Journal of Articles
Faculty of Education, Užice
Volume 28 • No 27 • December 2025
https://doi.org/10.5937/ZRPFU2527251T

UDC 37.091.3::51
Original research paper
Pages 251–294
Submitted: 14th March 2025
Accepted: 8th September 2025

DOES TASK FORMAT MATTER? AN EMPIRICAL STUDY OF THE USE OF MULTIPLE-CHOICE AND CONSTRUCTED-RESPONSE TASKS IN GEOMETRY TEACHING

Dragana Lj. Trnavac*
Primary School "Kneginja Milica", Belgrade
Zorana L. Lužanin*
University of Novi Sad, Faculty of Sciences

ABSTRACT

This paper aims to examine the use of multiple-choice (MC) and constructed-response (CR) tasks in the teaching of geometry in the older grades of primary school, with a special focus on the solving strategies that students apply. This topic is significant because the choice of task format can affect students' success, motivation, and development of metacognitive skills, which are the essential elements for effective mathematics learning. The main research was conducted on a sample of 486 seventh and eighth-grade students from three primary schools, using 12 geometrical tasks. The results show that students solve tasks significantly better in the multiple-choice format, which can be partly explained by the use of a guessing strategy in this format. Analysis that took into account both student and class performance confirmed that task format remains a significant performance factor. However, observations revealed that students predominantly employed direct problem-solving strategies, with inverse approaches being used infrequently, even when such strategies were available within the task format. In addition, task formats are associated with differences in motivation and approach to solving, which points to the need for conscious selection of task formats in teaching. The research emphasizes that both formats have their advantages and that the joint application of both formats can contribute to adapting the teaching process to the students' needs.

Key words

format of math task, multiple-choice tasks and constructed-response tasks, solving strategies, older grades of primary school.

* dragana.savic@dmi.uns.ac.rs, https://orcid.org/0000-0002-0919-5049
* zorana@dmi.uns.ac.rs, https://orcid.org/0000-0002-7215-2252

Introduction

The teaching of mathematics is mainly based on the use of tasks, which are the basic instrument for learning, monitoring, evaluation and assessment. Assignments are defined as any tasks that students solve independently or with minimal help, and whose context and presentation method have a significant impact on student achievement. A well-designed task can motivate students and encourage them to think creatively, while teachers monitor their abilities as well as their understanding of mathematical concepts and reasoning (Canogullari & Radmehr, 2025; Watson & Ohtani, 2015).

The choice of tasks in mathematics education plays an important role because it shapes students' cognitive processes, improves conceptual understanding, supports differentiated instruction, develops basic skills and ultimately contributes to a more interesting and effective learning experience. Using a balanced mix of different tasks, teachers can optimize student learning outcomes and foster deeper understanding of mathematics. A variety of tasks is essential for learning, as they serve different functions. It is therefore important to provide students with the opportunity to have contact with different types of mathematics tasks (Martins & Martinho, 2024).

We can divide tasks according to several characteristics, such as context, complexity (cognitive demand) and format. The focus of this article is on task formats. By task format, we mean the way in which the task is structured and presented to students, including the type of response that is expected from them and the form in which they should provide it. Different task formats allow for the assessment of different skills, knowledge and competences.

Tasks can be categorized based on their format as follows: 1. selected-response tasks, which require students to select (circle, underline or otherwise mark) one or more offered answers, and 2. constructed-response tasks, which require students to independently design and write their answers.

In the presented research, multiple-choice (MC), multiple-choice with two correct answers (MCm), and multiple binary choice (MCb) tasks were observed as representatives of selected-response tasks. Constructed-response (CR) tasks can be further classified into extended-response tasks, short-answer tasks, and completion tasks.

There are a number of significant differences in the preparation, use and assessment of tasks in these two formats. One of the key differences between the tasks of these two formats concerns the solving strategies. With MC tasks, students can guess the solution even when they do not have enough knowledge about the task's context. Also, they often use a strategy of eliminating incorrect alternatives or recognizing the correct answer from the options offered. In contrast, CR tasks require students to independently create an answer, thereby encouraging deeper cognitive functions and thinking.

Classifying tasks into MC and CR formats can be useful, but it may create a misleading impression if the variations within each format are overlooked. While MC tasks provide an effective way to assess basic and intermediate levels of thinking, their application is limited when it comes to more complex skills and cognitive processes, for which CR tasks are more appropriate (Martinez, 1999).

The research presented in this paper examines the possibilities of applying MC and CR tasks in the teaching of geometry in the upper grades of primary school. Students' successs in solving tasks, their behavior, including task-skipping, as well as their choice of strategies when solving tasks of the same or similar context, but in different formats, were analyzed. The analysis of students' answers to both formats of the tasks allowed for an examination of the advantages and disadvantages of each format. The results obtained provide a foundation for improving teaching practice and optimization of the choice of tasks in teaching.

Review of the literature

The advantages and disadvantages of using different task formats, particularly MC and CR tasks, have been the subject of considerable research. One of the important differences between the mentioned formats of tasks is related to their composition. CR tasks are simpler to compose compared to MC tasks. While the CR task requires a clear and precise formulation of the question, the MC task consists of two main components, the tree representing the question or problem and the alternatives. The alternatives consist of a correct answer (key) and several plausible but incorrect options (distractors), posing a significant challenge in task design (Clay, 2001; Kusumawati & Hadi, 2018). Haladyna and colleagues (Haladyna, Downing, Rodriguez, 2002) in "A Revised Taxonomy of Multiple-Choice (MC) Item-Writing Guidelines" emphasize that all distractors should be plausible and based on students' typical errors. A study (Özdemir & Toker, 2025) provides insights into distractor design in math problems, showing variations in effectiveness, some lead to misunder-standings, while others accurately assess student understanding. CR and MC tasks often have the same tree, i.e. they differ only in that the answers offered are represented in the MC tasks. However, it is not always possible to compose the task so that both formats have the same tree.

A key advantage of CR tasks is that students can identify their mistakes (Stankous, 2016), allowing for insights into their thinking process (Chaoui, 2011). Thus, the student's focus is not only on the final answer but also on the steps that led the students to the solution (O'Neil Jr. & Brown, 1998). However, some students tend to simply skip the task, i.e. do not attempt to solve it, and thus we do not have information about the students' thinking (Hollingworth, Beard, Proctor, 2007). Bridgeman (Bridgeman, 1992) states the advantages of CR tasks compared to MC tasks. First, constructed-response (CR) tasks reduce measurement error by eliminating the possibility of random guessing. Second, CR tasks eliminate the possibility of correcting the solution through feedback; in contrast, in MC tasks, the absence of the correct answer among the options signals an error, prompting students to revise their strategy. Third, CR tasks cannot be solved backwards, starting from the provided answers. However, the author states that the inability to correct via feedback can also be considered a shortcoming of CR tasks to the extent that feedback can reduce trivial computational errors. A study (Bonner, 2013) examined the impact of format on students' performance and strategy use in solving MC and CR problems of equivalent trees. A statistically significant main effect of format on performance was found, with CR tasks being more difficult. MC tasks were associated with more diverse strategies, such as backtracking and guessing. The author emphasizes that, depending on the purpose of the test, MC tasks are not suitable for describing mathematical achievement and for strengthening the problem-solving strategy that starts with the essence of the mathematical task. According to Haladyna (Haladyna, 2004), good distractors should be chosen by low achievers and ignored by high achievers. Sangwin and Jones (Sangwin & Jones, 2017) examined whether there was an interaction between the procedural task format (MC and CR) and the solution process (direct and inverse). Students' achievements are higher when answering MC tasks in comparison to CR tasks. Moreover, across both task formats, student performance was higher on tasks that required a direct approach to problem-solving compared to those involving inverse reasoning. The findings suggest that when confronted with tasks requiring inverse procedures, students tend to approach multiple-choice (MC) items by verifying the provided answer options directly rather than executing the actual inverse calculation. In a study (Katz, Bennett, Berger, 2000), students solved problems in algebra, arithmetic, and geometry. In order to reduce sources of differences in performance, the CR tasks and the MC tasks had the same tree. Strategies are classified as traditional, those typically associated with solving CR tasks, and non-traditional, those typically associated with solving MC tasks (eg, evaluation of a potential solution). Unexpectedly, the students sometimes used non-traditional strategies in the CR tasks as well. In some problems, the format of the task affects the difficulty, but not the choice of strategy, while in others, the reverse is true. According to Prídavková (Prídavková, 2023), employing diverse problem-solving strategies is crucial for developing mathematical thinking.

A significant difference in the use of MC and CR tasks is reflected in the task evaluation process. Reviewing CR tasks requires more time, and it can depend on the individual who reviews them, that is, there can be inconsistencies in the evaluation (Livingston, 2009; Stankous, 2016; Štěpánková & Emanovský, 2011). In addition to these limitations, Chaoui (Chaoui, 2011) states another one, that tests with constructed-response tasks contain only a few questions, which in some cases is not enough for the given topic. Therefore, an individual's output will differ depending on the group of questions (Livingston, 2009). When evaluating MC tasks, less subjectivity is achieved, and the process itself is shorter; that is, the tasks are easily evaluated (Štěpánková & Emanovský, 2011). This can also be explained by the fact that the review of MC tasks can be done by machine, which makes the scoring process quick and cheap, without disagreements in opinions (Livingston, 2009). In addition to the mentioned advantages of MC tasks, the ability to cover a large number of topics and better test results are also mentioned (Stankous, 2016). A large number of topics implies the use of a large number of questions, so the effect of luck when choosing questions on the result of the examinee is reduced (Livingston, 2009).

MC tasks may not allow recognition of students' misconceptions (Birenbaum & Tatsuoka, 1987) and only the result of the solution process is visible (Andrà & Magnano, 2011). Barton (Barton, 2018) defines and uses diagnostic questions written as MC questions with four provided answers of which one is correct, and each incorrect answer indicates a specific error or misconception. Therefore, when the MC task is an integral part of the test, it is important that none of the alternatives is obviously incorrect, i.e. that it cannot be eliminated as an incorrect answer. When the MC task is used in mathematics teaching, alternatives can provide the teacher with a better insight into the knowledge and understanding that students possess. Choosing the correct answer in MC questions is related more to short-term memory than to long-term understanding of the topic (Stankous, 2016), so these questions are intended to be solved faster, which means that they are not too complex (Zhouf, 2013). However, Goecke and the authors (Goecke, Staab, Schittenhelm, Wilhelm, 2022) considered the possibility of using CR tasks for the same purpose and concluded that it is possible to measure declarative knowledge using different response formats, without losing information about the ranking of individuals. They recommend that, given the choice, researchers and practitioners should use tasks in the MC format, as this format is simpler to administer and less expensive to assess. As one of the reasons for using CR questions, Livingston (Livingston, 2009) states that the respondent can choose the correct answer among the offered answers only if he sees the answer beforehand. It is generally accepted that CR tasks (tasks requiring recall) ask subjects to retrieve information from their memory, while MC tasks (recognition tasks) provide an opportunity for the subject to distinguish the correct answer among the information provided (Birenbaum & Tatsuoka, 1987). Verbić (Verbić, 2013) points out that there is a difference between the number of students who know the answer to a certain question and the number of students who provide the correct answer. This difference is influenced by the possibility of guessing the correct answer, the way the question is presented, and others.

In practice, a combination of these two task formats is often used, which was especially used in large-scale international research, which allows for a deeper insight into the similarities or differences in the application of different formats. The study (Marcq, Donayre, Braeken, 2024) used the answers of 260.000 students from 71 countries who participated in mathematics testing as part of the Program for International Student Assessment (PISA) 2018. The authors investigated how the format of the task (MC and CR) affects the probability of a correct answer and came to the conclusion that the format of the task contributes 12% of the variance in task difficulty. This influence was not the same across countries, while in lower-performing countries task format accounted for 30% of the variance, in higher-performing countries it was 10%. These results highlight challenges in comparability of educational outcomes between different countries. In the paper (Schult & Sparfeldt, 2018), the authors studied how MC and CR tasks differ in terms of reliability and criterion validity, using PIRLS 2006 and TIMSS 2007 scores for fourth grade students. It was concluded that in PIRLS and TIMSS tests, reliability and validity did not significantly depend on the answer format, and that when choosing between MC and CR tasks, other characteristics should be considered, such as development, administration and assessment costs. Using TIMSS 2011 data, a study (Liou & Bulut, 2020) showed that CR tasks were more difficult than MC tasks for students in Taiwan. According to research (Photopoulos, Tsonos, Stavrakas, Triantis, 2021), students prefer MC tasks, although they do not consider them fairer, easier, or less stressful. Students who prefer multiple-choice (MC) tasks appreciate the simplicity of having predefined answer options. In contrast, those who favor constructed-response (CR) tasks feel that the MC format limits their ability to express what they have learned, demonstrate their thinking, and generate original solutions. Hubbard and the authors (Hubbard, Potts, Couch, 2017) investigated the effect of task format by comparing CR tasks and multiple-true-false, which are a special case of MC tasks. Their findings indicate that correct response rates correlate across the two formats but that a higher percentage of students provide correct responses for multiple-true-false tasks.

Research methodology

Research questions. The research questions were designed to provide deeper insight into the differences that task format can have on student achievement and problem-solving processes, as well as potential implications for the teaching process and evaluation of knowledge in mathematics.

The following research questions were asked as part of the research:

1. The influence of the format on the success of the solution: How does task format (MC or CR) affect students’ performance in solving the same mathematical tasks?
2. Answering and Explanation: How does task format influence the frequency of solution attempts?
3. Problem-Solving Strategies: To what extent do students use direct and inverse problem-solving strategies depending on the task format?

The pattern. A total of 889 students from grades five through eight, attending three primary schools in the municipality of Novi Beograd, participated in the study. A detailed analysis was conducted on a subsample of seventh- and eighth-grade students, as they had already acquired the relevant mathematical content, in accordance with the curriculum, enabling them to apply various procedures for solving tasks directly. Students from grades five and six did not have access to all tasks, as part of the mathematical material required for solving them had not yet been covered in their education. Moreover, their flexibility in problem-solving was limited due to a smaller number of available strategies. For this reason, fifth- and sixth-grade students were excluded from the main performance analysis but were included in regression models to examine the effects of grade level and academic achievement on success in solving tasks of different formats. In the first phase of the study, 491 students from grades seven and eight participated (266 seventh-grade and 225 eighth-grade students). Five students withdrew from the test, resulting in a final sample of 486 students, including 227 boys and 259 girls. The average grade in mathematics for this group was 3.84. The second phase involved 398 students from grades five and six (232 fifth-grade and 166 sixth-grade students), including 206 boys and 192 girls. The average grade of the students was 3.86.

The research was approved by the authorities of all three schools and was conducted during the 2021/22 school year. Testing of seventh- and eighth-grade students was conducted between March and April, while testing of fifth- and sixth-grade students took place between May and June. Students completed the test during regular mathematics classes and were informed that participation was anonymous and ungraded.

Research instrument. Twelve tasks from the field of geometry were selected for the purposes of the research. Each task was prepared in both MC and CR formats. Task texts and alternatives were tested in a pilot study, in which 8 students participated. The pilot study was used to correct and improve the formulation of the tasks. The final versions of the tasks are presented in Table 1A in the Appendix. Three tasks were created by the author, while the remaining ones were sourced from existing materials and modified or adapted as needed. During the creation of the tasks, special attention was paid to ensure that the tree in the MC task and the text in the CR task were as similar as possible. When choosing alternatives for MC tasks, recommendations from Haladyna and colleagues (Haladyna et al., 2002) were applied to a large extent. The tasks are grouped into four classes according to the complexity of the requirements.

Chord and *Angle*. These two tasks belong to the simplest cognitive requirements and test factual knowledge of elementary geometric concepts. In the CR task it is necessary to enter only one word or number, while the MC task requires recognizing the correct answer.

Rhombus (Vidović, Stanojević, Stuparević, Stanojević, Vračar, Stančić, 2015: 99), *Trapezoid* (Foy, Arrora, Stanco, 2013: 122), and *Color* (*Final exam at the end of primary education* 2014/2015, task 5). All three tasks refer to the concept of the area of simple geometric figures. Only the *Color* task was put in a real context, where the word "area" is not explicitly mentioned in the task setting. In the MC task *Color*, the provided answer choices indicate the volume of the packages being sold. Therefore, the text of the CR task was modified to explicitly state the volume of the package. The

Rhombus and *Trapezoid* tasks contain images, which simplifies solving because the solution can be obtained by breaking the figure into parts and using only the formula for the area of a rectangle. The answer choices in the *Rhombus* and *Trapezoid* tasks mainly represent the areas of individual figures visible in the image.

Hypotenuse, Staircase (https://aplusclick.org/t.htm?level=9;q=715) and *Triangles* (Matematički list za učenike osnovnih škola, 2017: 17). All three tasks are related to triangles. While the *Hypotenuse* task tests elementary knowledge of the Pythagorean Theorem, the *Staircase* and *Triangles* tasks require a higher level of thinking. In the *Staircase* and *Triangle* tasks, the data are not given in the text, but the student must read them from the pictures and then apply the corresponding statements. The distractors in the MC version of the *Hypotenuse* task are clearly incorrect if the student understands the properties of a triangle, whereas in the MC version of the *Staircase* task, distractors can be eliminated solely by analyzing the image. This selection of distractors allowed testing the application of the elimination strategy in solving MC tasks.

Tangram (BIGZ, Second written assignment A, Question 6), *Tetris* (Final exam at the end of primary education 2018/2019, task 15), *Gardener* (Pavlović Babić & Baucal, 2009: 37) and *Mondrian* (Anić, Košanin, Stanković, 2020: 127) . These tasks are characteristic in that the difficulty of the task changes depending on the format, that is, CR tasks are more difficult. This is especially pronounced in the *Mondrian* task, where the MC variant requires recognition of the Mondrian square, and the CR variant requires the creation of the Mondrian square. The remaining three tasks from this group were used for testing other variants of the MC tasks. The MC variants of the *Tetris* and *Gardener* tasks represent MCb tasks, while the MC *Tangram* task represents an MCm task.

Fifth-grade students can solve the *Rhombus* and *Trapezoid* tasks by breaking a figure into parts and using the formula for calculating the area of a rectangle, but the geometric figures rhombus and trapezoid are covered only later. The *Chord, Angle, Color*, *Tangram, Tetris*, and *Mondrian* tasks can be solved in the fifth grade, the *Staircase* and *Triangles* in the sixth grade, and the *Hypotenuse* and *Gardener* in the seventh grade. Seventh and eighth grade students are given the opportunity to apply multiple problem-solving procedures.

Procedure. To assess students, the twelve tasks (see Table 1A) were divided into three tests (Test 1, Test 2, Test 3) for seventh and eighth grade students, each containing four different tasks. Each test had two versions, version A and version B, which differs from each other in the format of the task, i.e. each version contained two CR and two MC tasks. In this way, it is ensured that each student solves both formats of the task. In each grade, students were divided into two groups that took two different tests. Within the groups, students were divided so that some took version A and others took version B.

Since eight of the twelve tasks include elementary content from geometry, which is covered already in the fifth grade, in the second stage of the research we prepared two tests (Test 4, Test 5), for fifth and sixth grade students according to the same principle as for seventh and eighth grade students. The structure of the five tests used and the number of students who solved them are given in Table 1.

Table 1. Distribution of tasks in tests

The name of the test		Number of students	Test tasks	
			CR format	MC format
TEST 1				
	Version A	74	Hypotenuse Staircase	Chord Rhombus
	Version B	76	Chord Rhombus	Hypotenuse Staircase
TEST 2				
	Version A	92	Tangram Tetris	Color Gardener
	Version B	90	Color Gardener	Tangram Tetris
TEST 3				
	Version A	79	Trapezoid Mondrian	Angle Triangles
	Version B	75	Angle Triangles	Trapezoid Mondrian
TEST 4				
	Version A	95	Trapezoid Mondrian	Chord Color
	Version B	98	Chord Color	Trapezoid Mondrian
TEST 5				
	Version A	101	Tangram Tetris	Angle Rhombus
	Version B	104	Angle Rhombus	Tangram Tetris

The tests are formatted so that students have enough room to work on both task formats. Students were not required to show their work but were instructed to attempt each task to the best of their ability.

Students answered questions about gender, grade and end-of-semester math grade. Teachers were present in the classroom during the 30-minute test.

Data analysis. When solving MC tasks, students have the opportunity to guess the correct answer, so a procedure was used that corrects the success rate and thus eliminates the influence of guessing.

A correction was applied to tasks with four answer choices, only one of which was correct. The correction is calculated according to the following formula: p = q + (100 - q) · 1/4, where p is the obtained percentage of students with the correct result and q is the percentage of students who know the solution (Chan & Kennedy, 2002), so it is

$$q = \frac{4p - 100}{3} \qquad (1).$$

The following statistical analyzes were used in data processing: descriptive statistics, z-test for testing the equality of proportions of two sets at a significance level of .05, logistic regression to evaluate the probability of success of the solution and chi-square test of independence. The statistical package SPSS (version 25) was used for data analysis.

Results

Table 2 shows the percentages of seventh and eighth-grade students who correctly solved nine tasks, in CR and MC format. Each MC task contained four offered answers, only one of which was correct. The fourth and fifth columns of the table show the results of the test of equality of proportions, which examined the statistical significance of the difference in the success of solving between the two formats. In eight of the nine tasks, students achieved significantly higher success in the MC format. The only exception was the *Angle* task, where the difference in the percentage of correct responses between the two formats was not statistically significant.

To neutralize the potential effect of guessing in the MC tasks, the results were additionally corrected by applying formula (1), in accordance with the model (Chan & Kennedy, 2002). After correction, the difference in success between the formats remained statistically significant only in the *Triangles* task, in favor of the CR format (columns 6 and 7).

The last two columns in Table 2 show the percentage of students who did not try to solve the tasks, i.e., they left the task unanswered. The results show that a statistically significantly higher percentage of students skipped tasks in the CR format compared to tasks in the MC format. While the number of students who did not answer the MC tasks was negligibly small, the rate of non-answers for the CR tasks ranged between 9.3% and 42.1%.

Table 2. Distribution of correct answers and non-answers of 7th and 8th grade students on nine tasks in relation to the format of the task

task	format	n	correct answer [%]	z-values	correct with correction	z-values	non-attempted [%]	z-values
Chord	CR	76	19.7	-3.269***	19.7	-0.933	25.0	3.935***
	MC	74	44.6		26.1		2.7	
Angle	CR	75	82.7	0.851	82.7	1.902	9.3	2.774***
	MC	79	77.2		69.6		0	
Rhombus	CR	76	11.8	-4.008***	11.8	-1.479	42.1	4.107***
	MC	74	40.5		20.7		12.2	
Trapezoid	CR	79	24.1	-2.61***	24.1	-0.173	25.3	3.422***
	MC	75	44.0		25.3		5.3	
Color	CR	90	33.3	-3.437***	33.3	-1.603	30.0	5.402***
	MC	92	58.7		44.9		1.1	
Hypotenuse	CR	74	52.7	-2.845***	52.7	-1.748	18.9	3.595***
	MC	76	75.0		66.7		1.3	
Staircase	CR	74	20.3	-2.877***	20.3	-0.372	40.5	5.668***
	MC	76	42.1		22.8		2.6	
Triangles	CR	75	10.7	-2.348***	10.7	2.823***	38.7	5.334***
	MC	79	25.3		0.4		3.8	
Mondrian	CR	79	11.4	-2.600***	11.4	1.712	25.3	3.422***
	MC	75	28.0		4.0		5.3	

Note: *p < .05. **p < .01. ***p < .001.

To examine the extent to which factors such as school success and class can influence the success of solving tasks, in addition to the format of the task, the grade in mathematics, and the class the students attend was also included in the logistic regression model, whereby these factors were treated as potential confounders (Table 3).

The results of the analysis showed that the grade in mathematics is a statistically significant predictor of success in eight of the nine observed tasks. Since all regression coefficients are positive, it can be concluded that students with better academic performance were more likely to solve the tasks correctly, regardless of the format in which the tasks were presented.

When it comes to task format, it was a significant factor in seven of the nine tasks. In six of these seven cases, the success of solving was higher in the multiple choice (MC) formats.

The grade effect was statistically significant on all tasks except the *Angle* task, with older students generally performing better. This finding indicates that the difference in educational level, that is the degree of mastery of the relevant material affects the probability of correctly solving the tasks, regardless of the format.

Table 3. Logistic regression models for the nine tasks

Factor	Chord	Angle	Rhombus	Trapezoid	Color	Hypotenuse	Staircase	Triangles	Mondrian
Format									
CR	-.698*** (.244) [b]	-.017 (.262)	-1.881*** (.352)	-1.161*** (.294)	.552* (.308)	-1.216*** [a] (.412)	-1.073*** (.391)	-1.146** (.468)	-.882*** (.312)
Math grade	.351*** (.117)	.401*** (.116)	.426*** (.148)	.802*** (.159)	1.356*** (.211)	.965*** (.188)	.606*** (.183)	.134 (.208)	.783*** (.181)
Grade level									
7th	-.851* (.341)	-.232 (.413)	-.800* (.451)	-.788** (.389)	-1.190*** (.407)	-.565 (.410)	-.709* (.386)	.535 (.440)	-.914** (.445)
6th	-.515 (.372)	-.066 (.406)	-1.164*** (.442)	-.906** (.414)	-.704 (.440)				-.377 (.449)
5th	-.629** (.368)	-.484 (.376)	-1.734*** (.442)	-1.972*** (0.414)	-2.427*** (.453)				-.969** (.407)
const.	-1.328** (.539)	-.005 (.510)	-1.528** (.655)	-2.976*** (.704)	-6.396*** (.968)	-2.012*** (.695)	-2.328*** (.782)	-1.858** (.852)	-3.833*** (.840)
HL [c]	10.202 (.251)	4.548 (.805)	11.031 (.137)	3.852 (.870)	19.758 (.011)	1.455 (.984)	12.393 (.135)	14.840 (.062)	8.965 (.345)

Note: baseline category: MC (format); 8th (Grade level)
[a] parameter estimate, [b] Standard error, [c] HL – Hosmer and Lemeshow’s test, Statistic-value of chi-square statistic. *p < .05. **p < .01. ***p < .001.

To examine whether there is a relationship between the format of the task and the giving of an explanation, regardless of whether the answer is correct or incorrect, a chi-square test of independence was applied. The results show that there is a statistically significant relationship between the formats of the tasks *Rhombus* ($\chi^2 = 5.37$, $p < .001$), *Color* ($\chi^2 = 17.75$, $p < .001$) and *Triangles* ($\chi^2 = 10.95$, $p < .001$) and giving explanations in the task, i.e. in all three CR format tasks, students gave explanations more often.

Table 4 shows the representation of the selection of alternatives in solving MC tasks that contain four alternatives, of which only one is correct. The analysis shows that students most often eliminate one, less often two alternatives, but do not systematically check the offered answers in relation to the task conditions.

Students were significantly less likely to choose those alternatives that were obviously incorrect. The *Hypotenuse*, *Chord*, and *Mondrian* problems contain at least one such clearly incorrect alternative, with less than 6% of students choosing those alternatives. In contrast, in the *Triangle* task, where none of the incorrect alternatives were easily eliminated, it was observed that each incorrect alternative offered was chosen by more than 10% of the students. Students rarely use the picture to evaluate the solutions offered as a way to choose the correct answer.

In the *Staircase* task, students could visually estimate the length of the segments, but more than 10% of students chose the incorrect alternatives. 24% of students applied the Pythagorean theorem in the *Rhombus* task, and 28.6% in the *Trapezoid* task, even though it was not necessary.

Table 4. Distribution of choice of alternatives of 7th and 8th grade students on nine MC tasks

Alternative	Chord n=74[a]	Angle n=79	Rhombus n=74	Trapezoid n=75	Color n=92	Hypotenuse n=76	Stairs n=76	Triangles n=79	Mondrian n=75
a	24.3[b]	7.6	8.1	14.7	18.5	3.9	11.8	34.2	**28.0**
b	24.3	3.8	18.9	9.3	16.3	**75.0**	21.1	**25.3**	46.7
v	**44.6**[c]	10.1	16.2	26.7	**58.7**	18.4	19.7	26.6	5.3
g	4.1	**77.2**	**40.5**	**44.0**	4.3	1.3	**42.1**	10.1	8.0

[a] number of students who solved the task in MC format
[b] percentage of students who chose the alternative, [c] the correct alternative is in bold

In the *Tangram* task, which has two correct answers, the CR variant was solved correctly by 60.9% of students, and the MCm variant by 41.1% of students. In the CR format, 6.5% of the students gave one correct answer, and in the MCm format, 45.6% of the students gave one correct answer. In the case of the *Tetris* CR variant, 7.6% of students solved it correctly, and the MCb variant, which has two correct answers, was solved by 26.7% of students. One correct answer was given by 11.1% of students by circling Figure 4. The *Gardener* task, with the solution procedure for both formats, was solved by 1.1% of students in the CR variant and 3.3% in the MCb variant. The MCb variant of the *Gardener* task shows that 54.4% of students filled in the table correctly, but did not show the solution procedure.

Discussion

The obtained results indicate that a higher percentage of correct answers in the MC format of the task cannot be interpreted solely as an indicator of a higher level of knowledge or understanding.

The high rate of solving MC tasks, while at the same time a significantly lower percentage of attempts in CR format tasks, suggests that students approach MC tasks even in situations where they are not sure of the solution, probably relying on a guessing strategy, which is in line with findings (Sangwin & Jones, 2017).

This interpretation is further supported by the fact that, after correcting the results to approximate the probability of correct responses due to guessing (Chan & Kennedy, 2002), the advantage of the MC format almost completely disappeared.

However, it is important to note that this correction is an approximation because it is based on assumptions about the probability of random selection and an even distribution of responses, which may not always correspond to the students' actual strategies. Therefore, the results after correction should be interpreted as an indicative rather than an absolute indicator of actual knowledge.

The behavior of students in skipping tasks, i.e., missing answers, can be considered expected considering that MC tasks are not negatively scored, so students, even when they did not know the correct solution, often chose an answer, which indicates the possible use of a guessing strategy.

Haladina (Haladyna, 2004) points out that in some tests, the subjects are distracted from guessing, so they usually skip a large number of tasks, which is not the case in this research. On the contrary, CR tasks require independent formulation of answers, and students, faced with uncertainty or lack of knowledge, more often decided to skip them.

Authors Chan and Kennedy (Chan & Kennedy, 2002) state that there are types of MC questions that offer no advantage over equivalent CR questions because the answers to CR questions are obvious, and therefore, such CR questions are no different from MC questions.

The *Angle* task, which in the CR format required only the input of a numerical value, was probably perceived as simpler and more familiar, so the lower non-response rate indicates that students more often try to solve the task when its form is clear and when it seems accessible. This indicates that the structure and demands of a task, not just its format, can significantly influence students' motivation to approach solving. The *Angle* task shows that declarative knowledge does not change as a function of format (Goecke et al., 2022).

Only in the case of the *Triangle* problem, the students more successful in solving the CR format, but only after the correction of the results. The mentioned task requires several steps of solving, that is, it cannot be solved quickly, which is not a characteristic of selected-response tasks (Zhouf, 2013).

The findings also indicate the importance of motivational factors in solving tasks. Tasks that require independent formulation of answers can be perceived by students as more demanding, especially when there is no external framework to offer them guidelines (such as answers offered in MC format). Such tasks can cause some students to feel insecure and lead to not trying to solve them at all. In contrast, MC tasks may seem more accessible because they imply that it is possible to solve without complete understanding, which may increase students' willingness to try to answer, even if by guessing.

The inclusion of grade and prior knowledge as control variables allowed differentiation of the effects of educational level and prior knowledge on solving success.

As expected, students in the fifth and sixth grades had lower results, which can be explained by the discrepancy between the tasks and the level of knowledge they had acquired by then. In addition, students with a higher grade in mathematics performed better.

These findings confirm that differences in performance are not only determined by the format of the task, but also by available solving methods, prior knowledge, and achievement in mathematics. However, the effect of task format remains significant even when grade and math grade are controlled for, further confirming that the differences between MC and CR formats are not simply due to differences in the knowledge level of students in different grades.

With the CR tasks *Rhombus, Color,* and *Triangles,* the explanation, that is, the solution procedure, is shown more often. This indicates that the students use a direct solving strategy, starting from the conditions of the task towards the goal, and they give an explanation that indicates the way of thinking.

Students presented procedures when they had no alternatives offered, which may indicate that the CR format increases motivation to engage in solving.

These results indicate that students primarily use the direct solving strategy, especially in the CR format, while the inverse strategy in the MC format is not dominant. Instead of analyzing the offered alternatives, students often just circle one of the answers intuitively, without checking and explaining. The low frequency of visual evaluation and the absence of systematic elimination indicate a limited use of the inverse approach.

This finding partly deviates from the results of (Sangwin & Jones, 2017), which indicated an asymmetry in cognitive strategies in different question formats, where multiple-choice often encourages inverse and constructed-response tasks a direct approach. Based on the obtained results, it can be assumed that the students in this sample did not develop flexibility in solving tasks or awareness of the different strategies they can use depending on the structure of the task.

The application of the Pythagorean theorem in tasks where it is not necessary may indicate automaticity in the solving process, where students resort to the most familiar formulas instead of analyzing the task's requirements.

This is an important finding that can serve as a basis for the development of teaching interventions aimed at choosing appropriate strategies.

Conclusion

The results show that task format affects not only the accuracy of students' responses but also their willingness to attempt problem-solving. The high rate of skipped items in the constructed-response (CR) format, compared to the near-universal answering of tasks in the multiple-choice (MC) format, points to the need for systematic development of students' self-confidence and their readiness to engage with tasks even when uncertain about the outcome. Moreover, the findings suggest that both task format and question phrasing can shape students' perception of difficulty and accessibility, which has direct implications for their motivation and engagement in mathematics instruction.

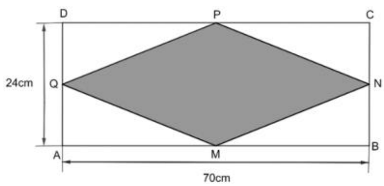
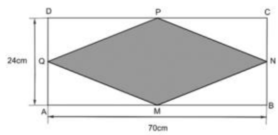
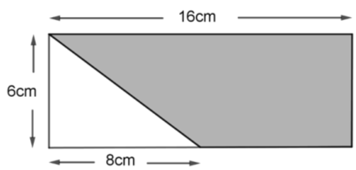
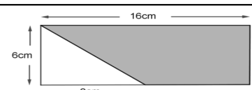
Students in this study predominantly rely on direct problem-solving processes, regardless of question format, both in multiple-choice (MC) and constructed-response (CR) tasks. Although MC tasks allow for the use of inverse strategies (e.g., evaluating the given answer options), students rarely adopt such approaches, even when they could potentially simplify the problem-solving process. The results underscore the need for mathematics instruction to systematically incorporate the examination and comparison of different problem-solving approaches, including both direct and inverse strategies.

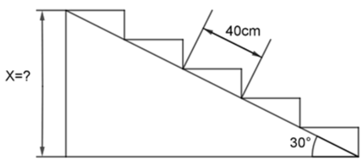
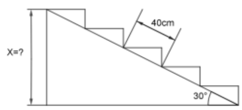
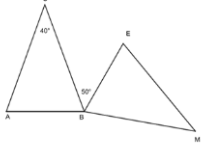
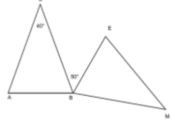
This pedagogical shift can be fostered through tasks that explicitly require students to analyze solutions, verify the accuracy of given answers, apply backward reasoning, and evaluate the efficiency of various procedures.

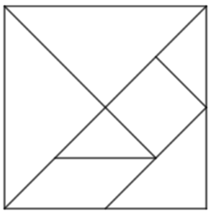
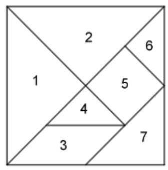
Particular emphasis should be placed on metacognitive guidance, enabling students to develop the ability to recognize when and why a specific problem-solving approach should be applied.Rather than relying solely on task format as a stimulus for problem-solving strategies, mathematics instruction should be designed to provide students with opportunities to actively compare different approaches, develop reflective understanding of their mathematical procedures, and become aware of the cognitive processes that lead to solutions.

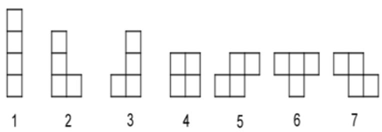
Future research should focus on a deeper analysis of motivational and metacognitive factors that influence students' choice of problem-solving strategies across different task formats. It is particularly important to examine how instructional programs and targeted interventions that promote the use of inverse approaches and foster students' self-confidence may affect their performance and attitudes toward mathematical tasks. In addition, further investigation is recommended into the impact of educational context and content on strategy selection and motivation. The inclusion of qualitative methods, such as interviews and classroom observations, could offer valuable insights into the underlying reasons and mechanisms that shape how students approach mathematical problem-solving.

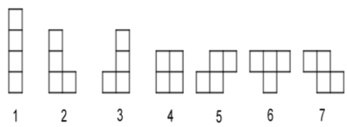
Table 1A. Tasks used in the study

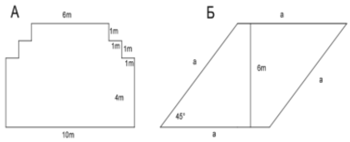
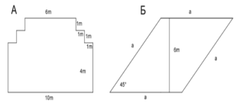
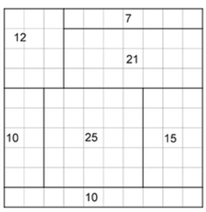
Task name	CR task text	MC task text
Chord	The line segment that joins any two points on the circle is called ___________________. Write the answer on the line.	Chord is: a) line segment which connects the center of the circle with a point on the circumference; b) line segment which has one point in common with the circle; v) line segment whose endpoints belong to the circumference; g) line segment which there are no points in common with the circumference. Circle the letter in front of the correct answer.
Angle	How many degrees does the straight angle have?	How many degrees does the straight angle have? a) $90°$ b) $360°$ v) $190°$ g) $180°$ Circle the letter in front of the correct answer.
Rhombus	Calculate the area of the shaded figure in the picture if the points M, N, P and Q are the midpoints of the sides of the rectangle ABCD. 	The points M,N,P and Q are the midpoints of the sides of the rectangle ABCD. What is the surface area of the shaded figure in the picture?  a) $210cm^2$ b) $1680cm^2$ v) $1470cm^2$ g) $840cm^2$ Circle the letter in front of the correct answer.
Trapezoid	 What is the area of the shaded part in cm^2 in the picture above?	 What is the area of the shaded part in cm^2 in the picture above? a) 24 b) 44 v) 48 g) 72 Circle the letter in front of the correct answer.

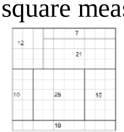
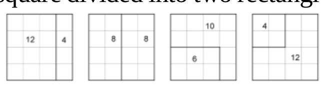
Task name	CR task text	MC task text
Color	Marko wants to paint a wall 4m long and 3.2m high. One liter of paint is required to paint a 5m^2 wall. The paint store sells paint in one-liter packages. What is the minimum number of liters of paint should Marko buy to paint the wall?	Marko wants to paint a wall 4m long and 3.2m high. One liter of paint is needed to paint a 5m^2 wall. The paint store sells different packs of paint and each has a volume written on it. Marko wants to buy the smallest package with which he can paint the wall. Which package will Marko buy? Circle the letter in front of the correct answer. a) 1l package. b) 2l package. v) 3l package. g) 4l package.
Hypotenuse	Calculate the length of the hypotenuse of a right-angled triangle whose legs are 4 cm and 5 cm.	The legs of a right-angled triangle are 4cm and 5 cm. What is the length of the hypotenuse? a) $3cm$ b) $\sqrt{41}cm$ v) $9cm$ g) $-\sqrt{41}cm$ Circle the letter in front of the correct answer.
Staircase	Nikola builds a staircase. All steps are of equal height and all treads are of equal width. What is the height X of the staircase? 	Nikola builds a staircase. All steps are of equal height and all treads are of equal width. What is the height X of the staircase? a) $50cm$ b) $80cm$ v) $200cm$ g) $100cm$  Circle the letter in front of the correct answer.
Triangles	Triangles *ABC* and *BEM* are isosceles (*AC*=*BC*, *BM*=*EM*) and congruent. Look at the picture! Determine the angle CAE. 	Triangles *ABC* and *BEM* are isosceles (*AC* = *BC*, *BM* = *EM*) and congruent. Look at the picture! Angle measure CAE is: a) 70°; b) 40°; v) 35°; g) 30°.  Circle the letter in front of the correct answer.

<table>
<tr><th>Task name</th><th>CR task text</th><th>MC task text</th></tr>
<tr><td>Tangram</td><td>One of the most famous Chinese puzzles is the tangram. Literally translated, tangram means seven boards of skill. It consists of seven parts called tani.
The goal of the tangram is to form a given shape from these seven tanas.
Mark the pairs of congruent triangles in the picture.
</td><td>One of the most famous Chinese puzzles is the tangram. Literally translated, tangram means seven boards of skill. It consists of seven parts called tani. The goal of the tangram is to form a given shape from these seven tanas.


Circle the letter in front of those statements that are correct.
a) Triangles 1 and 4 are congruent.
b) Triangles 4 and 6 are congruent.
v) Triangles 1 and 7 are congruent.
g) Triangles 1 and 2 are congruent.</td></tr>
<tr><td>Tetris</td><td>Tetris is a logical video game created on June 6, 1984.
The game uses pieces made up of four matching squares.
The figures used in the game of Tetris are shown in the picture.


How many axes of symmetry does each of the figures shown have? Write it in the table.
<table><tr><td>Figure</td><td>F1</td><td>F2</td><td>F3</td><td>F4</td><td>F5</td><td>F6</td><td>F7</td></tr><tr><td>Number of axes of symmetry</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table></td><td>Tetris is a logical video game created on June 6, 1984.
The game uses pieces made up of four matching squares.
The figures used in the game of Tetris are shown in the picture.
Circle the number below each of the figures shown that has more than one axis of symmetry.

</td></tr>
</table>

<table>
<tr><th>Task name</th><th>CR task text</th><th>MC task text</th></tr>
<tr><td>Gardener</td><td>The gardener wants to fence the garden. He is considering the following plans.
How many meters of fence are needed for the garden on plan A, and how many for the garden on plan Б? Which garden needs more meters of fence?

</td><td>A gardener has 32 meters of fence and wants to border the garden. He is considering the following plans.

Circle either Yes or No on both plans to indicate whether or not the gardener can border the garden with 32 meters of fencing.
<table><tr><th>Garden plan</th><th>Using this plan, can the border be made with 32 meters of fence?</th></tr><tr><td>Plan A</td><td>Yes / No</td></tr><tr><td>Plan Б</td><td>Yes / No</td></tr></table></td></tr>
<tr><td>Mondrian</td><td>The famous Dutch abstract painter and art theorist Piet Mondrian inspired mathematicians to come up with an interesting mathematical problem. A Mondrian square is a square whose side length is a natural number, and is divided into rectangles such that each of them has different dimensions, which are also natural numbers. Rectangles with dimensions 3×4 and 4×3 are considered rectangles of the same dimensions. The picture shows an example of a Mondrian square measuring 10×10.


Draw a 4×4 Mondrian square divided into two rectangles.</td><td>The famous Dutch abstract painter and art theorist Piet Mondrian inspired mathematicians to come up with an interesting mathematical problem. A Mondrian square is a square whose side length is a natural number, and is divided into rectangles such that each of them has different dimensions, which are also natural numbers. Rectangles with dimensions 3×4 and 4×3 are considered rectangles of the same dimensions. The picture shows an example of a Mondrian square measuring 10×10.

Circle the letter below the picture that represents Mondrian's 4x4 square divided into two rectangles.


a) b) v) g)</td></tr>
</table>

References

Andrà, C. & Magnano, G. (2011). Multiple-choice math tests: should we worry about guessing? *Quaderni Di Ricerca in Didattica (Mathematics)*, 21, 235–243.

Anić, I., Košanin, R. i Savković, J. (2020). *Matematika 7: udžbenik sa zbirkom zadataka. Deo 2* [*Mathematics 7: Textbook with Workbook. Part 2*]. Beograd: BIGZ školstvo d.o.o.

Barton, C. (2018). On Formative Assessment in Math: How Diagnostic Questions Can Help. *American Educator*, 42(2), 33–43. https://eric.ed.gov/?id=EJ1182085

Birenbaum, M. & Tatsuoka, K. K. (1987). Open-Ended Versus Multiple-Choice Response Formats – It Does Make a Difference for Diagnostic Purpose. *Applied Psychological Measurement*, 11(4), 385–395. https://doi.org/10.1177/014662168701100404

Bonner, S. M. (2013). Mathematics strategy use in solving test items in varied formats. *Journal of Experimental Education*, 81(3), 409–428. https://doi.org/10.1080/00220973.2012.727886

Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement*, 29(3), 253–271. https://doi.org/10.1111/j.1745-3984.1992.tb00377.x

Canogullari, A. & Radmehr, F. (2025). Task design principles in mathematics education: a literature review. *International Journal of Mathematical Education in Science and Technology*, 1–33. https://doi.org/10.1080/0020739X.2025.2457365

Chan, N. & Kennedy, P. E. (2002). Are Multiple-Choice Exams Easier for Economics Students? A Comparison of Multiple-Choice and "Equivalent" Constructed-Response Exam Questions. *Southern Economic Journal*, 68(4), 957–971. https://doi.org/10.1002/j.2325-8012.2002.tb00469.x

Chaoui, N. A. (2011). *Finding relationships between multiple-choice math tests and their stem-equivalent constructed responses*. The Claremont Graduate University. https://doi.org/10.5642/cguetd/21

Clay, B. (2001). *Is This a Trick Question? A Short Guide to Writing Effective Test Questions*. Kansas Curriculum Center.

Foy, P., Arora, A. & Stanco, G. M. (Eds.) (2013). *TIMSS 2011 User Guide for the International Database*. TIMSS & PIRLS International Study Center and International Association for the Evaluation of Educational Achievement (IEA), Lynch School of Education, Boston College, Chestnut Hill. http://timssandpirls.bc.edu/timss2011/international-database.html

Goecke, B., Staab, M., Schittenhelm, C. & Wilhelm, O. (2022). Stop Worrying about Multiple-Choice: Fact Knowledge Does Not Change with Response Format. *Journal of Intelligence*, 10(4), 1–22. https://doi.org/10.3390/jintelligence10040102

Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items* (3rd ed.). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9780203825945

Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5

Hollingworth, L., Beard, J. J., Proctor, T. P. (2007). An investigation of item type in a standards-based assessment. *Practical Assessment, Research and Evaluation*, 12(1). https://doi.org/10.7275/6ggz-8837

Hubbard, J. K., Potts, M. A. & Couch, B. A. (2017). How question types reveal student thinking: An experimental comparison of multiple-true-false and free-response formats. *CBE-Life Sciences Education*, 16(2), 1–13. https://doi.org/10.1187/cbe.16-12-0339

Katz, I. R., Bennett, R. E. & Berger, A. E. (2000). Effects of response format on difficulty of SAT-mathematics items: It's not the strategy. *Journal of Educational Measurement*, 37(1), 39–57. https://doi.org/10.1111/j.1745-3984.2000.tb01075.x

Kusumawati, M. & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REID (Research and Evaluation in Education)*, 4(1), 70–78. https://doi.org/10.21831/reid.v4i1.20202

Liou, P. Y. & Bulut, O. (2020). The Effects of Item Format and Cognitive Domain on Students' Science Performance in TIMSS 2011. *Research in Science Education*, 50(1), 99–121. https://doi.org/10.1007/s11165-017-9682-7

Livingston, S. A. (2009). Constructed-Response Test Questions: Why We Use Them; How We Score Them. *R & D Connections*, 11, 1–8. https://eric.ed.gov/?id=ED507802

Marcq, K., Donayre, E. J. C. & Braeken, J. (2024). The role of item format in the PISA 2018 mathematics literacy assessment: A cross-country study. *Studies in Educational Evaluation*, 83. https://doi.org/10.1016/j.stueduc.2024.101401

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207–218. https://doi.org/10.1207/s15326985ep3404_2

Martins, L. G. & Martinho, M. H. (2024). Types of tasks in Mathematics textbooks: A study with Portuguese's textbooks of 10th and 11th grades. *Educacion Matematica*, 36(1), 66–91. https://doi.org/10.24844/EM3601.03

Matematički list za učenike osnovnih škola [*Mathematical Journal for Primary School Students*], 2017(2). Beograd: Društvo matematičara Srbije.

O'Neil Jr., H. F. & Brown, R. S. (1998). Differential Effects of Question Formats in Math Assessment on Metacognition and Affect. *Applied Measurement in Education*, 11(4), 331–351. https://doi.org/10.1207/s15324818ame1104_3

Özdemir, A. Z. & Toker, Z. (2025). Analysis of distractors in mathematics questions and their potential to lead misconceptions. *Thinking Skills and Creativity*, 56. https://doi.org/10.1016/j.tsc.2024.101730

Pavlović Babić, D. & Baucal, A. (2009). *Matematička pismenost PISA 2003 i PISA 2006* [*Mathematical Literacy in PISA 2003 and PISA 2006*]. Beograd: Ministarstvo prosvete Republike Srbije – Zavod za vrednovanje kvaliteta obrazovanja i vaspitanja – Institut za psihologiju Filozofskog fakulteta Univerziteta u Beogradu.

Photopoulos, P., Tsonos, C., Stavrakas, I. & Triantis, D. (2021). Preference for Multiple Choice and Constructed Response Exams for Engineering Students with and without Learning Difficulties. *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU 2021)*, 1(April), 220–231. https://doi.org/10.5220/0010462502200231

Prídavková, A. (2023). The Solving Strategies of Mathematical Tasks Used By Prospective Primary Education Teachers. *INTED2023 Proceedings*, 1(March), 4800–4805. https://doi.org/10.21125/inted.2023.1252

Sangwin, C. J. & Jones, I. (2017). Asymmetry in student achievement on multiple-choice and constructed-response items in reversible mathematics processes. *Educational Studies in Mathematics*, 94(2), 205–222. https://doi.org/10.1007/s10649-016-9725-4

Schult, J. & Sparfeldt, J. R. (2018). Reliability and Validity of PIRLS and TIMSS: Does the Response Format Matter? *European Journal of Psychological Assessment*, 34(4), 258–269. https://doi.org/10.1027/1015-5759/a000338

Stankous, N. V. (2016). Constructive Response Vs. Multiple-Choice Tests In Math: American Experience And Discussion (Review). *European Scientific Journal*, Special Edition, 308–316.

Štěpánková, B. & Emanovský, P. (2011). On Open-Ended and Closed-Ended Questions in Didactic Tests of Mathematics. *Problems of Education in the 21st Century*, 28, 114–122.

Verbić, S. Ž. (2013). *Heurisitke za maksimizaciju informacione vrednosti računarskih testova znanja* [*Heuristics for the maximization of computer knowledge tests information value*]. Doktorska disertacija. Univerzitet u Beogradu.

Vidović, N., Stanojević, G., Stuparević, Z., Stanojević, V., Vračar, L. i Stančić, M. (2015). *Zbirka zadataka za završni ispit – osnovna škola, MATEMATIKA, Vežbam za malu maturu* [*Mathematics Exercise Book for the Primary School Graduation Exam*]. Beograd: Kreativni centar.

Watson, A. & Ohtani, M. (2015). *New ICMI Study Series Task Design In Mathematics Education an ICMI study 22.* http://www.mathunion.org/ICMI/

Zhouf, J. (2013). Categorisation of Multiple-choice Problems. *Procedia – Social and Behavioral Sciences*, 93, 592–596. https://doi.org/10.1016/j.sbspro.2013.09.244