not intentional). (Figure 2).

Figure 2. Theoretical Frameworks' Model



Source: Ficová & Pichaničová (2023)

The characteristics of the framework for assigning tasks for the 1st cycle are shown in Table 1.

Table 1. The Characteristics of the Framework

	Range of requirements	**R1**	basic understanding, direct application of concepts
		R2	the need to determine a method, an algorithm
		R3	connection of several areas of mathematics, interdisciplinary relations
		R4	evaluation of correctness, truth
Mathematical competences	*Communication*	**C1**	understanding – 1st and 2nd level of Bloom`s taxonomy
		C2	application – 3rd level of Bloom`s taxonomy
		C3	analyse – 4th, 5th level of Bloom`s taxonomy
	Argumentation	**A1**	assess the correctness of calculation procedures
		A2	assess the relevance of the results
		A3	argue the truth
	Symbolic language	**S1**	standard notations, formulas
		S2	work with several formulas, find connections
		S3	complex statements, non-standard representations
	Modelling and mathematization	**M1**	recognize a known model
		M2	recognize a less known model
		M3	create, verify the model

Table 1 shows the two dimensions of the model from Figure 2. Table 2 focuses on the 3rd dimension – content standards for 1st cycle.

Table 2. The Characteristics of content Standards

Component		Content standard
I.	Numbers and operations with numbers	Natural numbers up to 10.000
		Addition and subtraction of natural numbers up to 1,000
		Multiplication and division of natural numbers up to 1000
		Numerical expressions in problem solving
II.	Dependencies, relationships and working with data	Basics of working with patterns and sequences
		Basics of working with dependencies and relationships
		Basics of solution elementary combinatorial problems
		Observation of elementary probabilistic situations
		Basics of working with data
III.	Geometry	Orientation in plane and space
		Basics of exploring 2D geometric shapes
		Basics of exploring 3D geometric shapes
		Elementary measurement procedures
		Basics of exploring and creation of symmetrical shapes

We used the IRT model for statistical evaluation of the criterion test. IRT (Item Response Theory) is an alternative to classical test theory (CTT) and allows to find out the properties of items that classical test theory does not provide. The IRT models were selected in designed and evaluating of tests as the main method for several reasons. First, one advantage of IRT models over linear factor analytic methods is that information from examinee response patterns is analysed as opposed to the more limited information from correlation matrices. Second, nonlinear models such as IRT models may better reflect the relationship between item performance and the latent ability. Third, the multidimensional IRT models have been used to assess dimensionality of tests in which items reflect different skills, knowledge, or cognitive processes (Švecová et al., 2022). However, it has stronger prerequisites for use. IRT models with one latent variable use two assumptions:

- unidimensionality – the respondent's probability of answering an item is influenced by only one of his characteristics and any other influences are excluded,
- local independence – apart from the characteristics of the person, the influence of other test items on the answer is not considered (Švecová et al., 2022).

The 2 – Parameter Logistic model generalizes the unidimensional model by adding a new item parameter known as the discrimination parameter (Tendeiro, 2017). We compared unidimensional and 2-PL models for evaluation our test:

- The unidimensional model: we assumed that behind the different types and contents of tasks there is one common ability.
- The 2-Parameter Logistic model: we assumed that there are two distinguishable abilities behind the different types and contents of tasks.
  The comparison of 1- and 2-dimensional model (full test):
- We modelled two dimensions Geometry, measurement and Addition, subtraction (thematic unit).
- Dimensions were equally represented (24 items in each)
- Adding the 2nd dimension will slightly improve the fit of the model with the data, however, the correlations between the dimensions are very high (r = 0.944)
- The implementation of factor analysis also confirms the assumption of one-dimensionality.
  The unidimensional model seems to be more suitable for our data.

When verifying the task, its unsolved, difficulty, sensitivity and success were determined.

Results and Discussion

We present the characteristics of two of our created tasks, which were included in the pilot test intended for 3rd grade students (end of 1st cycle).

Table 3. Characteristics of the task Shopping in Diagon Alley

Shopping in Diagon Alley	
Level:	1st cycle
Performance standard	Solving application problem related to orientation in a table
Type of task	open
Theoretical framework	
Range of requirements: R2 – the need to determine a method, an algorithm. <u>Mathematical competences</u> C2 – understand simple content and apply calculation. S1 – identify relevant information in the text. M1 – acquire and use known models.	
Task's characteristics: Knowledge of numbers and operations with their is required when solving the problem. A pupil is expected to find, link, and analyze data from the text and the table. A basic understanding in a real-world context is required. The task is aimed at identifying the necessary data from the table. The table also contains redundant information.	

As can be seen from Table 3 the task was focused on working with a table and text. Three hundred and forty-six pupils from 39 schools throughout Slovakia solved the test with this task. Of the total number of pupils, there were 179 boys and 167 girls. Likewise, during the statistical processing, the seat of the school was also considered, i.e. how many pupils attended a rural school and how many pupils attended an urban school. Based on this division, 219 students from urban schools and 127 students from rural schools solved the test with the given task.

Statistical processing was carried out within NIVAM (National Institute of Education and Youth). During the evaluation of the task, it was found that the gender and the location of the school did not statistically significantly influence the solution of the task itself.

As we already mentioned in the methodology, we determined the unsolvedness, difficulty, sensitivity and success of mathematical problems.

Unsolved task means how many pupils did not even try to solve the task. In our case, it was less than 5% of all pupils, which means that only 17 pupils did not even try to solve the given task. Based on these data, we can conclude that the task was attractive enough for the students. Failure to solve the task in general can also affect its order in the test. This task was in 13th place out of 30 tasks, from which we can conclude that its position does not need to be changed.

Difficulty is a complex data from a didactic point of view, the structure of which is conditioned by the structure of the teaching material and the requirements for logical-recognition operations when assigning tasks. The intended difficulty of the test must be based on the properties of the components that make up the learning material and on the task itself:

1. degree of importance of the learning element in the curriculum structure,
2. difficulty and age appropriateness, logical-cognitive operation,
3. complexity of the required intellectual and manipulative skills.

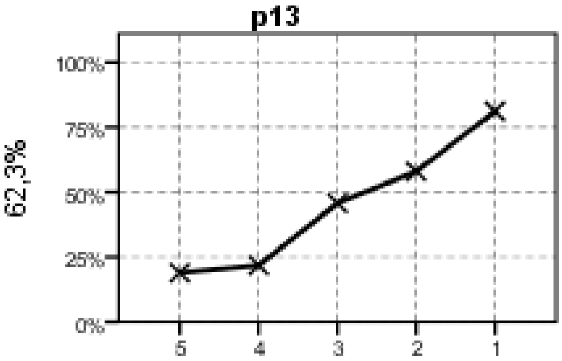
The difficulty index can have a value from 0 to 100 %. The more pupils solve the problem correctly, the higher the percentage. Easier tasks have a higher difficulty index. (Kubiš, 2015).

In the statistical processing of the individual tasks of the test, the task created by us received approximately 45%, which means that it was a moderately difficult task. This level may mean that the third graders have some difficulty in identifying data according to the given criteria, as it requires linking reading comprehension skills and chart orientation.

The sensitivity of the test expresses the extent to which the test can distinguish between students with good and poor knowledge (discriminative power of the test): it is especially important for NR tests (discriminatory). Here, task piloting is essential, after which each task is analysed separately. It is necessary to determine the distinguishing value of the tasks, that is, the sensitivity. The high sensitivity test contains tasks of different difficulty (easy, medium, hard). Sensitivity refers to the ability of a test to distinguish between students with different levels of actual knowledge and skill. In a sensitive test, students' results should be adequately distributed across the entire point scale. The test is not sensitive if all pupils achieve an excellent or poor result in the test - it does not differentiate between pupils (Kubiš, 2015).

The statistical evaluation of the task is shown in Graph 1 and subsequent Table 4.

Graph 1. Sensitivity of the Task with Respect to Performance Groups



Performance group (PG)

It can be seen from the Graph 1 that the success of solving the task with the performance group is increasing. This means that students with a higher level of knowledge achieved better results in solving the given task than students with a lower level of knowledge. The performance group is divided on Graph 1 according to the assessment that the tested students achieved in mathematics. Table 4 is also related to Graph 1.

Table 4. Sensitivity of the Task with Respect to Performance Groups

	Percentile Group of number				
	PG1	PG2	PG3	PG4	PG5
p13	81,2	58,0	45,7	21,7	18,8

From Table 4, we can see that more than 81% of pupils - excellent mark (PG1) were successful in solving the given task, i.e. they solved the task correctly. As the evaluation worsens, the percentage of successful solvers also decreases. So, for example, in performance group 3 (VS3) - mark good, there were only less than 46% of successful solvers.

Graph 1 and Table 4 show that this task differentiates pupils well according to their performance. Therefore, it can be used without change as a task for other tests.

The 2nd presented mathematical problem was from 2nd test. This test was solved by 340 pupils from 39 schools throughout Slovakia. Of the total number of pupils, there were 163 boys and 177 girls, 217 pupils from urban schools and 123 pupils from rural schools solved the test with the given task.

Table 5. Characteristics of the Task Toys

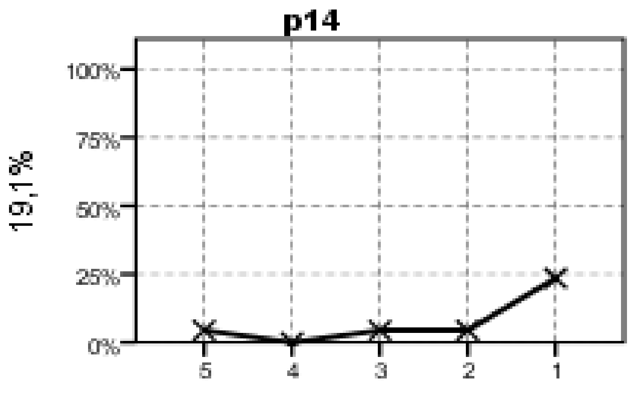
Toys	
Level:	1st cycle
Performance standard	The number of possibilities to pay the amount
Type of task	open
Theoretical framework	
Range of requirements: R3– connection of several areas of mathematics. Mathematical competences C2 – understand content and apply calculation, analyze options. S2 – connect, find connections in the data, choose, and use the representation of mathematical objects. M2 – recognize a model that is not common to students, interpret the results.	
Task's characteristic: When solving the task, knowledge from several areas is needed: the solution of simple combinatorial situations and numbers and variables. The pupils are expected to find and connect data from text and image. It requires a basic understanding in a real context. The problem is aimed at adding two or three natural numbers with and without a transition to 20 from memory. It requires a basic understanding in a real context. The expected solution from the student is to list all the possibilities of one, two or three numbers, the sum of which is less than 20. The expected solution is finding the number of possibilities.	

In the assignment there are pictures of four toys with prices – 18€, 9€, 8€, and 3€. Girl has 20€. The question is: How many of all options did she have to buy at least two toys?

This problem was for pupils very difficult. In the statistical processing of the test tasks, this task received only about 7%.

This task was in 14th place out of 30 tasks. Unsolved of this problem was higher, 12%. Based on these data, we can conclude that the task was less attractive for the pupils. We assume it has something to do with combinatorics. Third - graders in Slovakia have almost no experience with this area, despite it being part of the curriculum. The most common incorrect answer was: 3 options. We assume that the pupils forgot the possibility to buy three toys at the price of 9€, 8€, and 3€. Graph 2 shows sensitivity of the task with respect to performance groups.

Graph 2. Sensitivity of the Task with Respect to Performance Groups



Performance group (PG)

Unlike the first task, this one does not differentiate well. As can be seen from the graph, almost all income groups achieved low success rates. Table 6 shows that in the highest performance group the success of the solution was only 25%.

Table 6. Sensitivity of the Task with Respect to Performance Groups

	Percentile Group of number				
	PG1	PG2	PG3	PG4	PG5
p14	23.5	4.4	4.4	0.0	4.4

A task with such parameters cannot be included in the test. Therefore, we decided to change its assignment in the second pilot testing. New question is: How many different pairs of toys can she buy if she has 20€?

We expect that changing the word "options" to the word "pairs" will increase the success of the solution, as well as the sensitivity and reduce the unsolved. We also simplified the task by using the word pairs, as we omitted the possibility of buying three toys. The new wording of the task will be piloted in 2024.

When evaluating two selected tasks as part of pilot testing, we focused on their difficulty, sensitivity, success and, finally, their unsolved. Statistical evaluation of individual tasks was done through IRT. Based on this processing, we found out what the parameters of the test tasks are. Tasks with good parameters were moved to further pilot testing without changes, and tasks with insufficient parameters were changed based on the analysis. We will pilot test the changed tasks again to find out if they have improved their parameters.

Of course, it is not enough to change only the testing, it is also necessary to change the training of students, but also of future teachers. They must be ready to independently create mathematical problems for their students and thus develop their mathematical literacy and critical thinking. However, it is necessary to avoid the fact that teachers do not focus only on preparation for testing, but above all on the comprehensive development of students. (Göloğlu & Kaplan, 2021).

Mathematical tasks influence the teaching and learning character of learners. The nature of student learning is determined in two ways, among others:

1. the type of task,
2. the way it is used (Clarke & Roche, 2010)

Research shows that the structure of lessons and the types of tasks used in mathematics lessons are different around the world (Bangtho et al., 2015). In most European countries, teachers focus on skill development, with students focusing on practicing routines, rules, and patterns, while in Japan, teachers focus on developing conceptual understanding, spending as much time solving challenging problems as practicing skills (Stigler&Hiebert, 1999, Phu-Udon, 2001). Japanese students have a creative way of solving mathematical problems, they invent and present new methods of solving mathematical problems (Shimizu et al., 2010). This may also be one of the reasons why Japanese students achieve above-average results in PISA testing.

Conclusion

Criterion tests are designed to determine whether a pupil has a set of skills, competencies, or knowledge. Their goal is not to compare the pupil with other pupils of the same age. Test makers analyse components of specific academic skills, such as number comprehension and then write test items that measure whether the child has all components of the skill. The tests are standardized in terms of the level of skills a child should have. Nevertheless, the tests are constructed to measure how a child is acquiring specific skills (Berger, 2013, Webster, 2021).

According to (Webster, 2021) criterion math test should reflect the range and sequence of state standards (such as the common core national standards). It reflects the skills needed at any age: for young pupils, understanding one-to-one correspondence, counting and at least addition as an operation. As the child grows, he is expected to acquire new skills in an appropriate sequence that builds on earlier levels of skill acquisition (numerical operations with natural numbers; relations, functions, tables, and diagrams; geometry and measurement; combinatorics, probability, statistics; logic, reasoning, evidence). A criterion math test should reflect the range and sequence of state standards (such as the common core national standards). It reflects the skills needed at any age: for young pupils, understanding one-to-one correspondence, counting and at least addition as an operation. As the child grows, he is expected to acquire new skills in an appropriate sequence that builds on earlier levels of skill acquisition (numerical operations with natural numbers; relations, functions, tables, and diagrams; geometry and measurement; combinatorics, probability, statistics; logic, reasoning, evidence).

According to (Kantorov, 2000) it is not appropriate to use only one method of evaluating the performance of a pupil (and thus also of a teacher, class, school) on a national scale. school principals, teachers, parents, and students should learn as much as possible about the tests used to assess mathematics achievement.

The more we know about what is on the national tests that students are required to take and how those tests are scored, the better you can determine the meaning and usefulness of the results. Criterion testing in Slovakia is only in the beginning. Nationwide testing in 3 cycles of mathematics (after the 3rd, 5th and 8th grade of elementary school) is planned from 2026. Until then, pilot testing will take place in these grades of elementary school. These will provide test makers with feedback on tasks as well as entire tests.

The teaching of mathematics and especially its content has been reduced over time not only in Slovakia, but the prevailing opinion is also that creating high-quality mathematical tasks is simple and straightforward. We consider this assumption to be erroneous. Mathematics is required for accurate reasoning, and errors easily creep into the words, numbers, and symbols that express assessment tasks (National Research Council, 1993).

The created methodology represents an imaginary "starting line" that will allow, through the gradual automation of individual processes over time, to implement this tool in the environment of educational reality within the framework of planned innovations in the field of evaluation of educational results following the curricular reform implemented from the Recovery and Resilience Plan. The introduction of criterion-referenced testing in schools will provide another dimension of viewing the pupil's performance, which should primarily serve for the individual assessment of the pupil's performance within the tested content and the possibility to clearly identify his strengths and weaknesses. Test tasks linked to learning outcomes in mathematical literacy could eventually be incorporated into the new curriculum as Shepard, Daro & Stancavage (2013) states. The results of several research show that the mathematical literacy of elementary school pupils is still insufficient. As the authors state, this may also be due to a limited number of problems based on mathematical literacy. These researchers also state that research analysing mathematical literacy also requires a higher level of critical thinking (Rachman & Amir, 2022, Little et all., 2020, Little et al., 2022). Criterion tests mainly contain mathematical problems based on real life, and therefore it is necessary to pay attention to the process of their creation. Based on the results of several research, each student has a different level of literacy. That is why tasks in criterion tests must be created in such a way that this level can be identified and defined (Nurutami et al., 2018).

Creating mathematical problems should be part of teaching (Lieberman, 2018). For some students and teachers, it is difficult to find several correct solutions to the task (Roth & Ames, 2014). Finding different solutions to open mathematical problems helps to develop creativity not only among students but also among teachers. Through the development of creativity, students' relationship to mathematics can also be improved (Švecová & Pavlovičová, 2012, Švecová et al., 2014). If we want to focus on the development of critical thinking and mathematical literacy of primary school pupils, it is necessary to develop these components in teachers of the primary level of education as well. Papanikos (2023) also draws attention to the need to focus on the education and evaluation of future teachers at the primary level of education

Acknowledgements

This work was also supported by the Cultural and Educational Grant Agency under contract No. 014UKF-4-2023 Innovation of Education in the Training of Mathematics Teachers in the Field of Task Creation with Emphasis on Criteria Testing.

References

Andersen, S.C., & Nielsen, H.S. (2016). *The Positive Effects of Nationwide Testing on Student Achievement in a Low-Stakes System* (July 9, 2016). Available at SSRN: https://ssrn.com/abstract=2628809 or http://dx.doi.org/10.2139/ssrn.2628809

Armstrong, P. (2010). *Bloom's Taxonomy*. Vanderbilt University Center for Teaching. Retrieved [todaysdate] from https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/.

Bangtho, K., Inprasitha, N., & Inprasitha, N. (2015). Process Designing Mathematical Tasks on Addition of Teachers Using Lesson Study. *Creative Education*, 6, pp.1691-1696. doi: 10.4236/ce.2015.616170

Berger, M. (2013). Criterion-Referenced Testing. In: Volkmar, F.R. (eds) *Encyclopedia of Autism Spectrum Disorders.* Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1698-3_146

Brown, S. I., & Walter, M. I. (2005). *The art of problem posing*. Psychology Press.

Cai, J., & Hwang, S. (2002). Generalized and generative thinking in US and Chinese students' mathematical problem solving and problem posing. *The Journal of mathematical behavior, 21*(4), 401-421.

Cai, J., & Hwang, S. (2020) Learning to teach through mathematical problem posing: Theoretical considerations, methodology, and directions for future research. *International Journal of Educational Research, 102,*

Clarke, D., & Roche, A. (2010)/ *Teachers' Extent of the Use of Particular Task Types in Mathematics and Choices behind That Use.* http://files.eric.ed.gov/fulltext/ED520857.pdf

Drew, C. (February 17, 2023). *All 6 Levels of Understanding (on Bloom's Taxonomy)*. https://helpfulprofessor.com/levels-of-understanding/

Ferrini-Mundy, J. (2000). Principles and standards for school mathematics: A guide for mathematicians. *Notices of the American Mathematical Society, 47*(8).

Ficová, L., & Pichaničová I. (2023). Kriteriálne testovanie ako nástroj na mapovanie absolútneho výkonu žiaka. *Pedagogické rozhľady*, 2023(3), 20–27. Dostupné z http://archiv.mpc-edu.sk/sites/default/files/rozhlady-casopis/pr-3-2022.pdf

Göloğlu D.C., & Kaplan K.Ö. (2021). The impact of high stakes testing on the teaching and learning processes of mathematics. *Journal of Pedagogical Research, 5*(2), 119-137.

Hamidi, F., Soleymani, S., Dazy, S., & Meshkat, M. (2024). Teaching Mathematics based on Integrating Reading Strategies and Working Memory in Elementary School. *Athens Journal of Education*. 11. 9-22. 10.30958/aje.11-1-1.

Healy, C.C. (1993). *Creating miracles: A story of student discovery*. Key Curriculum Press.

Kamp, A. (2016). *Engineering Education in the Rapidly Changing World*: Rethiking the Vision for Higher Engineering Education. 2. vydanie, TU DEFT, 2016, 88sSrgs

Kantrov, I. (2000) Assessing Students' Mathematics Learning, *Issues in Mathematics Education.*

Kotsopoulos, D., & Cordy, M. (2009). Investigating imagination as a cognitive space for learning mathematics. *Educational Studies in Mathematics*, *70*, 259-274.

Kubiš, T. (2015). *Metodika tvorby testov a testových úloh*. Bratislava: Národný ústav certifikovaných meraní, 153s., ISBN 978-80-89638-28-4

Lieberman, J. (2009). Using Lesson Study to Develop an Appreciation of and Competence in Task Design." In B. Clark, B. Grevholm, & R. Millmand (Eds.), *Task in Primary Mathematics Teacher Education: Purpose, Use and Exemplars, Mathematics Teacher Education* (pp. 11-24), Springer Science + Business Media LLC.

Little, C.W., Lonigan, C.J., & Phillips, B.M. (2020). Differential patterns of growth in reading and math skills during elementary school. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000635

Little, C.W., Erbeli, F., Francis, D.J., & Tynan, J. (2022). Developmental trajectories for literacy and math skills from primary to secondary school. *Journal of Research in Reading, 45*(1), 65–82. https://doi.org/10.1111/1467-9817.12382

Mann, L., E. (2006) Creativity: The Essence of Mathematics In *Journal for the Education of the Gifted*, vol. 30, no 2. dostupné na internete https://journals.sagepub.com/doi/abs/10.4219/ jeg-2006-26

The Ministry of Education, Science, Research and Sport of the Slovak Republic (MESRS SR) (2023) Komponent 7: Vzdelávanie pre 21. storočie. https://www.minedu.sk/komponent-7-vzdelavanie-pre-21-storocie/

National Academies of Sciences, Engineering, and Medicine (1993). *Measuring What Counts: A Conceptual Guide for Mathematics Assessment*. Washington, DC: The National Academies Press. https://doi.org/10.17226/2235.

Nurutami, A., Riyadi, & Subanti, S. (2018). The Analysis of Students' Mathematical Literacy Based on Mathematical Ability. *Atlantis Press*, 157(Miseic), 162–166. https://doi.org/10.2991/miseic-18.2018.40

Papadopoulos, I., et al. (2021). Multiple Approaches to Problem Posing: Theoretical Considerations Regarding its Definition, Conceptualisation, and Implementation. *Center for Educational Policy Studies Journal*, 1-22.

Papanikos, G.T. (2023). The Evaluation and the Accreditation Process of Greek HEIs with an Emphasis on Primary Education Departments. *Athens Journal of Education*. 10. 173-186. 10.30958/aje.10-1-10.

PISA (2018). Technical Report chapter 15 https://www.oecd.org/pisa/pisa-for-development/pisafordevelopment2018technicalreport/PISA-D%20TR%20Chapter%2015%20-%20Proficiency%20Scale%20Construction%20-%20final.pdf

Phu-Udom, J. (2001). *Development of Mathematics Teaching Models with Students as Knowledge Creators. Doctor of Philosophy Program in Curriculum and Instruction*, Bangkok: Srinakharinwirot University, 2001.

Rachman, B.A., & Amir, M.F. (2022). Primary Students' Math Literacy in terms of Higher Order Thinking Skill. *Mimbar Ilmu*, *27*(2), 188–199. https://doi.org/10.23887/mi.v27i2.48435

Roth, J., & Ames, J. (2014). *Beiträgezum Mathematikunterrich*t (pp. 1043-1046). Münster: WTM-Verlag

Shepard, L., Daro, P., & Stancavage, F. (2013). The Relevance of Learning Progressions for NAEP A Study of NAEP Reading and Writing Frameworks and Assessments in Relation to the Common Core State Standards in English Language Arts.

Shimizu, Y., Kaur, B., Huang, R., & Clark, D.J. (Eds.) (2010). *Mathematical Task in Classrooms around the World*. Sense Publisher.

Stigler, J.W., & Hiebert, J. (1999). *The Teaching Gap; Best Ideas from the World's Teachers for Improving Education in the Classroom.* New York, NY: The Free Press.

Švecová, V., & and Pavlovičová, G. (2012). Reflection of Pupil's Models of Numbers in Mental Map. *Procedia - Social and Behavioral Sciences*. 46. pp. 5311-5314, 2012. 10.1016/j.sbspro.2012.06.429.

Švecová, V., Rumanová, L., & Pavlovičová, G. (2014) Support of Pupil's Creative Thinking in Mathematical Education. *Procedia-Social and Behavioral Sciences,* 116, pp. 1715-1719, 2014.

Švecová, V., Rybanský, Ľ. & Pavlovičová, G. (2022). Personal Need for Structure and Fractions in Mathematical Education. *Eur. J. Investig. Health Psychol. Educ.* 2022, *12*, 448-457. https://doi.org/10.3390/ejihpe12050033

Švecová, V., & Balgová, M. (2023). Inovácia hodnotenia výsledkov vzdelávania prostredníctvom nových spôsobov testovania v kurikulárnej reforme. Edukácia, 6 (2), 72-82.

Tendeiro, J.N. (2017). IRT (GMMSGE01): Parametric IRT (dichotomous data). https://bookdown.org/jorgetendeiro/ParametricIRT/

Tichá, M., & Hošpesová, A. (2013). Developing teachers' subject didactic competence through problem posing. *Educational Studies in Mathematics*, 83(1), 133–143.

Webster, J. (2021). "*Criterion-Referenced Tests: Measuring Specific Academic Skills*" ThoughtCo, Jul. 31, 2021, thoughtco.com/criterion-referenced-tests-measuring-academic-skills-3110860.

Wiberg, M. (2019). The relationship between TIMSS mathematics achievements, grades, and national test scores. *Education Inquiry*, *10*(4), 328–343. https://doi.org/10.1080/20004508.2019.1579626